School of Psychology

Bangor University

STATISTICAL LEARNING OF ORTHOGRAPHIC PATTERNS IN TYPICALLY DEVELOPING AND DYSLEXIC POPULATIONS

Anna Samara

Supervisors:
Dr Markéta Caravolas and Dr George Houghton

Submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Bangor University

October 2013

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: Anna Samara

Title: Statistical Learning of Orthographic Patterns in Typically Developing and Dyslexic Populations

Supervisor/Department: Dr Markéta Caravolas (School of Psychology)

Funding body (if any): FP7-Marie Curie Initial Training Network (ITN) Fellowship

Qualification/Degree: PhD in Psychology

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …… ***Anna Samara***……………………. (candidate)

Date: October 31st, 2013

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed …… ***Anna Samara***……………………. (candidate)

Date: October 31st, 2013

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic repositories, and for the title and summary to be made available to outside organisations.

Signed …… ***Anna Samara***……………………. (candidate)

Date: October 31st, 2013

<u>Statement 4</u>:

Choose **<u>one</u>** of the following options

a)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	√
b)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c)	I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: …… ***Anna Samara***……………………. Date: October 31st, 2013

Summary

There is growing interest in children's sensitivity to orthographic probabilistic constraints governing well-formed letter sequences in print. A few studies have demonstrated that young children's spellings conform to some untaught orthographic restrictions and have postulated that distributional statistical learning processes may underlie this ability. However, there are no studies investigating whether similar patterns can be learnt under experimental conditions whereby participants are not instructed to learn. Therefore, an incidental learning task was used in this thesis to investigate whether novel constraints on letter positions and letter contexts can be exploited and used in a subsequent legality discrimination task by 7-year-old children and adults. Results indicated that (a) novel letter positions and contexts were reliably learnt by children and adults. (b) Adults were, by and large, superior learners. (c) Children's and adults' ability to learn was similarly affected by pattern complexity. These findings confirm the statistical nature of children's sensitivity to general properties of their orthography. The next question addressed was whether implicit sensitivity to frequency-based information that is widely embedded in written language (allowable letter positions; probable/improbable letter pairs/triplets and larger units) is impaired in dyslexia. To this aim, variants of an implicit artificial grammar learning task were used among groups of dyslexic adults and skilled adult readers. It was demonstrated that (a) in letter versions of the task, most aspects of dyslexics' performance were spared relative to skilled readers' performance. (b) Importantly, in a nonlinguistic version, dyslexics' performance matched that of their skilled reader counterparts in every aspect. These results contrast with some previous studies of implicit learning in dyslexia and challenge the claim that a general learning deficit contributes to reading and spelling disability.

Acknowledgments

First, I would like to thank my supervisor, Dr Markéta Caravolas, for her academic advice and support throughout these years. I have benefitted tremendously from her academic expertise and I am grateful for the all the time she took to teach me. I greatly appreciate her trust in me and her understanding in the most challenging periods of this long journey.

I would also like to thank my second supervisor, Dr George Houghton, and the chair of my PhD committee, Dr Marie-Josephe Tainturier, for valuable comments over these years. I am thankful to Liz Du Pré from the Miles Dyslexia Center for help in recruiting participants. Many thanks are due to Sam Crewe, Laura Patricia Haggett and Sarah Whitehead for assistance in collecting part of the child data and help with the data entry. Andrea Reynolds has been a lovely and patient officemate in the most challenging months of the PhD.

Many good friends have contributed to the completion of this PhD. Eleni Magklara and George O'Leary, I really don't think I would have made it sane to the end without you. I'm truly indebted for your support and encouragement during these years, particularly in difficulties times (including but not limited to the last couple of weeks before the submission). I am grateful to Betty Mousikou and Naymé Salas for being a phone call away whenever I needed their help, support, or advice. May Sabry, Yougan Saman and Nishana Saman, thank you so much for your faith in me and for being there whenever I needed a city break. Further thanks to my good friends Elli Tsotsoli, Chara Oraiopoulou, Vasilis Tsiplakis, Zoe Nouvaki, and Elena Papagianopoulou, for insisting on checking on me and being very understanding when I was not answering emails or picking up the phone for weeks!

This huge undertaking would not have been possible without the help and support of my parents, Dimitris and Katerina, and my brother Vasilis. Thank you for believing in me, for being there and for listening to everything I had to say, even when it didn't make much sense. I would not have made it without your love.

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

The ability of humans to detect environmental structure and to adjust their behavior accordingly has been of long interest to cognitive psychologists. For several decades it has been known that people can and do encode basic statistical information (e.g., frequency of event occurrences), even in cases when they are not explicitly told or encouraged to do so (Hasher & Zacks, 1979, 1984). There are many early empirical demonstrations of this phenomenon, some of which are relevant to the domain of language. Attneave (1953), for example, demonstrated a high degree of correlation between adults' estimated frequencies of single letter occurrences (e.g., *how often does "s" occur in text*?) and actual letter frequencies. Along similar lines, Shapiro (1969) showed that close to 85% of the variance in adults' relative estimates of frequency of single word occurrences was accounted by measures of actual word frequency. It is now agreed that such demonstrations are but one example of the various statistical computations performed by humans. Not only are humans naturally predisposed to compute the simpler statistics (e.g., estimating frequency of element occurrences), but they are also adept at performing more sophisticated (e.g., conditional probabilistic) computations from the first few months of life. Infants' and adults' ability to exploit statistical structure embedded within various linguistic and nonlinguistic domains is now extensively studied under the rubric of *statistical learning* research. It is also becoming clear that sensitivity to structural properties is not necessarily a result of deliberate "mathematical" calculations, an idea supported by the findings of an extensive body of *implicit learning* research—initiated by Arthur Reber's (1967) pioneering work. Written language is a prime example of a patterned domain of knowledge, embedding "regularities" and constraints that are well explained in statistical terms. While some of these are explicitly taught as mnemonics for spelling (e.g., i before e except after c), as pattern complexity increases, knowledge of this sort may be less available to verbalization, even among the most proficient linguists (Kessler, 2009). Consequently, explicit skill learning may prove a less suited means of acquiring complex orthographic patterns (Steffler, 2001). This thesis combines insights from the statistical and implicit learning research streams to investigate developing children's, "typical" (i.e., literacy unimpaired) adults', and "atypical" (i.e., dyslexic) adults' ability to learn patterns like those embedded in written language.

Four lines of research relevant to the questions addressed and the methods employed in this thesis are discussed in the following sections. The chapter begins with a general overview of the statistical learning stream of research highlighting (a) the impressive range of statistical computations that can be performed by infants and adults (*what is learnable?)* and (b) their usefulness for different aspects of language acquisition. The second part of the chapter documents the emergence of a statistical learning perspective in spelling development research and presents some evidence showing that children are sensitive to statistically defined orthographic patterns, as well as conditional probabilities governing phoneme (sound)-to-grapheme (letter) correspondences from the very beginnings of their literate lives. Part three of this chapter summarizes key findings from a relevant, but somewhat independent line of research (implicit learning) that has also sought to assess "learning without knowing that you are learning" among adults, but has been more directly pre-occupied with the nature of the mental representations underlying learning performance. The fourth and final part of the chapter reviews a line of research that has sought to draw links between the ability to learn implicitly and the literacy difficulties experienced by dyslexic individuals.

1.1 Statistical Learning Research

The statistical learning approach to language acquisition originates in Saffran and colleagues' demonstration that preverbal infants may be able to detect which sound sequences are words in their native language—and discriminate them from those that are not—by adopting an impressively sophisticated statistical approach: Attending to the transitional probability statistic, that is, the conditional probability of Y given X in the sequence XY (Miller & Selfridge, 1950). The frequently used *prettybaby* example provides a nice illustration of how this statistic may serve as an informative cue to boundaries in the spoken sequence “pretty baby” (/prɪti/ /beɪbi/; (*which one is a word, pretty or tyba?*)[1]. /ti/ is word-final syllable, and as such, it can be followed by any of the syllables that begin words in English. As a result, the probability in which /ti/ is followed by the syllable /beɪ/ is very low (thus, uninformative). Critically though, the probability that the syllable /prɪ/ is followed by the syllable /ti/ is very high—and especially so in an infant-directed speech corpus. Therefore, attending to this difference reliably indicates that /prɪti/ is more likely to be a word than /tibeɪ/ (Saffran, 2003).

In the following sections, a number of artificial language learning studies are reviewed, demonstrating that infants are indeed able to learn patterns as complex as that illustrated above—despite the presence of some constraints on what is learnable. While a few different descriptions and definitions of statistical learning have been proposed over the last twenty years (for examples, see Perruchet & Pacton, 2006; Turk-Browne, Scholl, Chun, & Johnson, 2009), the one adopted here is that of a learning device forming the basis of humans'—among other species'—ability to extract and represent environmental regularities embedded in different domains and modalities (Aslin & Newport, 2012; Conway & Christiansen, 2002).

1.1.1 (Auditory) Statistical Language Learning

Identifying words in continuous speech. Word segmentation is a notoriously complicated problem: Everyday natural speech contains several cues to word beginnings and endings, yet they are all partially unreliable (Cole & Jakimik, 1980). Nevertheless, infants are able to infer word boundaries in fluent speech passages from their first 7.5 months (Jusczyk & Aslin, 1995). To explore the hypothesis that some transitional probabilistic (i.e., statistical) knowledge is brought to this surprisingly early emerging ability, Saffran, Aslin and Newport (1996a) exposed 8-month old infants to a continuous artificial language stream composed of four trisyllabic nonwords (e.g., bidakupadotigolabubidaku…) over a brief (2-minute) period. The language contained no audible pauses or intonation contour. In fact, the only pattern that served as a cue for discriminating between “words” such as bidaku and part-words such as dakupa was

---

[1]Phonemes are denoted with IPA symbols; see International Phonetic Association (1996, 1999).

the transitional probability between the syllables bi-da and the da-ku (transitional probability = 1.0), which was significantly higher than the probability among any other adjacent syllables, such as ku-pa or bu-bi (transitional probability = 0.33). Once the familiarization training phase was complete, infants were presented with two sets of three-syllable sequences. In one set, sequences spanned the word boundaries (part words), whereas sequences in the other set fell within the word boundaries (words). When infants' listening times to the sequences were analyzed, a statistically significant difference in favor of the less probable (unfamiliar) three-syllable part-word sequences was observed (novelty effect), showing that words were reliably segmented.

A large body of subsequent work has replicated and extended this important finding in several different ways. An important question concerns the exact statistical computations underlying segmentation performance in Saffran et al.'s (1996a) study and was addressed by Aslin, Saffran, and Newport (1998). High transitional probabilities between syllable pairs (i.e., syllable X predicted with high probability that syllable Y would occur) systematically co varied with frequent syllable pair occurrences (i.e., XY co occurred frequently) in Saffran and colleagues' (1996a) original material. Thus, infants' longer listening times for part words may have reflected sensitivity to either or both statistical cues. To establish that infants were capable of extracting the more complex conditional probability statistic, Aslin et al. (1998) presented similar-age (i.e., 8-month-old) infants with a new speech stream embedding syllable- and part-words that were equated in terms of the (simpler) frequency of occurrence statistic. Word/part word discrimination accuracy was once again reliable, confirming the unique contribution of transitional probability sensitivity to infants' word segmentation performance.

It is not only infants who are sensitive to transitional probabilities and use them for the purpose of word segmentation. Saffran, Newport, and Aslin (1996b) tested adults' ability to compute and exploit similar information following 21 minutes of exposure to an artificial speech stream. Adults' ability to segment words was assessed by means of a forced 2 choice alternative test. Furthermore, the segmentation task was somewhat more complicated than the one used with preverbal infants. While the transitional probabilities among syllables spanning boundaries were close to zero (.10 - .20), word-internal transitional probabilities were not consistently 1.00, as in the infant studies, but varied between .31 and 1.00. Adults performed with significantly better than chance accuracy. Furthermore, words containing higher transitional probabilities (ranging from .75 to 1.00) were responded to more accurately than words with relatively lower transitional probabilities (ranging from .37 to .50). While adults are capable of performing the word segmentation task on the basis of conditional probabilistic information, there are not necessarily the most superior learners, as it may be expected intuitively. As demonstrated by Saffran, Newport, Aslin, Tunick, and Barrueco (1997), when adults and 6- to 7-year-old children are confronted with the same segmentation task, children are able to perform as accurately as adults.

Mastering natural language involves more than extracting statistical patterns among adjacent elements such as those originally studied by Saffran and collaborators. Languages also embed several remotely connected dependencies. For example, vowels in C_1VC_2e words where "e" is silent (e.g., r*i*d*e*, c*a*r*e*) receive a long pronunciation *regardless* of the identity of the intermediate C_2 consonant (Stanback, 1992). Successful syntactic processing also relies on the discovery of nonadjacent dependencies, such as those between auxiliary verbs and inflectional morphemes (e.g., the boy *is* play*ing*), or those between sentence subjects and verbs far away from each other (e.g., *the rocks* on the beach *are* jagged) (Gómez, 2002; Newport & Aslin, 2004). Therefore, a question pertinent to statistical language learning involves the extent to which nonadjacent dependencies are also attended to, and can be utilized in the context of word boundary extraction. Newport and Aslin (2004) constructed twenty trisyllabic words of the form $CV_1CV_{2x}CV_3$ and mixed them into a pseudorandom stream of continuous speech where only the relationships between nonadjacent syllables (e.g., CV_1CV_3) were sufficient for syllable-word discovery. The transitional conditional probabilities among adjacent syllables—within- *as well as* between-words—were in fact all very low (between .20 and .25). Transitional probabilities among nonadjacent syllables between words (i.e., those syllables spanning word boundaries) were also low (again, ranging from .20 to .25), but those among the first and last syllable within words were systematically 1. Given that this was the only cue to word boundaries, reliable segmentation between words and part words would indicate that adults are sensitive to nonadjacent probabilistic information. Against this expectation, following approximately 20 minutes of familiarization, participants' performance did not exceed chance levels (~ 48% mean accuracy) and several manipulations (e.g., increasing the length of exposure; simplifying the language by reducing the number of syllable frames, etc.) proved unsuccessful in inducing learning. Learning was, on the other hand, possible when the nonadjacent contingencies were built among phonemic segments rather than syllables (consonants separated by vowels: experiment 2; vowels separated by consonants: experiment 3), as it also frequently the case in natural languages such Hebrew or Arabic. It appears that while statistical learning processes are not limited to the computation of information among neighboring elements, it may best suited for the acquisition of patterns that resemble those found in different linguistic inputs (see Newport & Aslin, 2004, for an extensive discussion of alternative explanations for the relative ease of nonadjacent segment learning compared to nonadjacent syllable learning).

Statistical learning of other aspects of linguistic structure. The studies reviewed above suggest that infants, as well as adults, can identify lexical elements on the basis of adjacent and some nonadjacent transitional probabilities. More recent artificial language research suggests that phonotactic patterns—that is, patterns defining how phonemes occur and co-occur in one's spoken language—are also readily acquired via similar statistical learning processes (e.g., Chambers, Onishi, & Fisher, 2003; see Chapter 2 for an extensive review) and can be also used by infants as young as 9-months-old for the purpose of parsing words in continuous speech streams (e.g., Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001).

Statistical cues convey information that may be useful for several tasks other than word segmentation. Recently, statistical learning accounts have been put forward to explain the acquisition of other (low-level, as well as high-level) aspects of linguistic structure. For example, there is considerable evidence that 1-year-old infants' ability to discriminate only between their native language phonetic units (e.g., /r/ vs. /l/ is a "valid" phonetic unit contrast that can alter word meaning for English speakers, but not Japanese speakers) and form phonemic categories (e.g., for Japanese—but not English speakers—/r/ and /l/ are grouped into the same phonemic category) is shaped by their sensitivity to the statistical distribution of sounds in their native language (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002). Statistical cues are, in addition, powerful tools for uncovering higher-level structure, such as extracting grammatical category information (e.g., truck is a noun) or rudimentary syntax (the ordering of words in a sentence; Gómez & Gerken, 1999; Saffran & Wilson, 2003). Redington, Chater and Finch (1998, among others, e.g., Mintz, Newport, & Bever, 2002) provided computational evidence for the former, by showing that analyzing the distribution of contexts within which words occur in a corpus of adult speech can lead to reliable grouping reflecting grammatical category membership. Regarding the acquisition of rudimentary syntax, Saffran and Wilson (2003) showed that 1-year-old infants familiarized with multiword sentences (each sentence was created by concatenating five disyllabic words into a continuous speech stream, e.g., datopidubutobadudipa; sentences were separated by a brief pause), succeeded in segmenting the words on the basis of transitional probabilities (1.00 vs. .25) and further extracted allowable word orders by tracking transitional probabilities at the word level (which were independent of the syllabic transitional probability information).

1.1.2 Other Auditory Statistical Learning

Further research has examined human's ability to perform statistical computations over nonlinguistic auditory sequences. In recent studies, the ability to extract "tone words"—rather than syllable words—has been reliably observed with materials of varying complexity (e.g., Creel, Newport, & Aslin, 2004; Gebhart, Newport, & Aslin, 2009; Saffran, Johnson, Aslin, & Newport, 1999; Tillmann & McAdams, 2004). Using a direct nonlinguistic analogue of the original speech segmentation task (Saffran et al., 1996a), Saffran et al. (1999) demonstrated that 8-month-old infants and adults are capable parsers not only of linguistic material (continuous speech stream), but also of tone material (continuous tone stream). Extending Newport and Aslin's (2004) findings, Creel et al. (2004) provided evidence of some contingency learning among remotely connected musical tones. However, as in the Newport and Aslin's study (2004), nonadjacent learning was found to be highly selective. When the musical elements cueing the boundaries were perceptually dissimilar, i.e., varied significantly in terms of pitch class and timbre, adults' ability to discriminate between nonadjacent predictable tone sequences and random tone triplets did not reliably exceed chance levels (experiment 1). Even though the contingencies appearing every two tones were not learnt, it worth noting that participants reliably discriminated between adjacent tone sequences and random triplets on the basis of low transitional probabilities among tones occurring one next to another; therefore, adjacent contingency learning did occur. Evidence of nonadjacent learning was obtained when the remote tones cueing boundaries were perceptually similar among the auditory dimensions of pitch (e.g., statistical cues were paralleled by a change in tones' octave; experiment 2) or timbre (experiment 3).

Studies such as those by Creel et al. (2004) suggest that (a) statistical learning processes operate at least across two domains (language; music) in the auditory modality; and (b) some of the constraints on what is easily learnable are shared across linguistic and nonlinguistic domains (and may not simply reflect sensitivity to common linguistic structure, as Newport and Aslin, 2004, originally speculated). The domain generality of statistical learning is further suggested by studies showing that infants and adults can learn patterned relationships in a different modality. Some demonstrations of visual statistical learning are presented below.

1.1.3 Visual Statistical Learning

Another "hint" that statistical learning processes are not uniquely sensitive to language-related statistical cues comes from findings of learning in the visual domain. Drawing on the findings of auditory statistical learning research, several studies have investigated whether similar computations may underlie the acquisition of nonlinguistic structure embedded in the visual modality. Kirkham, Slemmer, and Johnson (2002) familiarized preverbal infants (2-, 5-, and 8-month-old) with a continuous visual stream embedding high transitional probabilities between certain pairs of adjacent shapes (e.g., the probability of a yellow circle being followed by a pink diamond was 1.00; transitional probabilities between the remaining [i.e., random] shapes pairs was 0.33). At test, infants showed a preference for the low probability (novel) pairs rather than the highly probable pairs, which was reliable even among the youngest (2-month-old) group of infants. In a more stringent examination of infants' statistical learning abilities in the visual domain, Marcovitch and Lewkowicz (2009) provided evidence of significant learning among 4.5-month-old infants (but not 2.5-month olds) when pair conditional probabilities were manipulated independently of pair frequencies (i.e., similar to Aslin et al., 1998). Test-phase pairs varying solely in terms of the simpler statistic (i.e., pair frequencies over and above pair conditional probabilities) were also reliably discriminated by infants in the same group. Sensitivity to both temporal visual statistical cues appears to emerge early on in life.

Another two well-cited studies have assessed adults' sensitivity to temporal and spatial relationships among visually presented items (Fiser & Aslin 2001, 2002). Fiser and Aslin (2002) showed that, following passive viewing of a continuous temporal sequence of visually presented abstract shapes (i.e., A-B-C-D-E-F-G-H-I), adults were able to discriminate, on the basis of joint probabilistic contingencies between (a) "base triplets" such as ABC and impossible triplets (e.g., AFG); and (b) "base triplets" (i.e., ABC) and some less probable "part triplets" (e.g., CDE). In a final experiment, conditional probabilistic contingencies were also reliably learnt and used to discriminate between "base shape pairs" (B will occur given that A has occurred) and less frequent "part pairs" (E will occur given that D has occurred) (Fiser & Aslin, 2002). To investigate adults' sensitivity to spatial structure, Fiser and Aslin (2001) familiarized adult participants with 3 x 3 grids (i.e., complex multi-element scenes created by using shapes similar to those used in the previous study) embedding co-occurrence and predictive shape-shape relationships. For example, during familiarization, shapes A and B constituted a base pair and always appeared in a certain spatial (e.g., horizontal) configuration (base pair) within the grid. Following a few minutes of passive viewing (a total of 144 multi-element scenes presented for 2 second each), a two alternative forced choice test was used to compare participants' sense of familiarity of probable/improbable shape pairs—organized horizontally, vertically, or obliquely within a similar grid. Participants showed a preference for base pairs embedding high joint probabilities of shape co-occurrences (experiments 1 and 2), even when an additional cue to probable/unprobable absolute shape positions (i.e., their configuration within the grid) was removed. Furthermore, when the test phase pairs differed only in terms of the higher-order conditional probability statistic (and the co-occurrence frequency information statistic was made unreliable; experiment 3), adults reliably discriminated between cross- and base-shape pairs and favored the latter in their judgments of familiarity.

Combined with the evidence from the tone segmentation studies, visual statistical learning research confirms the idea that both linguistic and nonlinguistic structure can be extracted on the basis of a general learning mechanism operating across different domains and sensory modalities. The implications of these findings for humans' ability to learn patterned relationships in the *visual orthographic* domain are considered in section 1.2.

1.1.4 Statistics versus Abstract Rules

An overview of the empirical literature on statistical learning would be incomplete without referring to a body of research investigating whether infants can learn *more* than transitional or conditional probability statistics. Put simply: is there evidence of abstract rule learning—allowing generalization of learnt structure to new instances—in infancy? This question was addressed by a few studies using test phase stimuli that either conformed or did not conform to the underlying structure governing element orders in the familiarization set, but in addition, they *all* differed on a surface level from the familiarization set (i.e., they were novel instances of conforming/nonconforming stimuli). Gómez and Gerken (1999) assessed rule learning among 1-year-old infants using stimuli generated from a miniature artificial (finite-state) grammar (see Figure 1.1 for an example of such a synthetic system; an extensive body of research demonstrating patterned learning among adults using similar stimuli is reviewed in section 1.3). In one experiment (Experiment 4), infants were presented over a brief period with lists of 3- to 6-syllable auditory sequences such as JED FIM TUP or JED FIM TUP DAK SOG DAK (in a sense, these can be conceived as "sentences" embedding word-order relationships, that is the syllable-word JED can be followed by the syllable-word FIM; similar to Saffran and Wilson's, 2003 material). At test, half of the sequences were "grammatically ordered" (i.e., followed one of the transitions of the finite-state system), whereas the other half violated the "rules" of the system governing sequence order. Critically, both grammatical and ungrammatical items were generated by mapping the training syllables into a different vocabulary (e.g., JED was mapped onto VOT; FIM was mapped onto PEL etc.). Therefore, children's ability to discriminate between sequences such as PEL TAM JIC RUD TAM (grammatical) and VOT RUD JIC VOT RUD (ungrammatical) could not be accounted for their sensitivity to joint transitional probabilities, all of which were zero. Gómez and Gerken (1999) demonstrated above chance discrimination of such strings instantiated in a new vocabulary among 1-year-old infants and suggested that learning was abstracted beyond specific pairs of words.

Marcus, Vijayan, Rao and Vishton (1999) further tested 7-month-old infants' ability to acquire abstract "algebra-like" rules (as opposed to statistics) by exposing them to a few repetitions of three-word sentences (e.g., ga ti ti, li na na, ni gi gi), all of which were generated from a simple ABB finite-state grammar. Taking a similar approach to Gómez and Gerken's (1999) study, none of the stimuli that were presented during the familiarization phase were heard at test; half of the test phase sequences were consistent with the ABB grammar (e.g., wo fe fe), and the other half violated the grammar (e.g., wo few wo; illegal strings conformed to an ABA grammar). Even when the stimuli were carefully selected to avoid an overlap in phonetic features between the familiarization and test phase sequences (e.g., grammar-consistent strings did not share the voiced-unvoiced-voiced pattern of the familiarization strings; in Experiment 2), infants oriented longer to the inconsistent items at test. Marcus et al. (1999) suggested that their demonstration should not be taken to deny the importance of statistical computations for the purpose of language learning, but instead, to suggest that abstract rules can be also acquired in infancy.

Increasing evidence over the last years suggests that demonstrating pure rule learning is not as straightforward as originally thought (see also section 1.1.3 for a similar conclusion from studies of implicit learning by means of the artificial grammar learning task). As Seidenberg, MacDonald, and Saffran (2002)—among several other researchers—propose, establishing boundaries between rules and statistical cues can be extremely difficult, especially with material as simple as that used by Marcus et al. (1999). For example, while the transitional probabilities underlying syllable order in Marcus et al.'s (1999) familiarization stream could not have be used per se to discriminate between novel conforming/nonconforming stimuli, an alternative more statistical interpretation of the type "ga ti ti and li na na created contingencies of the type *different-same-same*" is hard to disprove (Seidenberg & Elman, 1999). The extent to which generalizations to untrained items and sensitivity to stimulus-specific statistical patterns arise from two qualitatively distinct learning mechanisms, or alternatively, whether these are qualitatively different outcomes of the same learning mechanism is an important topic of ongoing research (Aslin & Newport, 2012).

1.1.5 Further Research and Concluding Remarks

The literature reviewed in the previous sections suggests that statistical learning is a powerful mechanism that may operate from as early as 2 months of age (e.g., Kirkham et al., 2002), following exposure as brief as 2 minutes (e.g., Saffran et al., 1996a); importantly, statistical computations may underlie performance in a variety of language-related and unrelated tasks (e.g., Saffran et al., 1996a, 1996b; Fiser & Aslin, 2001, 2002; Gómez & Gerken, 1999). Infants' and adults' sensitivity to probabilistic information is not just a laboratory-induced artifact: Statistical learners are able to extract artificial (i.e., nonsense) words from impoverished streams of synthesized speech (e.g., paboki, tibudo vs. tudaro; Aslin et al., 1998), as well as real (foreign) words from naturally produced linguistic sentences (e.g., fuga, melo, vs. casa; Pelucchi, Hay, & Saffran, 2009). There are some limitations, of course, to what can be learnt under highly controlled and minimal exposure experimental conditions. For example, while 5.5- and 8-month-olds are able to discriminate between similar-length (e.g., disyllabic) words and part words at test, this ability has been shown to be disrupted when the speech stream is created by concatenating words of varying (2-syllable and 3-syllable) length (Johnson & Tyler, 2010).

Furthermore, learners are not "drowned in a sea" of possible statistical computations, among which, only a small minority offers viable statistically-based solutions to language-related or more general learning problems. For example, cognitive and perceptual factors (e.g., Creel et al. 2004; Newport & Aslin, 2004), set constraints on what is learnable and how easily they (i.e., structures) are learned. As a final aside, it is worth noting that despite some legitimate concerns over the longevity of what can be learnt with surprising ease in the laboratory (e.g., Taylor & Houghton, 2005), recent evidence suggests that experimentally induced statistical learning is more robust over time than anticipated (Kim, Seitz, Feenstra, & Sham, 2009; Frank, Tenenbaum, & Gibson, 2013). For example, Kim et al. (2009) demonstrated that visual statistical learning of temporal relationships not only persists across a 24-hour delay following exposure, but also does not significantly deteriorate relative to when measured immediately post-exposure.

Statistical computations may have been implicated in wide range of language-related activities; yet, this is not to suggest that they can account for all aspects of language acquisition, or even more generally cognitive development (Kirkham et al., 2002). For example, there are other "nonstatistical" cues to word boundaries, such as the predominant stress pattern in one's native language (e.g., English is a trochaic, and as such, most of its words receive word-initial stress; Cutler & Carter, 1987), which are also acquired early on in infancy (e.g., Jusczyk, Cutler, & Redanz, 1993). When they are pitted against statistical cues in laboratory experiments, previous research has sometimes shown that they are preferably favored by infants (Johnson & Jusczyk, 2001; however, see Thiessen & Saffran, 2003). Children's weighting of statistical versus other types of cues is still a matter of some controversy, and is extremely likely to be open to the rapid developmental changes observed during the first year of life. Recent advances in statistical learning research corroborate the importance of considering the joint contribution of multiple (statistical-based and/or other) cues for a given task (Thiessen & Saffran, 2003).

1.2 A Statistical Learning Account of Learning to Spell

It is a logical and straightforward assumption, based on the evidence for a domain and modality general learning mechanism presented earlier, that statistical learning processes may be implicated in children's ability to learn similar structures in the orthographic visual domain. Of course, this would only be true, if written language per se—or the relationship between written language units and oral language units—embedded patterned structural relationships similar to those that seem to be readily acquired by statistical learning processes. In his pioneering work, Venezky (1970) provided evidence for the existence of such patterns and how they can be of use to readers and spellers. It was shown that, for example (see Treiman, 2004, for several other examples), some of the ambiguity associated with spelling the English sound /k/ (*k*: e.g., *k*itchen; *c*: e.g, *c*at; *q: q*uarter*; ch:* e.g., *ch*air; *ck*: e.g, pi*ck*), can be reduced by attending to the positional cue that *ck* never occurs in word beginnings.

In this section, a brief overview of similar probabilistic/deterministic (i.e., all-or-nothing) frequency-based patterns that are known to govern correct spelling is given (see also Chapter 3), and the recent emergence of a theoretical framework consistent with the proposal that children's spelling development may be, to some extent, facilitated by statistical learning processes is documented. Characteristic examples of the few studies that have set out to investigate children's sensitivity to frequency-based patterns in written language are also presented (see also Chapter 3 for a more extensive discussion). Two important clarifications need to be made with regards to the findings from the line of research discussed below. First, unlike the studies carried out in the context of statistical learning research, orthographic frequency-based learning has been, to date, studied by assessing children's sensitivity to the *actual* patterns that occur in written language. In studies of this sort, children are administered nonword completion tasks and/or presented with and asked to choose between more and less orthographically plausible nonwords. There are some well-acknowledged pitfalls in this approach (i.e., assessing sensitivity to existing patterns), such as the difficulty to control for children's prior pattern knowledge (e.g., Gómez & Gerken, 1999). Second, and perhaps more importantly, the statistically governed nature of children's sensitivity (i.e., the extent to which such patterns are actually acquired by statistical learning processes) is not necessarily confirmed within this approach. The patterns may be untaught via some/most spelling curricula, however, it needs to be confirmed that these can be learnt under incidental conditions. This was one of the main aims of the work reported in this thesis

The acquisition of spelling skill. Learning to spell is a complex developmental process (Frith, 1985) and a challenging task for children—perhaps, in particular those faced with the *deep* (Katz & Frost, 1992) or *inconsistent* English orthography. Only a small minority of English words—such as bed—can be spelled accurately by mapping phonemes (i.e., sounds) into their highest frequency graphemes (i.e., letters) (e.g., in English, the sound /ɛ/ is most commonly spelled with the letter e), and, in fact, a great deal of controversy surrounds the spelling of most English vowel sounds (e.g., /ɛ/ can be also spelled as *ay* [e.g., s*ay*s], *ea* [e.g., h*ea*d], *ie* [e.g., fr*ie*nd], *u* [e.g., b*u*ry], *eo* [e.g., l*eo*pard]; Hayes, Kessler, & Treiman, 2005; see Kessler & Treiman, 2001, for a quantification of English vowels' "ambiguity"). It is not until other aspects of linguistic structure are taken into account that some consistency can be brought into the task of spelling (Caravolas, 2004). Morphological relationships among words (e.g., /ɛ/ in health may not be spelt with *e*, but *ea* is somewhat more expected considering that health is a derivative of the verb heal) provide one such cue. More pertinently for the aims of this thesis, several orthographic probabilistic and deterministic "rules" proscribing letters in certain positions (such as the "c*k is never an onset*" example presented above) or relationships between adjacent letters in print (e.g., *dz* is an illegal letter combination in English) comes into play to help spellers choose between several different alternative options. Although such linguistic aids are not always taught as explicit "rules" through the formal curriculum, their contribution to correct spelling performance is well recognized. The question of when children become sensitive to them is on the other hand, more debatable. Are children able to extract them from the very beginnings of learning to spell? Or does this ability emerge "late", for example, once children have reached sufficient levels of understanding of phoneme-to-grapheme correspondences?

Two popular models of literacy development suggest the latter. According to Frith (1985) and Gentry (1982), the emergence of orthographic knowledge *requires* the mastery of some simpler strategies and *heavily relies* on children's extensive exposure to written language: it is not until the very last stage/phase of learning to read and spell that both of these are assumed to hold true. Frith's (1985) model posits that the ability to spell words with conventional accuracy emerges gradually in stage-like manner. While in the first *logographic* stage, children are, essentially, unable to spell. They may produce scribbles and drawings; however, these bear no relationship to spoken sounds, that is, the more beginner spellers lack awareness of phoneme-to-grapheme relationships. It is only when children enter the second *alphabetic* stage that knowledge of letter-sound connections begins to emerge and gradually forms children's predominant spelling strategy: words tend to be (mis)spelled as they sound. Orthographic probabilistic constraints governing which letters tend to go together in written language (and which do not) and probable/improbable phoneme-to-grapheme relationships begin to be appreciated during the final *orthographic* stage, that is, once complete phonological understanding has been reached. Gentry's (1982) five-stage (*precommunicative*; *semiphonetic*; *phonetic*; *transitional*; *conventional*) theoretical model of spelling development makes a similar claim. While some basic orthographic conventions begin to be recognized and exploited during the second to last transitional phase, these are extensively used only when children enter the final conventional stage of spelling development.

Both of these conceptualizations of spelling skill have had a remarkable impact on spelling research over the last 35 years, and it is therefore, not surprising that the large body of evidence generated by statistical learning researchers has not been explicitly taken into account in theories of written language acquisition until recently. Nevertheless, there are now several reviews documenting the emergence of a statistical learning perspective of spelling development (Deacon, Conrad, & Pacton, 2008; Pollo, Treiman, & Kessler, 2007; Treiman & Kessler, 2013; see also Pacton & Deacon, 2008). Pollo et al. (2007) conclude their discussion of two prominent accounts of children's spelling development (the *phonological perspective*; e.g., Frith, 1985;

Gentry, 1982; the *constructivist perspective*; e.g., Ferreiro & Teberosky, 1982), by highlighting the wealth and breadth of information contained even in the earliest (i.e., pre-phonological) of children's spellings. They suggest that these early productions are not random collections of letters—therefore, uninformative—as proponents of the phonological perspective hold; instead, they often reflect beginner spellers' sensitivity to patterns derived from their orthography. How are these patterns learnt? According to Pollo et al. (2007), mere exposure to text (e.g., children's early exposure to books, magazines, even street signs) is sufficient for frequency-based pattern extraction to occur, a proposal paralleling recent advances in computational "metaphors" of learning to spell (see Snowling, Hulme, & Nation, 1997), that is, connectionist simulations of spelling development (e.g., Houghton & Zorzi, 2003). Several examples of beginner spellers' statistical sensitivity are provided by Pollo et al. (2007), among which, children's well-documented propensity of producing words that incorrectly include letters from within their own names (e.g., Treiman, Kessler, & Bourassa, 2001). Pollo et al. (2007) offer a statistically based interpretation of this finding, by suggesting that children's overexposure to spellings of their name may be the cause. In fact, as children begin to get exposed to more representative samples of text, a decrease is observed in their tendency to use letters from their own name and other letters that appear frequently within such material tend to be produced.

Psycholinguistic work by Kessler and Treiman (2001; see also Kessler, 2009) further corroborates the plausibility of a statistical learning account of learning to read and spell. In their extensive analysis of English phoneme-to-grapheme relationships (and vice versa) within monosyllabic words that are familiarly encountered by American college students, they identified several probabilistic patterns that may help resolve some of the notorious inconsistencies in mapping English phonology into orthography. Taking consonantal phonemic context into account while spelling medial vowels in CVC words was shown to significantly improve their consistency, although this was true far more reliably when coda phonemes (i.e., the consonant(s) that followed the vowel) were taken into account, rather than when onset phonemes (i.e., the consonant(s) that preceded the vowel) were taken into account. To illustrate this idea with some examples, while /i/ can be spelled in several different ways in medial positions (among which the most common spelling is *ea*), *ee* spellings were far more probable within monosyllabic words ending with /d/ (e.g., reed) or /p/ (e.g., deep). Along similar lines, *igh* was shown to be a more predictable spelling of the sound /aɪ/ before /t/ (e.g., light), but not before other consonantal phonemes such as /b/ or /d/—which tended to predict i_e spellings (e.g., ride). Similarly, incorporating knowledge of conditional patterns between vowels and onset consonantal phonemes worked well for a few sounds such /ɝ/, which is typically spelled as *or* before /w/ (e.g., work), but not before /k/, where ur, ir, or er are more frequently encountered spellings (e.g, curd). An important note regarding these findings is made by Kessler (2009, p. 22):*"It bears remembering that these conditional consistencies were all derived by mathematically analysing the vocabulary itself: the data do not say that people take advantage of these patterns, but they do prove that the patterns are there. […].The next question is whether humans are sensitive to, learn, and use conditional patterns".* An emerging stream of research addressing this question is reviewed in the following section.

1.2.1 Graphotactic and Phono-graphotactic Pattern Sensitivity

Treiman's (1993) seminal naturalistic work on a corpus of early spelling attempts produced by American English children at the end of kindergarten, first and second grade is one of the first lines of evidence showing that some basic orthographic conventions are "taken into account" even by the most inexperienced spellers (see also Chapter 3). Several more experimental studies have continued and extended her work. There are two key studies by Treiman and colleagues (Treiman, Kessler, & Bick, 2002; Treiman & Kessler, 2006) investigating whether children and adults are sensitive to conditional probabilities governing sound-spelling correspondences (as anticipated on the basis of Kessler and Treiman's [2001] corpus-based analyses). Another study by Hayes, Treiman, and Kessler (2006) has assessed children's use of a contextually conditioned graphotactic—rather than phono-graphotactic—pattern. These, together with evidence for another statistical influence on children's spellings (*orthographic conventions governing double letters*; e.g., Cassar & Treiman, 1997; Pacton, Perruchet, Fayol, & Cleeremans, 2001) are discussed in Chapter 3.

Paralleling the advances in statistical language learning research has been the emergence of a literature exploring children's sensitivity to frequency-based information in conjunction with—or when contrasted with—other (e.g., morphological) cues, rather than in isolation (e.g., Deacon, LeBlanc, & Sabourin, 2011; Pacton, Fayol, & Perruchet, 2005). An extensive coverage of the literature assessing children's sensitivity to morphological rules versus frequency based patterns is beyond the scope of this thesis. However, two findings of relevance to children's orthographic pattern sensitivity are noted here. First, findings from a study conducted with French speaking children (Pacton et al., 2005) suggest that the transcription of ambiguous spellings (e.g., in French, /o/ can be spelled as o, au, eau, ot, aut; /ɛt/ can spelled as ette, ête, aite, ète) are early on informed by probabilistic, context-based graphotactic constraints. In Pacton, Fayol, and Perruchet's (2005) study, from as early as second grade, children were more likely to produce an _ette spelling within pseudowords such as /sorivɛt/ rather than /soritɛt/, a pattern of performance reflecting the French probabilistic graphotactic constraint that /ɛt/ is frequently transcribed as ette after /f/ but very rarely after /v/. Furthermore, they relied on context-based statistical patterns in their transcription of /o/ sounds, and rarely produced words as vita*f*eau that would be graphotactically illegal in French. Along similar lines, Kemp and Bryant (2003) showed that among developing English spellers (5- to 9-year-olds) and adults, /z/-ending plural nonwords such as /stɒgz/ were more likely to be spelled with an _s (rather than *z*, another allowable /z/-plural word spelling) when the previous consonant was voiced (e.g., /g/), a trend paralleling the fact that g and s frequently co-occur in English (e.g., bugs), in contrast to b and z that never occur together in the orthography.

1.3 Implicit Learning Research

The study of "effortless" learning under well-controlled experimental conditions that involve no instruction to learn (*implicit learning*) has been the main focus of another relevant stream of the literature, generated following Arthur Reber's (1967) artificial grammar learning experiments. Despite the agreement on the domain-general nature of implicit learning processes, defining what it means for learning to be implicit has proved a very difficult task. The term was initially coined to describe the process of acquiring knowledge that this not fully available to conscious awareness, the latter being typically measured by recognition tasks or participants' ability to report verbally their basis of performance on a given implicit learning task. A definition along these lines (that is, judging learning to be implicit or not depending on whether the *knowledge product* is available or unavailable to retrospection; Frensch & Rünger, 2003) soon proved inadequate in light of evidence highlighting drawbacks associated with the way conscious knowledge was inferred (Shanks & St. John, 1994). A definition of implicit learning as the process of learning under incidental conditions and that involves little intention to learn, or the use of direct hypothesis testing by the learner appears to be more agreed upon (Perruchet, 2008; Seger, 1994).

The artificical grammar learning task. Among the several tasks that been employed for the study of implicit skill learning (for a thorough review, see Seger, 1994), one of the best-established paradigms is Reber's (1967) prototypical artificial grammar learning task. In a standard version of the experiment, participants are warned of an impending memory task and are presented with a set of seemingly arbitrary stimuli (e.g., letter strings) to be observed, mentally rehearsed, reproduced on a piece of paper (e.g., Brooks & Vokey, 1991; Knowlton & Squire, 1996; Reber, 1967) or on the keyboard (e.g., Kinder, 2000; Kinder & Lotz, 2009). The stimuli are not arbitrary as participants are made to believe; in fact, they all conform to a Marcovian finite-state grammar (such as the one shown in Figure 1.1), and are created by traversing left-to-right (starting at node 1) through one of the possible pathways of the finite state grammar (PTVV is, for example, a permissible letter string according to the grammar shown in Figure 1.1; PTVS is not permissible). Following memorization, participants are made aware of the rule-governed nature of the stimuli (but not the actual rules), and are invited to take part in a task assessing their ability to discriminate between novel strings that either obey or disobey the "rules". Typically, it is further explained to participants that the rules are extremely complex to unravel, therefore, they should use their intuition or "gut feeling" to judge the stimuli's grammaticality. Despite the apparent difficulty of telling apart grammatical from ungrammatical items—given that participants do not explicitly try to learn rules during the training phase—it has been unequivocally demonstrated that their ability to discriminate between grammar-conforming and nonconforming stimuli reliably exceeds chance (i.e., 50% accuracy). Whatever it is that participants learn while memorizing the grammar conforming instances, it is clearly sufficient for judging the grammaticality of the new strings.

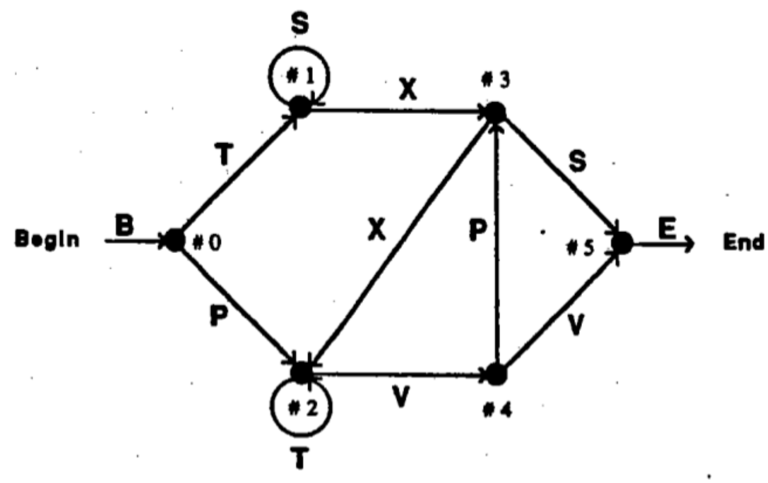


Figure 1.1. Example of a finite state grammar used for stimulus construction. Reproduced from Reber (1967).

This well-established result has been taken further in many different directions over the last decades (for a detailed review of advances in the implicit learning domain see Perruchet, 2008). For example, several studies have assessed (a) the validity of the claim that implicit learning processes are separate from the those involved in humans' ability to learn when instructed to do so (explicit learning), and (b) the extent to which participants' learning is followed by consciousness awareness, as well as how to sensitively measure the latter. The question of the representational form of the knowledge acquired in implicit learning tasks has generated another extensive body of research and, for the needs of this thesis, is reviewed in the following section. In fact, all of the above questions are largely interconnected in the implicit learning literature. For example, can participants learn abstract rule information/frequency-based string elements, and if so, is such knowledge available to conscious verbalization? [2].

---

[2]There are many well-acknowledged methodological problems related to the assessment of unconscious cognition (see Cleeremans, Destrebecqz, & Boyer, 1998; Perruchet, 2008; and Shanks & St. John, 1994, for relevant discussions), and therefore, the latter question was not addressed in the body of work reported in this thesis. This is a similar approach to taken by many statistical learning researchers. For example, according to Saffran et al. (1997), their primary interest was the laboratory study of learning under conditions of passive exposure (i.e., similar to those underlying natural language pattern extraction). Although the extent to which the product of such learning was available to conscious verbalization was not systematically assessed in their experiments, it was acknowledged that their adult participants may have varied largely on that dimension (Saffran et al., 1997).

1.3.1 Theories of Artificial Grammar Learning

The abstractionist account. While Reber's (1967) finding of above chance learning is well-accepted and well-replicated, his original claim (e.g., Reber, 1969, 1989) that the ability to discriminate between grammatical and ungrammatical strings relies on abstract rule knowledge was met with considerable skepticism. Transfer studies along the lines of Gómez and Gerken's (1999) study described earlier in the chapter (see section 1.1.5) were used by Reber himself (1969) as well other researchers (e.g., Manza & Reber, 1997; Mathews et al., 1989; Reber, 1969) and were taken to provide support for the abstractionist view. Reber (1969) used a standard training phase during which participants were presented a total of 18 grammatical items for a fixed amount of time (these were presented within 6 groups of 3 items each) and were asked to reproduce them on a piece of paper. Following correct generation of all items, a test phase was introduced, during which participants were asked once again to memorize an additional number of strings; however, they were allocated to one of the following four experimental conditions: (a) a *symbol transfer* condition whereby test phase items were generated from the same grammar that was used to construct the training items, but critically, the former were expressed in a different surface form (i.e., a different set of letters; R was mapped onto P etc.), (b) a *syntactic transfer* condition whereby the same set of letters but a different grammar was used at test, (c) a *different syntax-different symbols* condition whereby both grammar and surface form differed, and finally (d) a *null transfer* condition, whereby the same grammar that was used during training was also used at test. The critical demonstration taken to provide evidence of abstract rule learning was that a memorization advantage for grammar-conforming rather than non-conforming items was shown not only in the null transfer condition (i.e., same syntax and same surface form), but also in the symbol transfer condition. Reber (1969) concluded that participants are learning more than "*simply string together explicit symbols*" (p.119). Several more studies have demonstrated transfer to changed-letter items (e.g., Mathews et al., 1989; Gómez & Schvaneveldt, 1994), and more impressively, cross modality transfer has been also observed (e.g., Altmann, Dienes, & Good, 1995; Manza & Reber, 1997). For example, Altmann et al. (1995) demonstrated that participants exposed to spoken syllable strings were able to discriminate between grammatical and ungrammatical sequences constructed by mapping syllables onto unpronounceable visual shapes, even these were never explicitly taught to participants.

Although an extensive discussion of the extent to which transfer effects truly demonstrate abstract rule learning is beyond the scope of this review, two of the several objections that have been raised against this claim are noted: First, performance in transfer experiments is usually closer to the 50% (i.e., chance) level (mean accuracy usually ranges between .55 and .60), a finding raising the possibility that learning may be incorrectly concluded to be reliable; this marginally better than 50% discrimination ability may not be statistically different from that of untrained control participants. Although untrained, this latter group of participants may have achieved similar above chance levels of performance on the basis of some simple knowledge acquired at test, such as that all grammatical strings start with a few possible letters (Redington & Chater, 1996). Second, even when transfer is reliable, it may not necessitate an interpretation along the lines of complex abstract rule learning, but in fact, may reflect learning of quite simple local repetition structures (i.e., repetitions between adjacent string elements), as has been shown by previous studies (e.g., Lotz, Kinder, & Lachnit, 2009). A number of studies proposing partially or fully alternative explanations to Reber's claim that classification performance entails forming abstract rule representations are discussed below.

The (whole) instance-based account. Brooks and Vokey (1991; see also Brooks, 1978; McAndrews & Moscovitch, 1985) were among the first to propose and provide evidence for a partial nonrule-based account of artificial grammar learning performance. They argued against Reber's abstractionist theory by suggesting that the use of an analogical strategy—drawing on the abstract-level *similarity* between test phase strings and training items, the latter being memorized as wholes—could be sufficient for participants to transfer knowledge from one to another letter set. For example, given exposure to BDCCCB, participants may have been able to infer the grammaticality of MXVVVM on the basis that they both began and ended with the same letter and that they both included a repetitive structure in positions 3-5. To explore this hypothesis, Brooks and colleagues devised an edit measure, capturing the number of letters by which a test phase string differed from its most similar training item. Test phase items that would differ from their closest training item by one letter would be classified as similar ("close" items), whereas those that would differ from their closest training item by two or more letters would be classified as dissimilar ("far" items). On the basis of this categorization, four types of test phase material were created in Brooks and Vokey's (1991) study pitting the factor of grammaticality against the factor of specific similarity: there was a quarter of items that were both grammatical and similar to training items, and a quarter of items that were grammatical but dissimilar to training items; the remaining two quarters were either ungrammatical items that were similar or dissimilar to training items. Brooks and Vokey's (1991; see also Vokey & Brooks, 1992) predictions proved partially correct, in that adult participants showed a preference for similar rather than dissimilar items, regardless of their grammatical status. However notably, against their expectations (see Vokey & Brooks, 1994), an additional effect of grammaticality was observed, such that participants' rate of endorsement of grammatical strings was significantly higher than their endorsement of ungrammatical strings (for a replication, see Knowlton & Squire, 1994: Experiment 2a). Given that there was little evidence to disprove that grammaticality was another significant contributor to participants' performance, it was concluded that both factors may be independent determinants of participants' classification performance (Brooks & Vokey, 1991).

Fragment-based accounts. A critical assumption underlying the validity of Brooks and colleagues' demonstration that specific similarity is a determinant of participants' discrimination performance is that any "knowledge variables" that correlate with the construct's presence have been carefully removed. In a report documenting a re-analysis of Brooks and Vokey (1991; Vokey & Brooks, 1992) material, Perruchet (1994; see also Knowton & Squires, 1994) proved that this had not been appropriately ensured. When training and test phase strings were broken into their constituent initial/terminal trigrams (i.e., three-letter fragments or chunks), a considerably higher amount of training/test phase string overlap was observed among the items that were labeled as similar rather than those that were labeled as dissimilar. Furthermore, grammatical strings rather than ungrammatical strings shared more initial/final trigrams with the training set. Along the lines of this evidence, Perruchet (1994) argued that a plausible basis of participants' higher endorsement of grammatical over ungrammatical and similar over dissimilar items was their frequency-based sensitivity to these elementary fragments.

Further compelling evidence for the contribution of frequency-based bigram knowledge to participants' judgments of well-formedness was provided by Perruchet and Pacteau (1990; Experiment 1). They assessed participants' ability to learn under a "letter pair" condition which precluded the formation of abstract grammatical knowledge: That is, participants were exposed to a corpus of bigrams that would be, under normal circumstances, the constituent of the training strings. Bigram-trained participants' performance was compared against that of participants trained under standard presentation condition (exposure to the corpus of entire training strings). Both groups' judgments of grammaticality were significantly above chance, supporting Perruchet and Pacteau's (1990) view that fragment-based knowledge mediates artificial grammar learning performance. However, it should be noted that participants trained with letter pairs were slightly less accurate than participants trained with entire strings, partially due to the latter groups' ease of rejecting salient ungrammatical strings embedding an illegal starting letter. These were taken to constitute another source of knowledge contributing to performance.

A final well-documented line of evidence for fragment-based learning (and against the instance-based account of artificial grammar learning) is Knowlton and Squire's (1994; Experiment 2b) study pitting the specific similarity of strings, quantified along the same lines of Brooks and collaborators' reasoning, against the strings' *anchor and global associative chunk strength*, two metrics of the amount of bigram/trigram overlap between a specific test string and the entire corpus of training strings. The latter was calculated by summing and averaging, for *each* test phase item, how frequently each of its constituent elements occurred across the set of training items (see Chapter 4 for a more detailed description and specific examples of high vs. low associative chunk strength items). It was shown that when similar (i.e., only one letter different) and dissimilar (i.e., more than one letter different) items were balanced, that is, did no differ in terms of associative chunk strength, similarity no longer exerted an influence on adult participants' (typical and amnesic) classifications. On other hand, performance was substantially determined by test phase item differences in terms of Knowlton and Squire's (1994) associative chunk strength measures. High associative chunk strength items were endorsed significantly more than low associative chunk strength, a finding providing support for the idea that training/test phase item chunk overlap plays a significant role in grammaticality judgments (see also Johnstone & Shanks [2001; Experiment 2] for an interesting demonstration, using stimuli from a different type of grammar, of a “reversed” grammaticality effect when the set of ungrammatical, rather than grammatical items, was made to share more chunks with the training items).

Are grammar rules acquired over and above chunks? What is unclear—on the basis of the evidence reviewed up to this point—is whether sensitivity to bigram/trigram frequencies can account for all of participants’ classification performance. In other words, are fragment-based accounts a full alternative to the rule-based interpretations of artificial grammar learning performance? Knowlton and Squire’s (1996) manipulation of stimuli’s (a) adherence to “rules” and (b) chunk strength information in an orthogonal manner are of direct relevance to this question. In Experiment 1, they demonstrated that both factors had an effect on typical as well as amnesic participants’ judgments of grammaticality, such that (a) grammatical letter strings were endorsed more frequently than ungrammatical letter strings and (b) high associative chunk strength letter strings were endorsed more frequently than low associative chunk strength strings. However, an interaction between chunk strength and grammaticality indicated that the effect of chunk strength was not independent of grammaticality. It was only within the set of ungrammatical items that low chunk strength items were endorsed less frequently than high chunk strength items. On the basis of the grammaticality effect observed in Experiment 1 and additional evidence of reliable discrimination performance between grammatical and ungrammatical items in a letter transfer condition (Experiment 3), Knowlton and Squire (1996) suggested that participants may learn more than chunks in the artificial grammar learning task. Rules, although not necessarily veridical representations of grammar as Reber (1969) suggested, were also learnt by amnesic and control participants.

Further evidence for a dual learning mechanism interpretation of artificial grammar learning performance was obtained by Meulemans and Van der Linden (1997). First, they performed a re-analysis of Knowlton and Squire’s (1996) grammatical/ungrammatical items in light of another chunk-related variable (chunk novelty; the extent to which a test phase sequence contains bigrams/trigrams that have never appeared during string memorization) and revealed that a confound between grammaticality and chunk novelty could account, to some extent, for the grammaticality effect demonstrated by Knowlton and Squire (1996). Second, controlling for chunk novelty item differences within their material, they pitted chunk strength and grammaticality against each other under conditions contrasting the *range* of items used for memorization. In Experiment 2b, the memorization set covered most of the items that could be generated from the finite-state grammar (n = 125), whereas in Experiment 2a, participants were presented with only 32 of the strings that could be generated from the grammar. Interestingly, a differential contribution of frequency-based fragment and rule knowledge was obtained as a function of training set range. The effect of grammaticality (but not associative chunk strength) was significant when participants were presented with the more representative set of training strings, whereas, the opposite pattern (significant effect of associative chunk strength, but no effect of grammaticality; or grammaticality by chunk strength interaction, cf. Knowton and Squires, 1996) was obtained in the short range condition. Meulemans and Van der Linden (1997) proposed that, while bigram and trigram knowledge can account for participants' performance under exposure to a limited set of grammatical items, exposure to an extensive item range gives rise to abstract rule knowledge.

The list of factors that are frequently confounded with grammaticality seems to be endless, as in fact, even within Meulemans and Van der Linder's (1997) design, one source of knowledge was shown to co vary with changes in the grammatical status of strings (Johnstone & Shanks, 1999). That is, although both grammatical and ungrammatical test phase items contained to the same extent bigrams and trigrams that appeared during the training phase (i.e., they were equated in terms of global/anchor associative chunk strength and chunk novelty), in Experiment 2a (whereby a significant main effect of grammaticality was obtained), bigrams and trigrams appeared more frequently in *new* positions among ungrammatical items. A multiple regression analysis—following Lorch and Myers (1990) individual regression equation method—using the different chunk strength related measures, grammaticality and novel correct position as independent predictors, revealed that grammaticality was not a significant contributor to performance in Meulemans and Van der Linder's (1997) design (Johnstone & Shanks, 1999). As Johnstone and Shanks (1999) conclude, there appear to be serious difficulties in disentangling between rule-based and fragment (frequency- or statistically based) accounts of learning using stimuli generated from transitional finite-state grammars. This is not to suggest that the question of whether abstract rule learning occurs over and above sensitivity to statistics is of little theoretical importance (see also section 1.1.4). However, it may be better addressed by means of alternative grammars that provide a more precise quantification of rules independent of fragment knowledge (for an example of such a system, see Johnstone & Shanks, 2001).

Summary and conclusions. To sum up, the nature of the representations underlying artificial grammar learning performance has been a source of considerable debate, and several theories have been put forward to explain participants' better-than-chance classification accuracy. A full account of artificial grammar learning performance remains elusive. By one view, artificial grammar learners base their grammaticality judgments on an abstract representation of the rules used to construct the training stimuli, whereas, according to the instance-based theories, their performance may be further guided by the similarity between the stimuli encountered at test and individual training stimuli. Fragment-based theory proponents suggest that sensitivity to bigram and trigram information provides a more accurate account of participants' performance than sensitivity to whole training exemplars or rules, even though the presence of additional explained variance in participants' performance indicates that other string aspects may also be encoded and influence classification responses. Of necessity, the studies described above are only a selection of those investigating the basis of participants' classification responses in the artificial grammar learning task. One "omission" is the episodic-processing account of artificial grammar learning performance highlighting the processing-driven nature of implicit knowledge acquisition and suggesting that there no defaults in the representational form of knowledge acquired (e.g., Whittlesea & Dorken, 1993; Whittlesea & Wright, 1997, Wright & Whittlesea, 1998). Whittlesea and colleagues' theoretical framework and empirical evidence in support of their claims are covered in Chapter 5.

Despite the disagreement about the mental representations underlying learning performance in implicit tasks such as the artificial grammar learning task, it is ubiquitously agreed upon that implicit cognitive processes are important for a wide range of behavior and several domains of knowledge, including the acquisition of literacy skill (e.g., Pacton et al., 2001; Steffler, 2001). This proposal has been recently taken further by another body of literature that postulates implicit learning deficits in situations where unexpected reading and spelling difficulties emerge. Evidence for the presence of implicit learning impairments in developmental dyslexia and their relationship to the main behavioral difficulties that are commonly documented in the context of this disorder are presented in the next section.

1.3.2 Implicit Skill Learning in Dyslexia

Dyslexia is a neurodevelopmental disorder (Habib, 2000) characterized by inaccurate and/or dysfluent word reading and spelling (Rose, 2009). It is a hereditary disorder (see Grigorenko, 2001 for a review of twin, genetic and familial aggregation studies elucidating the role of genes in the etiology of dyslexia) that is highly prevalent among school-aged children (Fletcher, 2009, note that prevalence estimates vary considerably between reports due to different inclusionary criteria). Literacy difficulties may gradually improve as dyslexic children get older, but frequently persist into adulthood, and therefore are not the result of a temporary developmental lag (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Shaywitz et al., 1999).

Although several theories have been put forward to explain the causal factors behind this disorder, there is now increasing consensus that literacy impairment stems from a deficit in the phonological component of language which affects dyslexic children's ability to develop well-specified phonological representations (for reviews see, Bishop & Snowling, 2004; Snowling, 2000). It has been frequently shown that dyslexic children's ability to represent, store, and/or manipulate the speech sounds (phonemes) of words is seriously impaired relative to that of their age-matched peers (*phonemic awareness* deficits: e.g., Bradley & Bryant, 1978; Caravolas, Volín, & Hulme, 2005; Ramus et al., 2003). Marked impairments on other measures that draw heavily on phonological processing (e.g., verbal short-term memory, nonword repetition and rapid automatized naming tasks) have been also frequently observed among dyslexic populations (e.g., Pennington, Van Orden, Smith, Green, & Haith, 1990; Rack, Snowling, & Olson, 1992; Ramus et al., 2003).

The case for a causal role of impaired phonological processing in dyslexia (Bradley & Bryant, 1983; Ramus et al., 2003; Snowling, 2000; Stanovich, 1988) is supported by at least two complementary lines of evidence. The first is longitudinal, and builds on the now well-established finding children's phonemic awareness abilities—measured *before* literacy skills begin to emerge—are one of the most reliable longitudinal predictors of later reading and spelling skill in typically developing children (e.g., Caravolas, Hulme, & Snowling, 2001; Caravolas et al., 2012; Muter, Hulme, Snowling, & Stevenson, 2004). Second, carefully designed intervention studies have shown that phonological awareness training—combined with explicit instruction on the links between sounds and letters—results in significant gains in reading and spelling among dyslexic children (e.g., Bowyer-Crane et al., 2008; Hatcher, Hulme, & Ellis, 1994; Torgensen et al., 1999)

The presence of phonological deficits in dyslexia is at this point beyond doubt and seems largely undisputed by most theoretical accounts of dyslexia. However, extensive research has documented that dyslexic impairments are hardly ever limited to the domain of literacy. A wide range of additional cognitive deficits have been frequently observed and may include impairments in (a) low-level visual processing skill (e.g., Lovegrove, Bowling, Badcock, & Blackwood, 1980; for a review see Stein & Walsh, 1997), (b) auditory temporal order perception and/or the ability to process rapidly changing verbal acoustic elements (e.g., Tallal, 1980; for a review see: Farmer & Klein, 1995), (c) visual attentional skills (e.g., Valdois et al., 2003; for a review see: Valdois, Bosse, & Tainturier, 2004), and (d) motor skills leading to clumsiness, poor coordination and handwriting (Nicolson & Fawcett, 1990; for a review see: Nicolson, Fawcett, & Dean, 2001). However, researchers are divided on their interpretations of the causal roles of these additional deficits; they are viewed by some as direct or indirect (i.e., affecting phonological processing skill) causes of reading and spelling problems, and by others as the consequences of poorly developed literacy skills. For example, are children's auditory difficulties in the perception of speed sounds a cause or an effect of phonological processing deficits (Hulme & Snowling, 2009; Ramus, 2003)? Another important consideration concerns the extent to which the presence of co-occurring deficits may be related to the well-known fact that co morbidity between dyslexia and other developmental conditions (for example, attention deficit hyperactivity disorder, developmental coordination disorder and dyscalculia) is the norm rather than the exception (Kaplan, Dewey, Crawford, & Wilson, 2001; Landerl & Moll, 2010; Pennington, Willcutt, & Rhee, 2005).

The view that the broad heterogeneity of dyslexic symptoms is not to be ignored is core to the automaticity deficit hypothesis of dyslexia discussed below—a framework raising the possibility that dyslexia may be better explained along the lines of a general and pervasive impairment in skill acquisition processes (e.g., Nicolson & Fawcett, 1990, 1994, 1995; Nicolson et al., 2001). Two claims and one clear prediction are central to Nicolson and Fawcett's proposal. First, impaired skill automatization may be able, at least partially, to account for the well-documented reading aloud and spelling deficits in dyslexia: Both of these skills become fully automatic and require little of skilled learners' conscious control (e.g., Ehri, 1976). Second, impaired automatization may be *also* account for deficits in the acquisition of other skills (cognitive or not) that are unexplained by the presence of a phonological processing impairment. The prediction is that, dyslexic children will be able to conceal some of their automatization difficulties (after all, their performance is not impaired in absolutely every domain, as Nicolson and Fawcett [1990] acknowledge) whenever allowed to resort to conscious compensation strategies. On the other hand, when prevented from doing so, automatization deficits become particularly prominent. While this preliminary framework was not intended to provide a causal explanation of literacy failure in dyslexia, a more recent version of Nicolson and Fawcett's (1990) proposal draws causal links between a dysfunction in the cerebellum—a brain structure related to automatization and motor control—and speech articulation, leading in turn to poorly specified phonological representations (Nicolson et al., 2001). Consistent with this proposal is some neuroimaging evidence indicating abnormal anatomical and metabolic cerebellar function in dyslexic adults (e.g., Nicolson et al., 1999; Rae et al., 1998).

The automaticity deficit hypothesis. Incomplete automatization in dyslexics was first studied within the context of a skill that bears no relation to written or oral language and is, undoubtedly, highly over learnt: the gross motor skill of balance. In the first of a series of papers, Nicolson and Fawcett (1990) documented a significant balancing deficit among 13-year-old dyslexic adolescents relative to controls when a beaming walking (among other) task was performed in parallel to a counting backwards task (experiment 1), or a high versus low pitch discrimination task (experiment 2). On the other hand, under a single task "just balancing" condition—allowing, presumably, participants to compensate for any balancing difficulties—dyslexic adolescents had little trouble with any of tasks used to assess gross motor skill and performed at control levels. In follow-up studies with dyslexic children and adolescents, similar signs of a marked motor skill impairment were observed for performance on a “bead threading” task and a “peg moving” task (Nicolson & Fawcett, 1994), as well as when dyslexic participants were asked to balance while blindfolded (Fawcett & Nicolson, 1992; Nicolson & Fawcett, 1994). On the basis of these surprising but consistent findings, Nicolson and Fawcett (1990, 1994) criticized single-deficit accounts, such the phonological deficit theory of dyslexia, for being overly restrictive and able to predict only a few deficits: those that are directly reliant on the proposed single underlying cause of dyslexia; they argued that such accounts lack parsimony, in that they fail to account for deficits that are less consistent with the underlying cause they propose. In their words, “*it is hard to envisage how any of the reading specific* [i.e., phonological] *accounts of dyslexia could cope with these results*” (Nicolson & Fawcett, 1990, p. 175).

While the phonological deficit theory may indeed have a hard time explaining such findings, there are several points of caution to be noted with regards to the claim of impaired skill acquisition and automatization. First, several studies have failed to observe motor skill deficits among dyslexic children (e.g., Irannejad & Savage, 2012; Wimmer, Mayringer, & Landerl, 1998), and perhaps more importantly, alternative interpretations have been suggested to account for Nicolson, Fawcett and collaborators’ findings (Rochelle & Talcott, 2006; Ramus et al., 2003). Recently, Rochelle and Talcott (2006) conducted a meta-analysis of 17 studies published between 1985 and 2004 comparing balance function (in particular, postural instability) between dyslexic and control groups. The study revealed a high degree of heterogeneity in the significance of the standardized postural stability impairment across studies. Critically, it was shown that several studies had not screen for co-occurring disorders such as attention deficit hyperactive disorder—which is known to *also* co-occur with developmental coordination disorder (e.g., Kaplan et al., 2001), and may have included garden-variety poor readers in their dyslexic samples (literacy impaired individuals with below average IQ). When such variables were entered as moderators in the analyses, significant associations were observed, such that, for example, weaker postural stability deficits were observed within studies that had controlled for the presence of co-occurring attention deficit hyperactivity disorder. Stronger differences between dyslexic and controls' reading performance were not associated with more impaired postural stability.

Findings such as those reported by Rochelle and Talcott (2006) cast serious doubts on the claim that impaired automatization—affecting processes as basic as postural stability—is specifically related to reading impairment in dyslexia. However, emerging evidence from another line of research suggests that Nicolson's and Fawcett's claims should perhaps not be dismissed outright. Building on the prediction that learning deficits occur when dyslexic individuals are prevented from engaging all of their effort into compensating for their difficulties, the methods of implicit learning research have been seen by some researchers as a new window into the study of automatized learning in dyslexia. Impaired skill learning should be observed under incidental conditions, that is when participants are unaware of the fact that they are learning, and therefore, cannot mask their difficulties by using compensatory strategies (Kelly, Griffiths, & Frith, 2002; Roodenrys & Dunn, 2008).

1.3.3 Implicit Sequence Learning in Dyslexia

A recent and growing body of research has employed standard implicit learning tasks such as the artificial grammar learning and the serial reaction time task (a measure discussed below) to assess skill learning in dyslexia. Learning of artificial grammars occurs incidentally and without participants knowing that they are learning. It is only at the very last stage of the experiment that participants are informed about the rule governed nature of the stimuli. Even more so, in the serial reaction time task, learning is inferred indirectly (i.e., by means of RT differences between patterned and nonpatterned stimuli) and participants need never to be informed of the true purpose of the experiment. Partly due to this reason, the latter has been, to date, the most commonly used measure of implicit learning among dyslexic individuals. Studies using the former paradigm are considered in detail in Chapter 4, whereby our own experimental work sheds additional light on the hypothesized implicit skill learning deficit in dyslexia. In this section, we briefly describe the serial reaction time task procedure and discuss the mixed evidence of spared/impaired implicit sequence learning among dyslexic children and adults.

Evidence from the serial reaction time task. A frequently used alternative to the artificial grammar learning task is Nissen and Bullemer's (1987) visuospatial choice task assessing implicit sensitivity to sequential structure. In the serial reaction time task, participants are presented with a stimulus target (e.g., an asterisk) that appears in several possible locations of the screen and are asked to press a corresponding key in response to its location as fast and as accurately as possible. Unknown to them, the sequence of button presses follows a fixed predetermined order (e.g., D-B-C-A-C-B-D-C-B-A, where A, B, C, and D correspond to one out of four possible locations on the screen) or is governed by contingencies specifying possible transitions between the stimuli (i.e., probabilistic sequences; for examples, see Curran, 1997). Following several hundreds of "training" trials, during which response times tend to decrease as a result of non-specific practice (Kelly et al., 2002), participants are presented with one or more "test blocks" of predictable versus randomly generated sequences[3]. RT increments in response to the randomly generated sequences tend to occur and they are taken to indicate that participants have learnt the sequence element contingencies, even though they are usually unable to verbally recall in full what is the sequence practiced during the task.

Howard, Howard, Japikse, and Eden, (2006) used an alternating serial reaction time task embedding nonadjacent relationships among element locations to compare the performance of 11 dyslexic college students against that of 12 typical age-matched controls. During the task, a colored circle appeared in one out of four possible (horizontally distributed) locations and participants were asked to respond to its presence by pressing one out of four buttons (one for each location). Correct responses were not completely random. The circle's location every *two* trials varied on the basis of an 8-element long sequence, such that, for example, if the colored circle appeared in the leftmost position during trial 2, it would appear in the second to last position during trial 4, regardless of where it appeared during trial 3 (in fact, during trial 3, it could appear in any of the four locations with equal likelihood). Given that predictable and unpredictable sequence locations occurred throughout the entire experiment (8 blocks of trials), learning was also assessed throughout the task by comparing RTs for the random and probable circle locations. It was shown that (a) dyslexic adults were overall slower (but and not less

[3]It is noteworthy that the transition from the training to test phase of the experiment occurs without participants being informed.

accurate) than controls, (b) both groups' RTs decreased from the first to the last block, and (c) critically both groups' RTs were significantly faster for probable than improbable trials. Although some learning occurred among dyslexic participants (i.e., their sequence skill learning was not fully compromised), a significant interaction revealed that the amount of learning was significantly less for dyslexic than skilled readers. On the basis of these group differences, Howard et al. (2006) concluded that higher-order sequence learning is impaired in dyslexia. However, dyslexics' performance on another implicit learning task, Chun and Jiang's (1998) spatial contextual cueing task was similar to that of controls. On the basis of this double dissociation (impaired sequence learning, spared contextual cueing learning), Howard et al. (2006) ruled out the possibility of an overall skill learning deficit in dyslexia.

Against the assumption that dyslexic adults' ability to learn simpler sequential structure may be spared, an significant impairment was found among dyslexic adults by Stoodley, Harrison, and Stein (2006) in a study using a brief deterministic (fixed 10-element sequence) sequence learning variant. However, it was acknowledged that differences at the group level may have concealed evidence of spared learning among some dyslexic individuals. An fMRI study (Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006; see also Menghini et al., 2008 for a voxel-based morphometry investigation of grey matter volume abnormalities in implicit-learning-related brain regions in dyslexia) further replicated the behavioral finding of impaired sequence learning in dyslexic adults (a 9-element fixed sequence was used) and reported different patterns of activation throughout their task performance, and in particular, during the final stages of learning. For example, the left Supplementary Motor Area was significantly less activated in dyslexic adults, suggesting on the basis of evidence from previous neuroimaging studies (e.g., Hazeltine, Grafton, & Ivry, 1997) their difficulties in creating an internal representation of the sequence. Furthermore, unexpectedly high cerebellar activation was observed during the final stages of the experiment in dyslexic individuals, but not controls, signaling lacking of automatization, according to previous research such as the neuroimaging studies of Nicolson et al. (1999).

Demonstrations of impaired implicit sequence learning are by no means limited to adulthood (e.g., Gabay, Schiff, & Vakil, 2012; Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011; Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003; Vicari et al., 2005; Stoodley, Ray, Jack, & Stein, 2008). For example, Vicari et al. (2003) compared 10-year-old dyslexic children's and age- and IQ-matched controls' ability to learn a fixed 5-item-long sequence governing the serial order in which colored circles were flashing on the screen. Throughout the task, children were simply instructed to respond to the presence of one of three possible colors (e.g., green) and speed as well as accuracy of responses was recorded. An examination of children's RTs in the first block of the task revealed no reliable group differences. Dyslexic children were not slower performers, a finding that was taken to suggest that a pure motor impairment among dyslexic children would not suffice to explain any potential group differences. These were indeed observed. First, dyslexic children's RTs tended to increase while "passing" from the first to the last block, while the opposite held true for the typically developing children. More critically, the difference in RTs between the (second to last) sequence governed block and the (last) random block was significant and in the predicted direction (i.e., random block RTs were slower) only for children in the control group. While dyslexic children showed no evidence of implicit sequence learning, under *explicit* task instructions (children were asked to memorize and verbalize the 5-long colour sequence prior to the training period), no difference was found between dyslexic and typically developing children's performance: Both groups' RTs were significantly facilitated by the presence of a predictable sequence.

While all of the above studies have shown that dyslexic individuals' performance on the serial reaction time task is impaired, it is noteworthy that some studies have failed to replicate this finding. Kelly et al. (2002) were the first to document spared learning among a group of dyslexic adults confronted with a double deterministic sequence learning task: one of the sequences governed the allowable location of stimulus elements in a 2 x 2 grid, whereas the other one governed the identity of the sequence "objects" (one out of 4 "alien" figures). Relative to controls, dyslexic individuals' reaction times were overall slower in the task (replicating, for example, Howard et al., 2006); however, they were significantly faster in the pre final sequenced blocks in comparison to a "random spatial location" block (e.g., block 9 in one counterbalanced condition), a "random object sequence" blocks (e.g., block 12) and a random spatial location/random object sequence block (block 15). Not only was dyslexic adults' learning reliable, but it was also of similar magnitude to that observed among skilled readers. Therefore, there was no evidence to suggest that dyslexic adults are impaired in implicit sequence learning. Although some care was taken to prevent participants from becoming explicitly aware of the object and spatial location sequences, the findings of Kelly et al. (2002) have been criticized on the basis that, at least skilled adults' performance may have not been fully implicit (Stoodley et al., 2006). As further noted by Howard et al. (2006), while deterministic serial order learning may be spared among dyslexic adults, the simplicity of the sequence may have concealed the presence of deficit in higher-order learning. Using the serial reaction time task (among other tasks tapping on different domains of neurocognitive difficulty in dyslexia), Menghini et al. (2010) failed to replicate their previous findings of impaired sequence learning (e.g. Menghini et al., 2006) in a group of 11.5-year-old dyslexic children. It was suggested that implicit learning may be impaired among some, but not all dyslexic individuals. Intact sequence learning was also observed in another small scale study with dyslexic adults by Rüsseler, Gerth, and Münte (2006; see also Chapter 4 for evidence of spared artificial grammar learning among the same group of participants). Using a simpler and less motor-processing-reliant task (a simple cue-target pairing reaction time task; participants were asked to provide a motor response only when a specific target appeared), Roodenrys and Dunn (2008) showed that 8- to 10-year old dyslexic children were, once again, slower, but not significantly less advantaged by the presence of cued trials than age-matched controls. Sequence learning was also found to be spared among dyslexic adults in a study employing a "high versus low triplet frequency learning" task attenuating the influence of deficits in the execution of *sequential motor movements* (Bennett, Romano, Howard, & Howard, 2008). When participants were required to respond to the location of only one type of stimuli (indicate the location of green but not red circles; however, the red stimulus locations were probabilistically [remotely] predictive of where the green target would appear), no reliable differences were detected in skilled versus dyslexic adults' performance.

In sum, there is some—albeit partially conflicting—evidence of impaired sequence learning in children and adults with developmental dyslexia. This has been taken to suggest that one aspect of implicit learning, that underlying the ability to learn sequential patterned relationships, may be impaired and contributes to the wide range of difficulties that are typically associated with dyslexia. This is a relatively new stream of research and the finding of reduced or fully compromised implicit sequence learning among dyslexics is open to more systematic replications. The questions of whether learning impairments affect the implicit acquisition of simpler as well as more complex material is also open to further research aiming to control for large methodological differences between the studies reported above. The possibility that nonmotoric implicit skill learning may be spared in dyslexia has been directly considered by some researchers (e.g., Bennett et al., 2008) and indirectly implied by others (e.g., Roodenrys & Dunn, 2008; Stoodley et al., 2006), but has not been extensively studied to date. This is an important point. If learning deficits are as pervasive as hypothesized by the automaticity deficit framework (e.g., Nicolson & Fawcett, 1990) and result in less automatized decoding skill or reduced implicit orthographic pattern acquisition in dyslexia, as Sperling, Lu and Manis (2004) posit, learning impairment should be observed in motor as well as nonmotor implicit learning tasks such as the artificial grammar learning task. This question was addressed in this thesis.

1.4 Objective of the Thesis

This chapter provided a general overview of (a) the statistical and implicit learning research and their experimental methods; (b) theories of the representations that support implicit learning performance; and (c) the applicability of findings from both streams to theories of oral and written language acquisition. Some, albeit indirect, evidence for the recent proposal (e.g., Pollo et al., 2007) that spellers extract frequency-based orthographic patterns by means of the same statistical learning mechanisms that have been implicated in oral language acquisition was also presented. To date, there are no studies assessing directly whether orthographic patterns other than those found in learners' written language can be acquired by children and adults under well-controlled experimental conditions. This gap in the literature calls for a formal investigation of the mechanisms underlying orthographic sensitivity and was one of the two main aims of the body of work reported in this thesis. The second aim of the thesis was to investigate the extent to which an implicit learning impairment affecting dyslexic adults' ability to acquire chunks (i.e., pairwise associations between two/three adjacent letters), absolute single-letter positions and whole exemplars is associated with dyslexia. If a general learning deficit is partially responsible for dyslexic individuals' literacy difficulties (even indirectly, for example, by affecting the quality of their phonological representation; Gombert, 2003) their performance on a well-established implicit learning task (the artificial grammar learning task) should be impaired.

Finally, the work reported in this thesis was further informed by a few recent demonstrations of a link between children's and adults' ability to learn statistically defined patterns under laboratory conditions and their language skills (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kidd, 2012). Similarly, a handful of the implicit learning studies described earlier (section 1.3.3) have shown significant correlations between individual differences in literacy-skill and sequence learning task performance. These studies, including, to the best of our knowledge, the only statistical learning study that has examined the relationship between visual pattern learning performance and reading (Arciuli & Simpson, 2012) are discussed in the following chapters. A replication and extension of these results by demonstrating that individual variation in literacy and/or literacy-related skill correlates with our measures of learning skill was an additional aim of our experimental investigations. Therefore, all of our participants were administered measures of reading and spelling ability, among other cognitive and literacy-related measures, and correlational analyses are reported throughout the thesis.

1.5 Overview of the Experiments

Two experimental paradigms were employed throughout the thesis and were used in the cross-sectional implicit and statistical learning studies presented in Chapters 2 – 5. We began (Chapter 2) by considering whether zero- and first-order statistics governing allowable letter positions and letter co-occurrences within words can be incidentally acquired by adults and used in a subsequent legal/illegal word discrimination task. For this purpose, an incidental graphotactic learning task was developed following the methods of previous studies investigating novel phonotactic acquisition in adults (Onishi, Chambers, & Fisher, 2002). A slightly modified version of the task was used in Chapter 3 to investigate whether similar learning can be induced "early" in development. More specifically, a between-subject design was used to compare learning of (a) zero- and first-order constraints, (b) following shorter or longer exposure, (c) between two different learner groups (7-year-old children and adults).

Building on the artificial grammar learning literature, the work presented in Chapters 4 and 5 was set to investigate whether sensitivity to frequency-based information (bigrams and trigrams, letter positions and whole "training" stimuli) is spared or impaired in the presence of dyslexia. To this aim, dyslexic adults were administered one out of three artificial grammar learning variants (developed following Kinder & Lotz's 2009 design) and their learning performance was compared to that of age- and IQ-matched controls.

Chapter 2

Statistical Learning of Novel Graphotactic Constraints: A Pilot Study

2.1 Introduction

Statistical learning processes have been abundantly demonstrated to operate in the acquisition of natural languages from early infancy onwards. In seminal studies, Saffran and colleagues showed that 8-month-old infants, typically developing children and adults are able to compute and exploit syllabic transitional information in order to postulate word boundaries (e.g., Saffran et al., 1996a, 1997). Notably, probabilistic distributional learning has been shown to subserve other aspects of language acquisition, such as the ability to discover rudimentary syntax (e.g., Saffran & Wilson, 2003) and grammatical category membership (e.g., Mintz et al., 2002). Of particular relevance to the present study, several investigators have examined learning at the level of phonotactic constraints; that is, probabilistic constraints defining legal and illegal sound distributions (e.g., *zb* is an illegal sound combination in English; *str* is allowed in English, but not in word endings). In a series of well-documented experiments, Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993) provided evidence of 9-month-old infants' listening preference for unfamiliar words that complied with phonotactic constraints in their native language over words that were legal in another language to which they had never been exposed. For example, American infants favored words that complied with English phonotactics (e.g., /pɪ*tʃ*ɪz/), yet are illegal in Dutch. Conversely, Dutch infants listened longer to words embedding legal Dutch phoneme sequences (e.g., /opko*mst*/) which are not permitted in English. Given this early “tuning” to natural phonotactic constraints, recent studies have sought to examine the learning mechanisms that underlie this ability.

2.1.1 Incidental Phonotactic Learning Following Brief Listening Experience

Onishi and colleagues demonstrated that infants and adults become sensitive to *novel* phonotactic constraints embedded within aurally presented speech sequences, following brief listening experience (Chambers et al., 2003, 2010, 2011; Onishi et al., 2002). For example, in a speeded syllable repetition task, adult participants were faster at repeating novel syllables containing phonemes that conformed to experimental positional constraints (i.e., syllables with permissible phonemes such as /b/ in the onset position, e.g., /bæp/) than novel sequences containing phonemes that violated them (i.e., syllables with phonemes such as /p/ not permissible in the onset position, e.g., /pæb/). Furthermore, repetition RTs were faster for novel syllables embedding newly learnt contingencies between vowel phonemes and onset/ coda phonemes (context-based learning; e.g., /b/ can be an onset when the adjacent vowel is /æ/ but not /ɪ/). Under both experimental conditions, RTs yielded for syllables that were identical to those presented during exposure to the learning set were not significantly faster than RTs for novel syllables. That is, learning was not restricted to tokens in the learning set, and was not merely an artifact of familiarity with previously seen legal sequences. In a nutshell, the findings of Onishi et al.’s (2002) study suggest that learning about novel statistical distributions of speech sounds occurs rapidly and without engaging deliberate effort.

Experimental evidence from infant habituation studies suggests that such an unsupervised learning mechanism is present from infancy. In a study using the head-turn preference procedure, 10.5-month-old and 16.5-month-old infants who were habituated with critical syllables—similar to those presented to adults—demonstrated a novelty preference (i.e., longer listening times) for new items violating rather than conforming to the positional constraints (Chambers et al., 2003). In a follow-up study, 10.5-month-old infants also learnt context-based phonotactic constraints, following the same number of habituation trials. However, listening RTs were in the opposite preference direction, i.e., were reliably longer for legal syllables embedding the constraints (Chambers et al., 2011). Several factors, such as longer familiarization, are known to cause such a shift (familiarity → novelty) in the preference direction (Houston-Price & Nakai, 2004). Given that the number of familiarization trials was the same in both experimental conditions, the switch was taken to indicate the higher complexity of context-based constraints (Chambers et al., 2011).

What representations underlie the rapid acquisition of novel phonotactic constraints? Two different hypotheses have been put forward to explain the learning effects obtained in the previous studies with infants and adults. With regard to positional constraints, it is possible that learning involves forming representations of the absolute position of segments in words *independent* of context (e.g., /b/ is a legal onset). Nevertheless, learning about position at the segmental level needs not preclude learning of units larger than individual segments, and therefore the underlying representations may take context into account (e.g., /bæ/) even when such information is redundant. Representing units larger than segments is necessary for context-based pattern extraction. However, if units larger than segments are computed during positional learning, this may interfere with learners' ability to generalize their knowledge to syllables containing novel adjacent vowels. To address this question, Chambers et al. (2010) compared repetition RTs for legal and illegal syllables containing a transfer vowel that was different to the one used to create the training syllables. Following exposure to syllables such as /bæp/ and /bɛs/, participants were faster at repeating onset/coda-legal (e.g., /bɪp/, /bɪs/) than onset/coda-illegal syllables (e.g., /pɪb/, /sɪb/). Importantly, the legality advantage for test items containing the new, untrained vowel was comparable in size to the advantage observed for syllables containing a vowel that had been trained directly during the exposure phase. Generalization was not modulated by the training-transfer vowel similarity (e.g., similar vowels: /æ/ vs. /ɪ/; dissimilar vowels: /ɛ/ vs. /ʊ/). Thus, it was shown that newly learned positional constraints were represented, to some extent, abstractly, at the segmental level (Chambers et al., 2010). Converging evidence of generalized phonotactic learning to novel frames has been also obtained with longer strings of the form C_1VccVC_2 embedding positional constraints in word edges (Endress & Mehler, 2010). However, learning of *word-medial* phonotactic constraints of the type cVC_1C_2Vc (e.g., /ʅifsɑn/) versus cVC_2C_1Vc (e.g., ʅisfɑn/) was not possible when the grouping among C_1 or C_2 sounds was arbitrary, i.e., C_1 sounds did not share any phonetic features. This finding is important, as it highlights that edge-based positional regularities are easier to learn than middle-based positional regularities. What is more, it suggests that there are certain limits as to what can be learnt—at least in the context of brief laboratory experiments (Endress, Nespor, & Mehler, 2009).

2.1.2 Speech Error Studies

The studies reviewed above suggest that novel phonotactic learning can be induced in infants as well as adults, following brief listening experience and without engaging deliberate effort. Another line of research with direct relevance to the question of *how* learning occurs builds on the well-demonstrated phonotactic regularity effect (Fromkin, 1971). Involuntary natural speech errors tend to respect the phonotactic constraints of one's language and are rarely phonologically ill-formed. For example, English speakers rarely commit errors such as *dlorm* for dorm, *dl* being an illegal coronal-coronal onset in English (Stemberger, 1983 in Dell, Reed, Adams, & Meyer, 2000).

In a series of experiments, Dell et al. (2000) examined whether experimentally induced speech errors committed by English speakers respected artificial constraints on the allowable position of consonantal phonemes (e.g., the probability of /f/ as an onset was zero throughout the experiment; *experiment-wide constraints*). Participants were presented with and asked to repeat sequences of four CVC syllables (e.g., /hɛs/ /fɛŋ/ /mɛg/ /kɛn/) at a fast tempo. Exposure was given in four 96-trial sessions spanning 4 days, i.e., was much more extensive than the training provided in the Onishi studies. In line with the phonotactic regularity effect, there were virtually no speech errors that disobeyed the language-wide phonotactic constraints (i.e., in none of participants' errors /ŋ/ was misplaced as an onset). More importantly, following just one training session, speech errors resulting from a misplacement of an experimentally constrained phoneme (e.g., /f/) maintained their "legal" position more frequently than speech errors involving unrestricted consonantal phonemes. Similar rapid detection of positional constraints was evidenced in a more recent study by Taylor and Houghton (2005) using a slightly modified version of Dell et al.'s (2000) methods. A finding of relevance to our investigations was that adaptation to the experiment-wide constraints was as quick and impressive as was their "unlearning" in the presence of conflicting stimuli (Experiment 4; Taylor & Houghton, 2005).

Subsequent speech error studies extended Dell et al.'s (2000) findings by demonstrating learning of context-based constraints, such as those examined by Onishi and colleagues (Warker & Dell, 2006). Experiment-wide constraints (e.g., the vowel sound /æ/ is preceded by /k/ and followed by /g/; the vowel sound /ɪ/ is preceded by /g/ and followed by /k/) were upheld to high rates from training day 1; however, reliable learning was demonstrate only after two days of training, suggesting that longer exposure was needed for contextual constraint learning. What is more, Warker, Dell, Whalen, and Gereg (2008) provided evidence that more complex phonotactic-like dependencies (e.g., nonadjacent context-based constraints, such as, "if the medial consonant in a CVCVC sequence is /v/, /f/ is an onset and /s/ is a coda") can be also learnt using the same methods, although complexity largely affects the time course of pattern acquisition. To sum up, the evidence reviewed from listening and speech error studies is consistent with the theoretical claim that phonotactic constraints are rapidly encoded by speech production processes under incidental learning conditions. However, learning capacity is not unlimited (Endress & Mehler, 2010; Warker et al., 2008), at least under brief laboratory conditions.

2.1.3 The Case of Graphotactics

Similar to the speech domain, written language is subject to statistical regularities setting constraints on the possible (i.e., deterministic; acceptable vs. unacceptable) or probable (frequent vs. infrequent) successions of graphemes in script (Jaffré & Fayol, 1997). Graphotactic acceptability varies as a function of language (e.g., the digraph *ll* is illegal as an onset in English, but is frequently encountered in Welsh, e.g., Llan); however, most alphabetic orthographies embed constraints on the position and context of letter distributions. Positional constraints encode information on the legal position of particular segments within the syllables or words of a given orthography. For example, the probability of several plausible graphemes as a representation of the sound /k/ (*c*a*k*e, *ch*aracter, pi*ck*) is restricted in English by graphotactic constraints permitting *ck* in medial or word-final, but not word-initial positions. Contextual constraints relate information on a segment's position to another aspect of the orthographic context, such as the identity of its neighboring segments. For example, *s* is the only allowable spelling of a word-final /z/ sound when the latter is preceded by a consonant grapheme such as *d* or *g*. That is, *gz* or *dz* are graphotactically illegal spellings of sound combinations that occur frequently at the end of English words (e.g., /bagz/, /padz/).

What learning mechanisms underlie the acquisition of graphotactic pattern knowledge? While children are typically taught to spell through systematic explicit instruction in the classroom, learning to spell is not likely to be determined by explicit processes alone (Steffler, 2001), perhaps especially in highly inconsistent orthographies. Spelling rules that are more probabilistic in nature are hard to teach overtly or even to state (Kessler, 2009). Thus, a plausible hypothesis is that spellers additionally acquire graphotactic knowledge in an implicit fashion, i.e., through simple exposure to print or through their reading experiences. By this view, children are powerful "statistical learners" who observe and internalize the relative frequency with which letters or letter combinations occur and co-occur (Pollo et al., 2007). To date, this interesting and increasingly agreed upon proposal has not been tested under well-controlled experimental conditions, and several interesting questions arise in light of the evidence presented in the previous sections. Can graphotactic learning be induced under brief incidental experimental conditions? Does sensitivity to novel graphotactic constraints develop as rapidly as novel phonotactic learning?

2.1.4 Statistical Learning and Written Language

A recent line of research suggests that statistical learning ability may account for unique variance in different aspects of language-related performance. Two studies in the oral language domain have provided evidence of an association between (a) the ability to learn sequential structure implicitly and performance on a measure of sentence perception under degraded listening conditions (Conway et al., 2010); (b) performance on a similar measure of implicit learning and the ability to comprehend orally presented sequences embedding adjacent and nonadjacent language dependencies (Misyak & Christiansen, 2012). In a similar vein, a relationship has been found between adults' nonadjacent statistical learning performance (on an artificial grammar learning/serial reaction time task hybrid) and their ability to process syntactically complex sentences embedding distant dependencies (object-relative sentences such, *the reporter* that the senator attacked *admitted* the error) (Misyak, Christiansen, & Tomblin, 2010).

Reading performance has been also shown to positively correlate with visual statistical learning skill in experienced readers (adults) and typically developing children (kindergarten to Year 6) (Arciuli & Simpson, 2012). Arcuili and Simpson (2012) employed an embedded triplet paradigm (similar to the one used by Fiser & Aslin, 2001; section 1.1.3 of the Literature Review) assessing learning of temporal relationships among visually presented "alien" figures. It was shown that children and adults with higher scores on the Reading subtest of the Wide Range Achievement test (a standardized single word recognition measure; Wilkinson & Robertson, 2006) also scored higher on the learning task. A regression analysis partialling out age and participants' accuracy during the exposure phase of the experiment revealed that statistical learning accounted for a small but significant amount of unique variance in reading ability.

Results from another series of studies with children and adults with prominent literacy-related deficits provide converging evidence for the proposed link between statistical learning and reading/spelling ability. For example, modest to large positive correlations were observed between skilled/dyslexic participants' learning scores on Howard et al.'s (2006) alternating serial reaction time task (see section 1.5.2) and their performance on a standardized measure (Woodcock & Johnson, 1990) of real word (r = .59) and nonword reading (r = .52). Poor spelling skill was also significantly associated with reduced learning in the serial reaction time task, r = .42; however the relationship no longer remained when reading skill was partialled out. The same positive relationship between reading and learning scores was obtained in Bennett et al.'s (2008) follow-up study assessing learning by means of a similar but less motoric reliant sequence learning task. These findings are encouraging for drawing links between individual-to-individual variability in implicit/statistical learning skill and literacy but warrant further investigation as they are not always replicated. In a large scale study with garden variety poor readers (n = 422), implicit sequence learning performance was unrelated to reading skill (measured by the Basic Reading subtest of the Wechsler Individual Achievement Test; Wechsler, 1992) (Waber et al., 2003).

2.2 The Paradigm

The experimental paradigm introduced in this chapter sought to induce pattern learning following brief visual exposure. Our methodology built on a large body of work employing the two phase (study-test) design to investigate incidental learning in a laboratory setting. During phase 1 of experiments of this sort, participants are exposed to a set of stimuli which appear arbitrary, yet conform to a set of rules or embed relevant patterns. Following exposure to such "adequate" stimuli, participants are given a surprise test phase (phase 2) assessing their sensitivity to the relevant underlying statistics. Learning at test is typically measured by means of performance on two alternative force choice tests (statistical learning studies: e.g., Fiser & Aslin, 2001, 2002), well-formedness judgment tasks (artificial grammar learning studies: e.g., Johnstone & Shanks, 2001), or choice reaction time tasks (spatial context learning studies: e.g., Chun & Jiang, 1998; sequence learning studies: Nissen & Bullemer 1987).

Our task was an adaptation of the incidental phonotactic learning task introduced by Onishi et al. (2002). A typical exposure/test phase trial is shown in Figure 2.1. Exposure was given in the context of a colour detection task in which participants were asked to detect the position of a red letter. We used a colour detection task to ensure that participants were attending to the stimuli without necessarily reading them. More importantly, we sought to induce learning following brief stimulus examination, i.e., in a way that is somewhat similar to real-life graphotactic learning as conceived by the statistical learning perspective of spelling. In line with artificial grammar learning studies, we employed a well-formedness task, during which participants were asked to classify items based on whether they comply with the "rules" used to generate the items presented during exposure. The well-formedness task has been extensively used over the last 40 years in its original version (e.g., Kinder & Lotz, 2009; Knowlton & Squire, 1996) or in modified versions (e.g., Whittlesea & Dorken, 1993).

As in the Onishi et al. (2002) study, learning in our experiments was induced in a deterministic, "all-or-none" fashion. For example, with regard to positional constraints learning, letters from one set (e.g., *d*) were restricted to onset positions (e.g., *d*et) and never appeared as codas. Inversely, letters from set 2 (e.g., *t*) were restricted to codas and never appeared as onsets. It is worth highlighting that both constraints on onsets and codas appeared simultaneously. In other words, participants were given both cues to legality. For example, participants could infer legality in the positional constraints variant by attending to positional constraints on onsets, codas or both. Similarly, participants tested under the contextual constraints conditions could be tracking co-occurrence statistics between onsets and the adjacent following vowel, codas and the adjacent preceding vowel, or both.

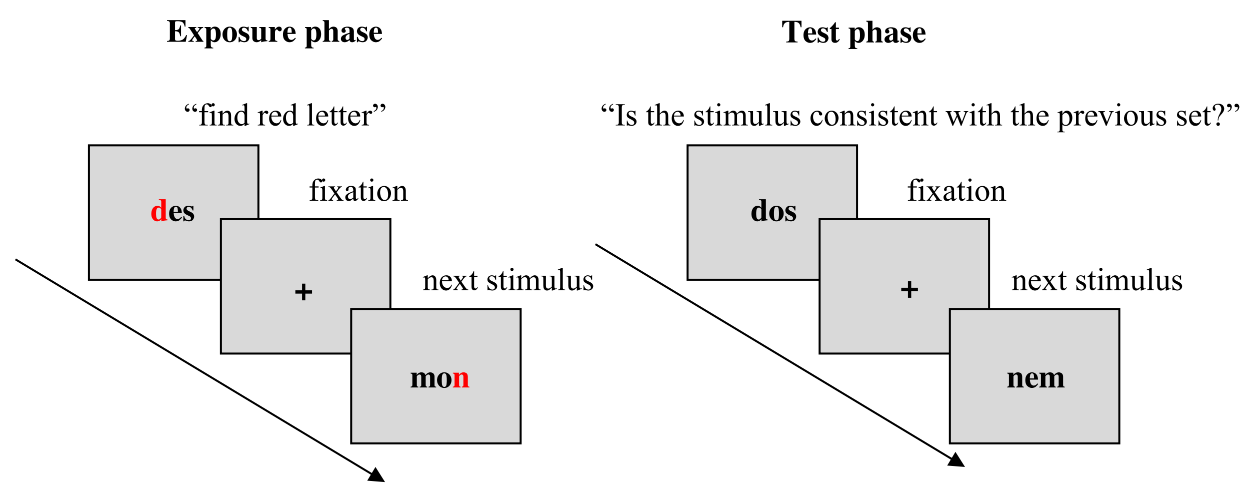


Figure 2.1. The incidental graphotactic learning task. Trial outline.

2.3 Experiments 1.1 – 1.3

2.3.1 Rationale of a Pilot Study

The aim of the pilot study was to (a) validate the incidental graphotactic learning task adapted from Onishi et al. (2002) and (b) replicate the effects observed by Onishi et al. (2002) (phonotactic learning) in the visual domain (graphotactic learning). In order to compare variations in learning as a function of different types of orthographic material (i.e., linguistic vs. nonlinguistic), two different alphabets were used for stimulus construction. For the so-called linguistic version, Latin letters were used to create stimuli consisting of graphotactically plausible word-like CVC strings. All of the stimuli contained onset, vowel and coda graphemes as well as bigrams (CV, VC) that are permissible in the English orthography (and phonology), and all conformed to English graphotactic rules for permissible onset, vowel, and coda spellings for monosyllabic words (although a small number of tokens occurred less frequently in the English orthography). Thus, participants were not exposed to unnatural graphotactic patterns, but merely to novel sequences and relative distributions of these. For the nonlinguistic version, strings were constructed using ten letters from the Greek alphabet. We assumed that our monolingual English participants would perceive the Latin letter strings as plausible English syllables, while the Greek letters would be perceived as visual symbols having no interpretable linguistic value. However, it is acknowledged that participants may have been, to some extent, familiar with the Greek alphabet. For example, students may have come across some of the letters through statistics classes; however, the Greek-letter strings were not verbalizable to these participants.

In order to address the issue of the type of the learning underlying task performance, we created one 2- and one 3- vowel version of the Latin-letter positional constraints learning variant. Participants in the 2-vowel positional constraints learning variant were exposed to 16 CVC strings made with two different vowels (o, e) presented 9 times each. For participants in the 3-vowel variant, syllable frames were combined with 3 different vowels (o, e, and i), resulting in 24 CVC strings presented 6 times each. This manipulation contrasted the influence of frequency of token repetitions with the influence of the variety of tokens available for learning. The underlying statistics were the same in both versions (see Material section). In other words, although participants were exposed to the same number of stimuli, the 3-vowel version included more tokens available for learning, whereas the 2-vowel version included fewer tokens which appeared more frequently. If participants' performance reflected more "abstract" learning of the relevant sub-lexical constraints, then discrimination ability should be higher in the 3-vowel positional constraints variant. Alternatively, if the underlying representations were more tied to the whole-syllable instances, performance should be higher in the 2-vowel positional constraints variant.

Taken together, these experiments allow us to assess whether statistical learning processes contribute to the acquisition of orthographic knowledge in skilled adult learners. In addition to the incidental graphotactic learning task, our experimental design included standardized measures of reading and spelling ability. Recent studies suggest that implicit skill learning correlates with a variety of cognitive and language measures. However, few replications have thus far been attempted. To this end, we sought to examine the relationship between incidental graphotactic learning and general reading or spelling ability.

2.3.2 Method

2.3.2.1 Participants

Seventy undergraduate students (17 male, 53 female) from Bangor University participated in the study in exchange for course credit. Thirteen participants were excluded from the analyses for one of the following reasons: i) native language other than English (n = 3), ii) reported history of dyslexia and/or other learning difficulties (n = 4), iii) failure to comply with the experimental instructions (n = 1), iv) English Welsh bilinguals (n = 5). The remaining 57 individuals (13 male; 44 female) were monolingual native English speakers with no learning difficulties and normal or corrected-to-normal visual acuity. Each participant was randomly allocated to one of three experimental groups (see details of Design below). All participants gave their informed consent before the start of the experimental session.

2.3.2.2 Material

2.3.2.2.1 Incidental Graphotactic Learning Task

Linguistic variants. Following the paradigm of Onishi et al. (2002), 32 Ci_Cj syllable frames were created by using two sets of four consonants (d, m, l, f [set 1]; t, n, p, s [set 2]) and combining them with each other. For half of the syllable frames, set 1 consonants were used as onsets and set 2 consonants as codas (C_1_C_2 sequences; e.g., d_t), whereas for the other half the reverse was true (C_2_C_1 sequences; e.g., t_d). In 2-vowel versions, the vowels *o* and *e* were used to fill in the syllable frames, giving rise to a total of 64 syllables; in the 3-vowel version, the vowel *i* was additionally used, resulting in a total of 96 syllables. The sequences were arranged in two pairs of lists—that is, four lists in total, and the presentation of these was counterbalanced within experimental groups[4]. Although the particular sets of consonants and vowels were chosen in an effort to keep the number of real CVC words to a minimum, 8 (out 64) and 13 (out of 96) strings were unavoidably real words in the 2- and 3-vowel version, respectively. In this study, these were not excluded from any of the analyses.

For positional constraints learning in one of the counterbalanced list conditions, one C_1VC_2 list (e.g., list 1 in the 2-vowel version: n = 16; 3-vowel version: n = 24) served as exposure items. The same list served as legal seen items at test, whereas the other C_1VC_2 list (e.g., list 2 in the 2-vowel version: n = 16; 3-vowel version: n = 24) served as legal unseen test items at test. C_2VC_1 items (list 3 in the 2-vowel version: n = 16; 3-vowel version: n = 24) served as illegal items during the test phase. Items presented during the exposure phase were governed by the following underlying zero-order statistics: i) in lists 1 and 2, for example, there was an equal probability of appearance of any C1 letter as an onset, e.g., p(d) = .25, whereas the probability of appearance of any C2 letter as an onset, e.g., p(t), was zero, ii) there was an equal probability of appearance of any C2 letter as a coda, e.g., p(t) = .25, whereas the probability of

---

[4]Unfortunately, due to experimenter's error, lists 3 and 4 did not serve as exposure/legal seen/legal unseen materials in Experiments 1.1 – 1.3; only list 1 and list 2 items were used for this purpose and were counterbalanced across participants. That is, approximately half of the participants in each experimental condition were exposed to list 1 items and provided legality judgments on list 1 (legal seen), list 2 (legal unseen) and list 3 (illegal) items; the remaining participants were exposed to list 2 items and provided legality judgments on list 2 (legal seen), list 1 (legal unseen) and list 3 (illegal) items.

appearance of any C1 letter as a coda, e.g., p(d), was zero. The full set of experimental stimuli used for positional constraints learning is shown in Appendix A.

For contextual constraints learning, stimuli were recombined in two pairs of lists in order to introduce a first-order constraint according to which consonant position depended on the vowel type. For one of the counterbalanced list conditions, stimuli with set 1 consonants as onsets and set 2 consonants as codas were combined with the vowel o (e.g., fos), whereas sequences with set 2 consonants as onsets and set 1 consonants as codas were combined with the vowel e (e.g., tem). Assignment to list was counterbalanced across participants in a similar manner to positional constraints learning. For example, in one condition, list 1 items (n = 16) served as exposure and legal seen test items, list 2 items (n = 16) were presented as legal unseen test items, and list 3 items (n = 16) were presented as illegal test items. The full set of experimental stimuli used for contextual constraints learning is shown in Appendix B. In contrast to the positional constraints variant, the probability of appearance of any C_1 or C_2 letter as an onset or coda was equated for contextual constraints learning, e.g., p(d) = p(t) = .125. Exposure items were governed by the following underlying first-order statistics: In lists 1 and 2, for example, there was an equal joint probability of occurrence of the vowel o and any C_1 letter as an onset, e.g., p(d, o) = p(m, o) = .125, and ii) the vowel e and any C_2 letter as an onset, e.g., p(t, e) = p(n, e) = .125. Consequently, the joint probability of appearance of any C1 letter and e, e.g., p(d, e), or any C_2 letter and o, e.g., p(m, o), was zero. A corollary of this design was that VC contingencies, e.g., p(o, t) or p(o, n) also occurred with a probability of .125. Thus, participants could benefit from first-order contingencies in the CV or VC portions of the stimuli. Importantly, these first-order contingencies occurred systematically and with the same statistical probabilities throughout the task.

Nonlinguistic variants. A positional constraints variant (2 vowels) and a contextual constraints variant (2 vowels) was devised using characters from the Greek alphabet. Letters were selected to bear the minimal possible resemblance to the Latin letters. The letters of each sequence were transposed using the following rules: d → θ, m → δ, l → τ, f → π, t → ξ, n → φ, p → γ, s → λ. Similarly, the vowels o and e were replaced with the Greek letters α and η. Every other aspect of the material used in these versions was identical to the Latin-letter-2-vowel versions. The full set of experimental stimuli used for the nonlinguistic versions of the positional and contextual constraints learning are shown in Table C1 and Table C2 in the Appendix.

2.3.2.2.2 Literacy Measures

All adult participants were assessed on standardized literacy ability measures in addition to the experimental task. As in Arciuli and Simpson's (2012) study, the Word Reading and Spelling subtests of the Wide Range Achievement Test-Forth Edition (WRAT-IV; Wilkinson & Robertson, 2006) were used to measure reading and spelling achievement.

2.3.2.3 Apparatus

Computer experiments were run on a Windows XP-based PC with a 15" CRT colour monitor. Stimulus displays were generated using E-prime software (Schneider, Eschman, & Zuccolotto, 2002). Presentation of stimuli and millisecond accurate response registration was achieved by means of the same software package. Participants gave manual responses using a standard QWERTY keyboard.

2.3.2.4 Design

Participants were randomly allocated to one of three experimental groups: Participants in group 1 (6 male; 14 female; $n = 20$) were administered the 2-vowel variant positional constraints learning with Latin letters, followed by administration of the contextual constraints learning Greek letter variant. Participants in group 2 (2 male; 16 female, $n = 18$) were administered the 3-vowel positional constraints learning Latin letter variant. Participants in group 3 (5 male; 14 female, $n = 19$) were tested on the contextual constraints learning Latin letter variant, followed by administration of the 2-vowel positional constraints learning Greek letter variant.

Variability of learning set (smaller vs. larger corpus of stimulus tokens) was a between-subject manipulation (group 1 vs. group 2). Type of letter set (Latin vs. Greek) and type of constraint (position vs. context) were manipulated within- as well as between-subjects. Given the pilot character of this study, we did not attempt to create a fully balanced design. Therefore, we did not assess learning of positional/contextual constraints in a 3-vowel variant with Greek letters or contextual constraints learning in a 3-vowel variant with Latin letters. Further potential limitations of our mixed design are discussion in the General Discussion section. At minimum, our design allowed us to partially control for effects of repetition in performance during the second task. That is, participants were never administered two variants embedding the same type of constraint. In addition, directly comparable versions (e.g., positional constraints learning with Latin letters vs. positional constraints learning with Greek letters) were always administered in a between-subject manner.

2.3.2.5 Procedure

Participants were tested individually in a quiet room. The incidental graphotactic learning tasks were administered prior to the administration of the Reading and Spelling subtests of the Wide Range Achievement Test (WRAT-IV; Wilkinson & Robertson, 2006). Testing took place in a single experimental session which took approximately 30 minutes to complete.

Incidental graphotactic learning task. The task consisted of an exposure phase, a distraction phase and a test phase. The exposure phase was presented as a (red) colour detection task. At the beginning of each exposure trial, a three-letter string with one red and two white letters was displayed at the center of a black background. Participants were asked to respond to the position of the red letter as fast as possible without sacrificing accuracy by pressing one of three keys (*b, n, m*) corresponding to left edge, middle and right edge of the string. Strings remained on the screen until a response was collected and no feedback was given. Stimulus presentation was followed by a fixation point (white cross) also appearing centrally for 500 ms. A practice block of 6 trials was given in order to familiarize participants with the task. This was followed by 3 blocks of 48 trials each (trial order was randomized for each participant), for a total of 144 (2 vowel versions: 9 repetitions/string; 3 vowel version: 6 repetitions/string) trials. The target red letter appeared in each of the first, mid, or final position one third of the time. After all exposure strings were presented, participants moved to a short distraction phase involving simple calculations. Participants were presented with 10 simple arithmetic facts (single-digit additions of the form [x + y = z], with $1 \leq x, y \leq 8$ and $2 \leq z \leq 9$) and were asked to press on the number key corresponding to the correct result. No feedback was given during this phase of the experiment.

Completion of the distraction phase was followed by a “surprise” test phase: Participants were informed that the strings that they had previously seen were constructed according to a set of rules. They were told that during this phase they would see a set of new strings and were to decide whether each one of them “goes well” with the rules. Similar to the instructions given in implicit learning experiments, participants were encouraged to trust their “gut feeling” whenever they were not sure about their response. Each CVC test item appeared in the center of the screen and participants pressed one button if they judged it to conform to the previously seen items and another if they judged to the contrary. Items remained on the screen until a response was given. During the test phase, the full set of exposure items (legal seen items; $n = 16$) was presented together with a set of new strings that embedded the same graphotactic constraints (legal unseen items; $n = 16$), and a set of strings that violated the constraints (illegal items; $n = 16$). Each trial was followed by a fixation point (white cross) appearing at the center of the computer screen for 500 ms. All stimuli appeared in a single block (order of presentation was randomized for each participant) and no feedback was given.

2.3.3 Data Analyses

All analyses on reaction times and accuracy-based measures were conducted separately. RTs below 300 ms and above 5000 ms were identified as extreme outliers and were removed from all RT analyses. RTs more than 2.5 standard deviations above or below the same participant’s mean RT were trimmed to the values at 2.5 standard deviations from the mean. A criterion of $\alpha = .05$ was used for all statistical tests and all reported p values are two-tailed unless stated otherwise. To interpret significance for data that did not meet the requirement of sphericity (Mauchley’s test of sphericity), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Significance values resulting from a priori and post hoc multiple comparisons were corrected for type I error using Bonferroni’s method. Effect size measures—eta squared for analyses of variance (ANOVAs) and Cohen’s d for one-sample t tests—are reported.

2.3.4 Results

2.3.4.1 Incidental Graphotactic Learning task

Exposure phase. The main objective of the exposure phase analyses was to ascertain the validity of the exposure format, in that, participants were duly attending to the stimuli, and were performing the colour detection task reliably across blocks. We expected (a) letter detection accuracy to be very high and (b) RT performance over blocks of trials to be relatively stable. Our aim was not to compare performance as a function of other factors (e.g., type of constraint/task order) for this validation exercise. Participants were at ceiling in detecting the red letter in every block (> .98), thus, accuracy data were not further analyzed. RT performance did not vary as a function of counterbalanced list; we therefore omit further discussion of this factor. Mean correct RTs in each experimental condition (Figure 2.2) were subjected to separate one-way repeated-measures ANOVA with block (1, 2, 3) as a within-subject variable. Correct exposure RTs were stable in the 2-vowel positional constraints learning with Latin letters, $F(2, 38) = 1.82$, $p > .05$, $\eta^2 = .09$, and in the 3-vowel positional constraints learning with Latin letters, $F(1.21, 20.58) = 3.67$, $p > .05$, $\eta^2 = .18$. There was a main effect of block in the 2-vowel positional constraints learning with Greek letters, $F(1.21, 20.58) = 3.67$, $p < .05$, $\eta^2 = .21$, negative linear trend, $F(1, 18) = 5.52$, $p = .030$, capturing a small effect of practice; however, the range of mean correct RTs across the three exposure blocks was very small, indicating that differences were probably negligible. The ANOVAs revealed two additional main effect of block, in the contextual constraints variant with Latin letters, $F(1.45, 26.18) = 4.61$, $p < .05$, $\eta^2 =.20$, quadratic trend, $F(1, 18) = 7.10$, $p = .016$; and in the contextual constraints variant with Greek letters, $F(1.03, 19.57) = 4.62$, $p < .05$, $\eta^2 = .20$, quadratic trend, $F(1, 19) = 7.66$, $p = .012$. Similar to the 2-vowel positional constraints learning with Greek letters, it is unlikely that such small RT difference across blocks (Figure 2.2) were meaningful.

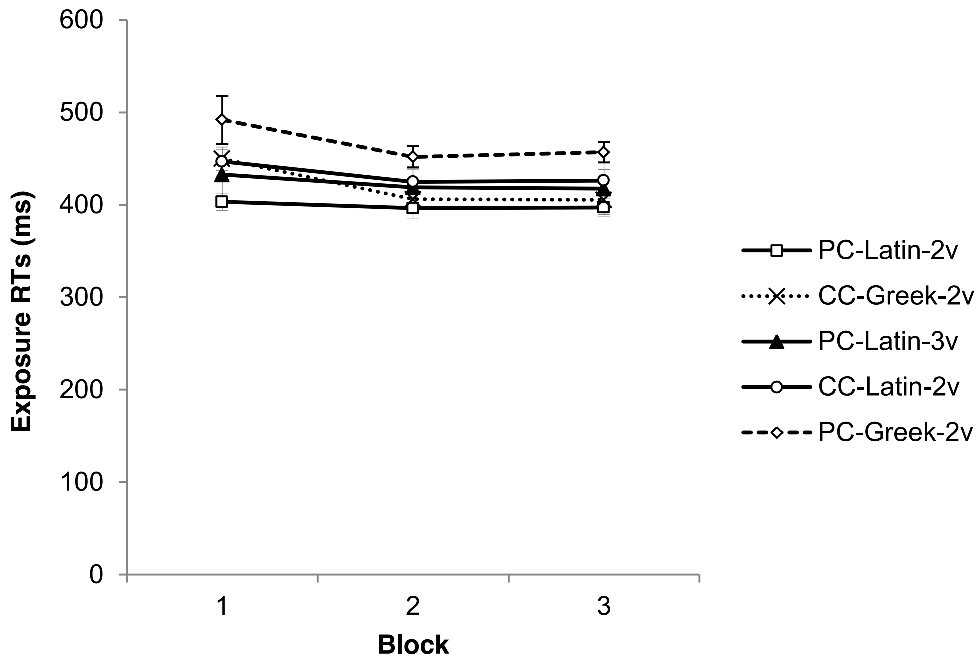


Figure 2.2. Exposure RTs in Experiments 1.1 – 1.3. Mean correct exposure RTs (ms) as a function of block plotted separately for each experimental condition. Error bars represent standard error of the mean.

Test phase. Discrimination accuracy data were subjected to signal detection theory analyses (Snodgrass & Corwin, 1988). In order to correct for any response bias towards rejecting or accepting items (i.e., criterion shifts), a measure of legality sensitivity (*d*') was calculated for each experimental condition. *d*' scores were computed by calculating the difference between the *z*-transformed proportion of "yes" responses to legal sequences (i.e., sequences that were correctly classified as legal; hit rate) and the *z*-transformed proportion of "yes" responses to illegal sequences (i.e., sequences that were mislabeled as legal; false alarm rate). Response bias was measured with the criterion score *c*, calculated by averaging *z*(hits) and *z*(false alarms) and multiplying the result by -1. Following Macmillan and Kaplan (1985), hit or false alarm rates of 0 were replaced with $1/2n$ where n corresponds to the number of signal or noise trials, respectively. Rates of 1 were replaced with $1 - 1/2n$.

Discrimination accuracy. All data points were included for analyses of discrimination accuracy. Table 2.1 presents the mean proportion of items endorsed as legal as a function of stimulus type. Mean *d*’ values in each experimental condition are shown in Figure 2.3. One-sample *t* tests (chance = 0) confirmed that participants reliably classified more test phase stimuli correctly than expected by chance in the 2-vowel positional constraints variant with Latin letters, $t(19) = 8.84$, $p < .001$, $d = 1.98$, in the 3-vowel positional constraints variant with Latin letters, $t(17) = 8.95$, $p < .001$, $d = 2.11$, and in the 2-vowel positional constraints variant with Greek letters, $t(18) = 6.47$, $p < .001$, $d = 1.49$. Sensitivity to first-order context-based constraints was also above chance in terms of performance in the contextual constraints Latin-letter variant, $t(18) = 2.28$, $p < .05$, $d = 0.50$, as well as in the contextual constraints Greek-letter variant, $t(19) = 2.22$, $p < .05$, $d = 0.50$. However, it is noticeable that effect sizes for learning in these conditions were only medium.

Table 2.1. Proportion of Items Endorsed as Legal (*SD*s) as a Function of Stimulus Type in Each Experimental Condition (Experiments 1.1 – 1.3).

Condition	Endorsement rates			
	Illegal (FAs)	Legal (Hits)		
		Seen	Unseen	Total
PC-Latin-2v	.30 (0.20)	.69 (0.15)	.62 (0.18)	.66 (0.15)
PC-Latin-3v	.17 (0.13)	.66 (0.18)	.55 (0.15)	.61 (0.15)
PC-Greek-2v	.29 (0.17)	.68 (0.16)	.66 (0.13)	.67 (0.14)
CC-Latin-2v	.43 (0.19)	.64 (0.19)	.45 (0.12)	.55 (0.12)
CC-Greek-2v	.48 (0.16)	.53 (0.18)	.55 (0.12)	.54 (0.13)

Note. PC= Positional Constraints. CC = Contextual Constraints. v = vowels. FA = False Alarms.

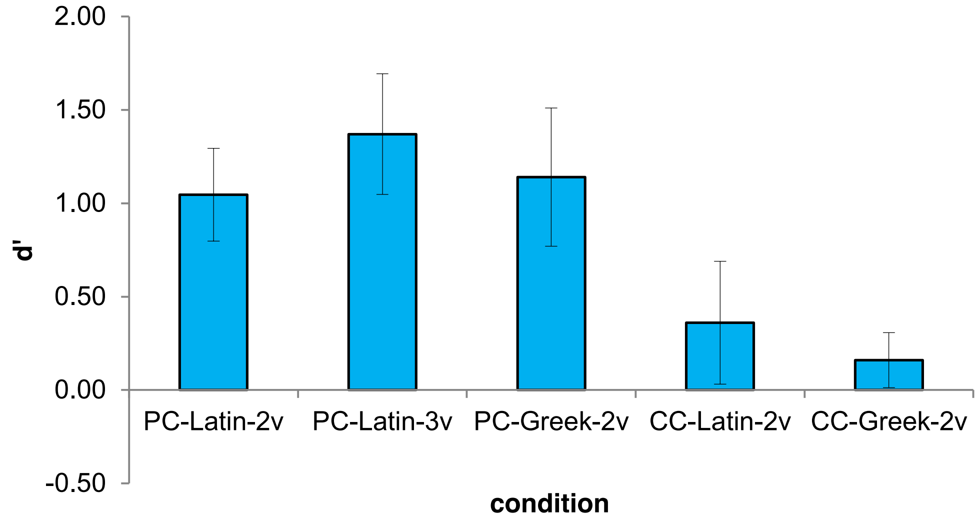


Figure 2.3. Discrimination accuracy in Experiments 1.1 – 1.3. Measures of discrimination accuracy, *d*', for detecting legal items across the five experimental conditions. Error bars represent 95% confidence intervals.

Next, Bonferroni-corrected independent sample *t* tests (*p* value adjusted to .005 to control for all possible post hoc comparisons) were carried out to explore differences in the magnitude of learning as a function of our experimental manipulations (position vs. context; linguistic vs. nonlinguistic material; lower token variety vs. higher token variety). Sensitivity to novel positional constraints was systematically greater than sensitivity to novel contextual constraints. This held true for (a) performance in the 2-vowel positional constraints learning variant with Latin letters relative to performance in the 2-vowel contextual constraints learning variant with Latin letters, $t(37) = 3.56$, $p = .001$, $d = 1.14$, (b) performance in the 3-vowel positional constraints learning variant with Latin letters relative to performance in the 2-vowel contextual constraints learning variant with Latin letters, $t(35) = 4.62$, $p < .001$, $d = 1.52$, and (c) performance in the 2-vowel positional constraints learning variant with Greek letters relative to performance in the 2-vowel contextual constraints learning variant with Greek letters, $t(37) =$

5.30, $p < .001$, $d = 1.68$. All effect sizes were large showing that the type of constraint manipulation resulted in a robust advantage of positional over contextual constraints learning.

With regard to the type of letter set manipulation (linguistic vs. nonlinguistic material), no difference was found between the 2-vowel positional constraints variants with Latin and Greek letters, $t(37) = 0.44$, $p > .05$, $d = 0.14$. Similarly, no difference was observed between the 2-vowel contextual constraints variants with Latin and Greek letters, $t(37) = 1.19$, $p > .05$, $d = 0.38$. Finally, pertinent to the issue of the type of learning underlying performance (potential learning of whole-syllable instances vs. "abstraction" of the relevant sub-lexical constraints), we observed a trend for higher *d*' values in the 3-vowel Latin-letter positional constraints variant relative to *d*' values in the 2-vowel positional constraints Latin-letter variant, however, the difference did not reach statistical significance, $t(36) = 1.67$, $p > .05$, $d = 0.54$.

Bias *c* analyses. Similar to *d*' scores, mean *c* values for each experimental group, shown in Figure 2.4, were subjected to one-sample *t* tests against the chance value of 0. According to the signal detection theory, bias *c* values below this value indicate liberal bias, i.e., participants' tendency to accept items, whereas bias *c* values above this value indicate conservative bias, i.e., a tendency to reject items (Macmillan & Kaplan, 1985). For participants in the 3-vowel positional constraint variant with Latin letters, the detection theory analysis revealed a significant bias towards conservatism, i.e., responding "no", $t(17) = 4.31$, $p < .001$, $d = 1.01$. There was no response bias in any other experimental condition: positional constraints 2-vowels with Latin letters, $t(19) = 0.79$, $p > .05$, $d = 0.18$; positional constraints 2-vowels with Greek letters, $t(18) = 1.10$, $p > .05$, $d = 0.25$; contextual constraints 2-vowels with Latin letters, $t(18) = 0.73$, $p > .05$, $d = 0.16$; contextual constraint 2-vowels with Greek letters, $t(19) = -0.38$, $p > .05$, $d = 0.09$.

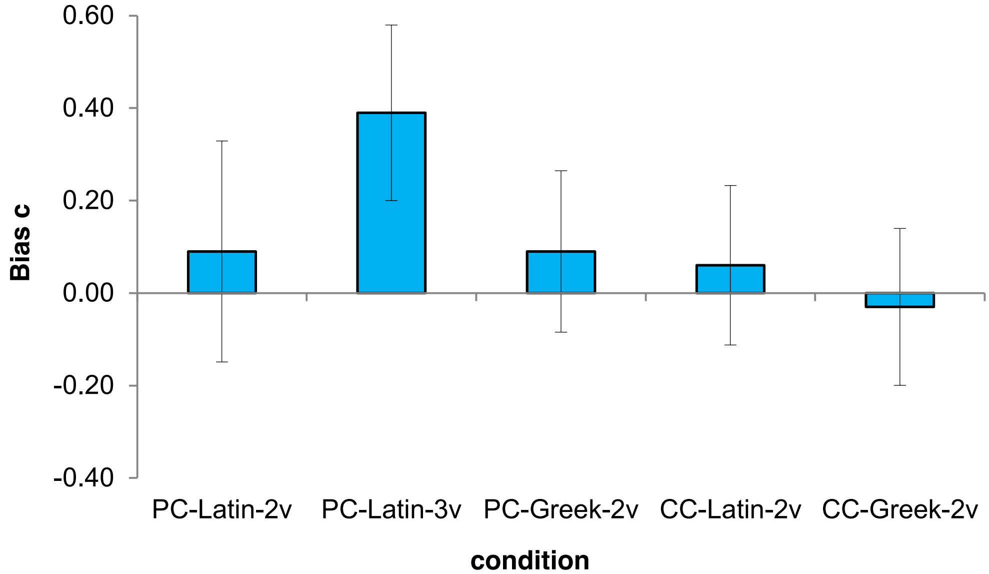


Figure 2.4. Response bias in Experiments 1.1 – 1.3. Measures of response bias, *c*, across the five experimental conditions. Error bars represent 95% confidence intervals.

Sensitivity to unseen vs. seen items. Analyses of discrimination accuracy indicated that learning was above the chance level across all five experimental conditions. However, given that participants were tested on their ability to judge previously seen as well as unseen items, we wished to determine whether learning was driven by successful recognition of items that were identical to those presented in the exposure phase of the experiment. To this aim, we calculated two additional, separate measures of discrimination ability: one by taking into account participants' rate of endorsement of legal unseen items, $d'_{\text{unseen}} = z(\text{hits legal unseen}) - z(\text{FAs})$, and one by taking into account participants' rate of endorsement of legal seen items, $d'_{\text{seen}} = z(\text{hits legal seen}) - z(\text{FAs})$. Mean d'_{unseen} and d'_{seen} are shown separately for each experimental condition in Table 2.2. These were subjected to one-sample *t* tests against zero, as well as paired *t* tests (reported in Table 2.2) to examine whether sensitivity was more reliable for one of the two types of items (legal unseen, legal seen).

One-sample *t* tests confirmed that in the 2-vowel positional constraints learning variant with Latin letters both legal unseen and legal seen items were reliably discriminated with better than chance accuracy, $t_1(19) = 6.98$, $p < .001$, $d = 1.56$, $t_2(19) = 9.38$, $p < .001$, $d = 2.10$. There was a trend for higher sensitivity to legal seen than unseen items, which did not reach significance. Participants in the 3-vowel positional constraints learning condition with Latin letters were also sensitive to legal unseen items, $t(17) = 8.91$, $p < .001$, $d = 2.10$, as well as legal seen items, $t(17) = 8.41$, $p < .001$, $d = 1.99$. In addition, they were significantly more sensitive to the latter. Performance in the 2-vowel positional constraints learning condition with Greek letters resembled that found in the 2-vowel positional constraints learning condition with Latin letters. Legal unseen items, $t(18) = 6.88$, $p < .001$, $d = 1.58$, and legal seen items, $t(18) = 5.87$, $p < .001$, $d = 1.35$, were discriminated with significantly better than chance accuracy, and there was no difference in *d*' scores between the two item subsets. In the 2-vowel contextual constraints learning condition with Latin letters, legal unseen items were not reliably discriminated with above chance accuracy, $t(18) = 0.97$, $p > .05$, $d = 0.22$. Only legal seen items were, $t(18) = 2.67$, $p < .05$, $d = 0.61$, and there was a significant difference in *d*' values in their favour. Finally, in the 2-vowel contextual constraints learning condition with Greek letters, neither legal unseen nor neither legal seen items were discriminated with better than chance accuracy, $t_1(19) = 1.98$, $p = .063$, $d = 0.61$, $t_2(19) = 1.68$, $p > .05$, $d = 0.38$. There was no difference in participants' sensitivity to the two subsets of items.

Table 2.2. Measures of Discrimination Accuracy for Legal Unseen and Legal Seen Items in Each Experimental Condition (Experiments 1.1 – 1.3).

Condition	SDT measures		paired *t* test results		
	d' unseen ± ME	*d*' seen ± ME	*t*	*p*	Cohen's *d*
PC-Latin-2v	0.96 ± 0.29	1.17 ± 0.26	1.93	.069	0.43
PC-Latin-3v	1.21 ± 0.29	1.57 ± 0.39	4.05	.001	0.96
PC-Greek-2v	1.10 ± 0.34	1.20 ± 0.43	1.25	.226	0.29
CC-Latin-2v	0.12 ± 0.26	0.61 ± 0.48	3.02	.007	0.69
CC-Greek-2v	0.18 ± 0.19	0.14 ± 0.17	0.40	.695	0.09

Note. SDT = Signal Detection Theory. ME = Margin of Error. PC= Positional Constraints. CC = Contextual Constraints. v = vowels.

Latency Analyses. Trimmed data accounted for a small number of all correct test phase responses in each experimental conditions (2-vowel positional constraints variant with Latin letters, $n = 12$; 3-vowel positional constraints variant with Latin letters, $n = 15$; 2-vowel positional constraints variant with Greek letters, $n = 17$; 2-vowel contextual constraints variant with Latin letters, $n = 13$; 2-vowel contextual constraints with Greek letters, $n = 10$). Mean correct trimmed RTs for illegal, legal seen and legal unseen items are shown separately for each experimental condition in Table 2.3. Data were subjected to separate one-way repeated measures ANOVA with stimulus type (3 levels: legal seen; legal unseen; illegal) as a within-subject factor.

The effect of stimulus type was not significant on mean correct latencies in the 2-vowel positional constraints variant with Latin letters, $F(2, 38) = .84$, $p > .05$, $\eta^2 = .04$. Similarly, no effect of stimulus type was found in the 3-vowel positional constraints with Latin letters, $F(2, 34) = 2.57$, $p > .05$, $\eta^2 = .13$. The effect of stimulus type was significant in the 2-vowel positional constraints variant with Greek letters, $F(2, 36) = 11.05$, $p < .001$, $\eta^2 = .38$. Planned contrasts revealed that correct RTs for illegal items were significantly longer than correct RTs for legal items, $F(1, 18) = 12.13$, $p < .001$. The difference in RTs between legal seen and legal unseen items did not reach significance, $F(1, 18) < 1$. The main effect of stimulus type was also significant for contextual constraints learning with Latin letters, $F(2, 34) = 3.85$, $p = .031$, $\eta^2 = .18$. Legal items were responded to faster than illegal items, $F(1, 17) = 5.17$, $p = .029$, and there was no difference in RTs between legal seen and unseen items, $F(1, 17) < 1$. No effect was found for contextual constraints learning with Greek letters, $F(2, 38) = 0.97$, $p > .05$, $\eta^2 = .05$.

Table 2.3. Mean Correct Test RTs in ms (*SD*s) for Illegal, Legal Unseen and Legal Seen Items in All Experimental Conditions (Experiments 1.1 – 1.3).

Condition	Illegal	Legal Seen	Legal Unseen
PC-Latin-2v	1178.73 (482.93)	1130.33 (452.67)	1123.22 (511.01)
PC-Latin-3v	1330.82 (458.57)	1232.62 (404.69)	1368.64 (516.88)
PC-Greek-2v	1514.25 (514.44)	1307.07 (367.02)	1318.00 (395.91)
CC-Latin-2v	1323.05 (558.25)	1178.79 (557.64)	1186.75 (480.82)
CC-Greek-2v	1348.83 (595.57)	1257.99 (492.68)	1267.14 (560.30)

Note. PC = Positional Constraints, CC = Contextual Constraints, v = vowels

2.3.4.2 Correlational Analyses

Descriptive statistics for WRAT-IV Reading and Spelling measures across all three experimental groups are shown in the left panel of Table 2.4. Pearson's correlation coefficients of the relationship between reading and spelling scores and their significance levels are shown in the right panel of Table 2.4. Inter correlations of medium size were observed in group 2 and 3, consistent with a large body of research showing that word-reading performance is closely associated to word-spelling performance in adults and—even more strongly so—in typically developing children (for a review, see Ehri, 1997). WRAT Reading performance did not correlate with WRAT Spelling performance in group 1. We cannot offer a ready explanation for this unexpected result[5].

Table 2.4. Mean Raw Scores (*SD*s) and Intercorrelations between WRAT-IV Reading and Spelling Ability for each Experimental Group (Experiments 1.1 – 1.3).

	n	WRAT Reading	WRAT Spelling	*r*	*p* values
Group 1	20	60.40 (2.72)	46.70 (3.21)	.04	.852
Group 2	18	59.50 (3.81)	46.56 (3.54)	.44	.067
Group 3	19	60.84 (4.10)	47.11 (4.05)	.58	.009

Note. WRAT = Wide Range Achievement Test

Preliminary screening of the literacy scores revealed small departures from normality. With regard to reading scores, a slight negative skew (skew = -1.26, *SE* = 0.54; *n* = 18) was observed in experimental group 2. Two extreme outliers were identified and removed from any further analyses (skew = -0.32, *SE* = 0.56). The new mean for WRAT Reading was 60.56 (*SD* = 2.34; *n* = 16). Scores were normally distributed following removal of the two outliers, confirmed by a Shapiro-Wilk test of normality, *p* = .316. With regard to spelling scores, an extreme outlier

[5] Group 1 participants were recruited from the same population of university students as group 2 and 3 participants. We anticipated all participant groups to be broadly normal in terms of their reading and spelling skills. Unfortunately, date of birth was not recorded during the pilot study. Thus, it was not possible to convert raw scores to standard scores.

was removed from experimental group 3. The new mean for WRAT Spelling was 46.61 (*SD* = 3.53; *n* = 18). Since departures from normality were small, no transformation was applied to the scores. Preliminary screening of *d*' scores across all five experimental conditions revealed no major departures from normality. In the 2-vowel positional constraints variant with Latin letters, the distribution was slightly leptokurtic (kurtosis = 2.32, *SE* = 0.99). Two extreme outliers were identified and removed from the analyses, which improved the kurtosis of the distribution (kurtosis = 0.54, *SE* = 1.04). The new mean was 1.07 (*SD* = 0.35; *n* =18). A slight positive skew was also observed for *d*' scores in the positional constraints variant with Greek letters (skew = 1.09, *SE* = 0.52), although the departure from normality was not significant in the Shapiro-Wilk test of normality, $p = .059$. As with literacy scores, *d*' scores were not transformed.

Similar to previous studies that examine the relationship between measures of implicit skill learning and reading ability (e.g., Arciuli & Simpson, 2012; Howard et al., 2006), a series of correlations was performed between *d*' scores and performance in WRAT Reading and Spelling. The full set of correlations performed is summarized in Table 2.5.

Table 2.5. Correlation Table between Measures of Discrimination Accuracy (Experiments 1.1 – 1.3), Reading and Spelling Ability.

IGL task	WRAT Reading	WRAT Spelling
PC-Latin-2v	-.11	-.40
PC-Latin-3v	.54*	.22
PC-Greek-2v	.00	.09
CC-Latin-2v	.03	.17
CC-Greek-2v	.51*	.05

Note. PC = Positional Constraints, CC = Contextual Constraints, v = vowels, FA = False Alarms, WRAT = Wide Range Achievement Test
* $p < .05$

These revealed significant positive correlations between learning of positional constraints (3-vowel variant with Latin letters) and reading ability, *r*(14) = .54, *p* = .032, as well as learning of contextual constraints (2-vowel variant with Greek letters) and reading ability, *r*(16) = .51, *p* = .021. Scatter plots showing significant associations between reading scores and task performance are shown in Figure 2.5.

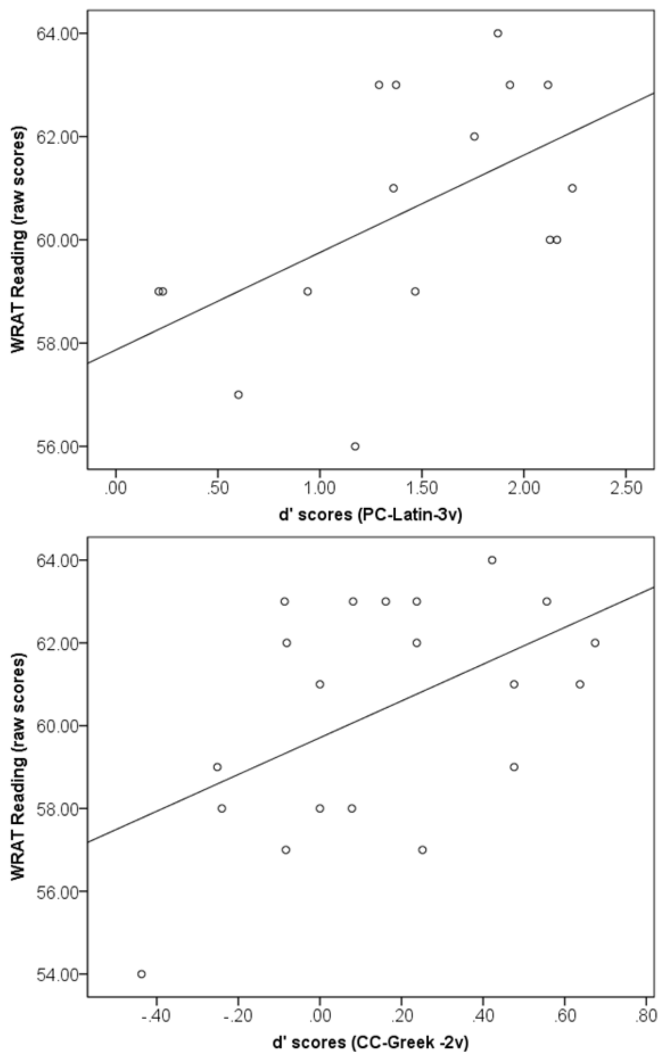


Figure 2.5. Scatter plots showing a positive relationship between (a) performance on the 3-vowel positional constraints learning variant with Latin letters and WRAT Reading performance, (b) performance on the 2-vowel positional constraints learning variant with Greek letters and WRAT Reading performance.

2.4 General Discussion

Previous research has demonstrated that brief listening (Chambers et al., 2003, 2010, 2011; Endress & Mehler, 2010; Onishi et al., 2002) or speaking experience (Dell et al., 2000; Taylor & Houghton, 2005; Warker & Dell, 2006) is sufficient to induce learning of novel phonotactic constraints in the auditory domain. The current study sought to induce similar graphotactic learning by means of a two-phase task adapted from Onishi et al. (2002). Our study focused on two graphotactic constraints that vary in complexity and that are similar to those frequently operating in the English orthography: Positional learning was induced by introducing constraints on the allowable position of letters (e.g., set 1 consonants could only appear in word-initial positions; set 2 consonants could only appear in word-final positions). Sensitivity to contextual constraints was investigated by relating information about letter identity to its neighbouring vowel (e.g., set 1 consonants were always followed by the vowel o, while set 2 consonants were always preceded by the vowel o). Following incidental exposure to pattern-embedding stimuli (phase 1), participants' learning was tested (phase 2) with legality judgments about novel conforming/nonconforming strings/sequences. In line with artificial grammar learning studies (e.g., Johnstone & Shanks, 2001), signal detection measures were used to assess whether sensitivity was significantly above the chance level.

Consistent with our hypothesis, participants, as a group, reliably discriminated between legal and illegal strings in all five experimental conditions. The experiments reported here validate Onishi et al.'s (2002) incidental learning task in the visual domain and extend previous findings with aurally presented phonotactically constrained syllables. Bias *c* analyses confirmed that participants' responses were generally balanced during the test phase (except for performance in the 3-vowel positional constraints variant, where a bias towards saying "no" was observed). Novel graphotactic learning was only partially evident in our latency analyses. Consistent with Onishi et al.'s (2002) demonstration of shorter repetition latencies for legal items, correct classification responses were made more quickly for legal than illegal test strings only in (a) the 2-vowel positional constraints variant with Greek letters and (b) the contextual constraints learning variant with Latin letters. There was no effect of stimulus type on participants' RTs in any of the remaining experimental conditions. This is not necessarily a surprising finding, as decision making reaction times may be less sensitive to newly acquired knowledge than repetition RTs in the speech production domain.

Comparisons between *d*' values across the different experimental conditions revealed that sensitivity to stimuli embedding constraints on position was systematically stronger than sensitivity to stimuli embedding context-based constraints. This clear and logical effect of level of complexity on learning performance helps to validate this type of paradigm for inducing fast graphotactic learning. Our finding of superior positional over contextual learning is in line with Warker and Dell's (2006) demonstration that, under laboratory conditions, context-based constraints are (a) harder to learn (i.e., learning requires a longer period of familiarization with critical stimuli) and (b) are learnt less strongly than positional constraints by human participants and connectionist implicit learning models. The finding is also consistent with developmental studies on the acquisition of language-wide graphotactic constraints—reviewed in more detail in the next chapter. Warker and Dell's (2006) demonstration of reliable context-based learning following longer training raises another interesting question addressed in Chapter 3. Does learning performance become stronger with increasing exposure?

Learning of both types of constraints was reliable with linguistic (i.e., pronounceable CVC strings) as well as nonlinguistic sequences. Interestingly, the difference between directly comparable variants with Latin and Greek letters was not statistically significant. Our finding supports the theoretical claim that the ability to detect statistical regularities in written language relies on the same domain-general learning mechanism that operates across different domains and even across different modalities of knowledge (e.g., human speech; Saffran et al., 1996a; musical tones, Saffran et al., 1999; visual shapes in the temporal domain; Kirkham et al., 2002; Fisher & Aslin, 2002; visual shapes in the spatial domain; Fiser & Aslin, 2001). We also investigated the extent to which learning in the positional constraints condition was relatively tied to the frames presented during exposure. To this aim, we manipulated between-subjects the frequency of token repetitions with the variety of tokens available for learning—while keeping the number of exposure trials identical. While we hypothesized that increasing the variety of CVC sequences comprising the learning tokens (3-vowel positional constraints learning condition) would result in better discrimination ability at test relative to a condition embedding a smaller corpus of tokens (2-vowel positional constraints learning condition) there was no statistical difference between performance in these conditions.

2.4.1 Sensitivity to Seen Versus Unseen Items

To investigate the extent to which participants' performance was driven by sensitivity to legal unseen (i.e., generalization) items and/or sensitivity to items presented during the exposure phase of the experiment, a set of analyses compared participants' *d*' values for each subset of items against chance levels. A comparison between *d*' values for seen and unseen items was also performed in an attempt to replicate Onishi et al.'s (2002) finding that previously seen legal items were not repeated more quickly than legal unseen items. In our study, there was no difference between sensitivity to seen and unseen items in (a) both nonlinguistic variants (i.e., in the positional and contextual constraints learning variants with Greek letters), and (b) in the 2-vowel positional constraints learning variants with Latin letters. However, learning in the 3-vowel positional constraints learning variant with Latin letters and in the 2-vowel contextual constraints learning variant with Latin letters was driven to a greater extent by participants' successful recognition of previously seen items. Thus, there was mixed evidence regarding the question of whether participants' discrimination ability benefitted from the presence of legal seen items.

More critically, while participants' classification performance of legal unseen (as well as legal seen) items was reliably better than expected by chance in all three positional constraints learning variants, a different picture emerged with regards to performance in the two contextual constraints learning variants. In the linguistic variant, only seen items were discriminated reliably, and more surprisingly, in the nonlinguistic variant, neither subset of items was classified with better than chance accuracy. These findings suggest that in the former experimental condition, learning was only driven by participants' successful recognition of legal seen items, while in the latter experimental condition learning was weak and unreliable when sensitivity to each item subtype was considered separately. This important limitation of our study is further discussed at the end of the chapter.

2.4.2 Correlational Analyses

Further to examining the learning mechanisms underlying the acquisition of novel graphotactic constraints, our study was also designed to investigate whether statistical learning skill is related to general reading and spelling ability. Significant correlations between statistical learning performance and measures of literacy-related skills would bolster the theoretical claim that such mechanisms are in the service of written language acquisition. For example, previous research has shown that implicit/statistical learning performance correlates with general reading and spelling ability in typical developing children (e.g., Arciuli & Simpson, 2012) and dyslexic adults (e.g., Howard et al., 2006). Performance measured by the *d*' estimate on the 3-vowel positional constraint variant with Latin letters was positively associated with reading performance. There was also a significant positive relationship between *d*' scores and reading scores in the contextual constraints learning condition with Greek letters, but this may have been spurious given the low performance levels found in this condition. There was no association between performance in the 3-vowel positional constraints variant with Latin letters—or in fact, any other variant—and spelling skill. The association between literacy and statistical learning skill is likely to be relatively weak (e.g., only modest correlations were obtained in Arciuli & Simpson's 2012 study). Therefore, lack of statistical power may be one explanation of the failure to obtain further significant correlations. Another potential explanation for the, by and large, null findings of our study is that WRAT Reading and Spelling may have not been sensitive enough to pick up variation in reading or spelling performance in our group of highly proficient (university student) readers and spellers. These issues were addressed in a follow-up study reported in Chapter 3.

2.4.3 Limitations and Conclusions

Five experiments addressed the question of whether participants are able to learn graphotactic constraints under incidental conditions and apply them to discriminate between legal unseen (as well as legal seen) and illegal items. Although overall discrimination performance was reliable under all experimental conditions, performance in the contextual constraints variant with Latin letters was mainly driven by participants' sensitivity to items that were identical to those presented during the exposure phase of the experiment. More puzzlingly, neither legal unseen nor legal seen items were classified with better than chance accuracy in the contextual constraints variant with Greek letter. This is an important limitation of our pilot graphotactic learning demonstrations and raises the important question of whether context-based learning is only reliable when familiar items are presented. This was addressed in the follow-up study reported in Chapter 3 by removing the subset of legal seen items from all experimental conditions (that is, balancing the number of legal and illegal items in line with previous artificial grammar learning studies; see also, Chambers et al., 2010) and investigating performance when only novel items (legal, illegal) were presented.

Our experimental investigations may have lacked statistical power. Further to this obvious issue, our results may be also limited by the nature of our mixed design. With the exception of group 2, participants in the other groups were administered two incidental learning tasks in a fixed order. Our design ensured that directly comparable versions were always administered between-subjects (e.g., participants were never administered two positional constraints learning variants). Nevertheless, it is not possible to rule out that participants may have handled the tasks that were administered second in a somewhat less implicit manner (see Reber et al., 1991 for a similar concern). For example, participants in group 1 may have employed a different (e.g., searching) strategy during the exposure phase of the 2-vowel contextual constraints learning task with Latin letters. Likewise, performance in the exposure phase of the 2-vowel positional constraints variant with Greek letters may have been influenced by the prior administration of the 2-vowel contextual constraints variant with Latin letters. To avoid contamination of the learning effects from the administration of previous tasks, a full between-subject design was employed in all of our subsequent studies. Participants tested under different experimental conditions were always naïve to the purpose of the study.

Chapter 3

Statistical Learning of Novel Graphotactic Constraints in Children and Adults[6]

3.1 Introduction

There is a growing interest in the development of orthographic knowledge and its contribution to skilled reading and spelling. Learners become sensitive to general properties of their orthography—such as the frequency of occurrence of individual letters/letter sequences in print and the legality of different spelling patterns—at an earlier developmental point than previously thought (Cassar & Treiman, 1997; Ferreiro & Teberosky, 1982; Hayes et al., 2006; Pacton et al., 2001; Treiman, 1993). These findings challenge the longstanding view that phonological information is the only resource available to beginning spellers (e.g., Frith, 1985; Gentry, 1982). In fact, recent theorizing about spelling development emphasizes children's early insights into different types and levels of linguistic knowledge, including knowledge of simple orthographic and morphological conventions (Deacon et al., 2008; Treiman & Bourassa, 2000).

Given that some information about legal orthographic forms is available to children before or as soon as literacy instruction begins, learning must be, to some extent, incidental in nature. Accordingly, several authors have argued that children's pattern extraction skills rely on implicit or statistical learning processes (e.g., Kessler, 2009; Pollo et al., 2007; Treiman &

---

[6]A modified version of Chapter 3 is in press at the Journal of Experimental Child Psychology: Samara, A., & Caravolas, M. (in press). Statistical Learning of Novel Graphotactic Constraints in Children and Adults. Journal of Experimental Child Psychology.

Kessler, 2013). However, to date, little empirical research has examined the mechanisms underpinning the acquisition of orthographic knowledge. Additionally, very few studies have investigated learning of written language patterns under well-controlled experimental conditions which allow manipulations of the type and amount of exposure given to the observer. The goal of this study was to address these issues in a group of 7-year-old children in comparison with a group of university students. In particular, we investigated whether sensitivity to novel orthographic constraints similar to those found in the English orthography develop under incidental learning conditions. Can learning be induced in 7-year-old typically developing children as well as in adults?

3.1.1 Development of Sensitivity to Orthographic Structure

Much research has been carried out on the developmental course of children's sensitivity to orthographic structure. In a seminal study, Treiman (1993) examined a large corpus of naturalistic writing data produced by a group of U.S. English-speaking children at the end of kindergarten/beginning of first grade. Detailed analyses of children's misspellings revealed that even the youngest children actively used some knowledge of simple spelling conventions in their own written productions, such as the positional constraint that "*ck* is not a legal onset in English". Moreover, first-graders rarely committed errors that were inconsistent with orthographic constraints on permissible letter doublets (e.g., *xx* or *kk*). Using a more experimental approach, Treiman (1993) further demonstrated that school beginners are sensitive to statistical probabilities in print in simple nonword judgment tasks (e.g., which one looks more like a real word, *moyl* or *moil*?).

The standard version of Treiman's (1993) now widely used orthographic constraints task presents participants with the oral pronunciation of a nonword and asks them to choose between two alternative written spellings, only one of which conforms to patterns in the English orthography. In one of the first experimental studies of this sort, Cassar and Treiman (1997) assessed children's sensitivity to three untaught orthographic constraints on allowable consonant and vowel doublets. Grade 1 children reliably preferred nonwords containing doublets in allowable positions (in word middles but not at word beginnings), as well as doublets composed of letters that are allowed to double (e.g., *ll* but not *xx*). Sensitivity to both types of constraints increased as a function of age. However, more complex relationships, such as the influence of phonological context on consonantal doubling, had an effect only in more advanced spellers' choices (Grade 6). Cassar and Treiman's (1997) key findings have been replicated in languages other than English (French: Pacton et al., 2001; Finnish: Lehtonen & Bryant, 2005). Moreover, Pacton et al. (2001) demonstrated in recognition (judgment) as well as fragment-completion production tasks that French-speaking first graders have a preference for nonwords embedding letters that are frequently doubled in their orthography (e.g., *illaro* > *ivvaro*), and for stimuli containing a doublet in a medial legal position than stimuli where the doublet is illegally situated (e.g., *nnulor* > *nullor*). This suggests that children's knowledge of commonly encountered orthographic constraints may generalize, to some extent, to their own written productions.

Young learners' appreciation of positional constraints, such as those governing the legality of consonantal doublets, provides them with but one cue to correct spelling. Taking surrounding context into account further facilitates the process of translating speech to print, in inconsistent (Alegria & Mousty, 1996; Kessler & Treiman, 2001; Pacton, Fayol, & Perruchet, 2002, 2005) as well as more consistent orthographies, as Slovak (Caravolas & Mikulajová, 2008). Several studies in English have demonstrated that children and adults take advantage of contextually conditioned phonological patterns in order to spell consonants (Treiman & Kessler, 2006; Treiman et al., 2002) or vowels (Hayes et al., 2006), although this type of knowledge seems to develop gradually, with some contexts being learned more quickly than others. For example, Treiman and Kessler (2006) showed that, by 7th grade level, children's spellings reliably reflected sensitivity to contingencies between vowel-coda contexts, such as that the vowel /ɛ/ is commonly spelled as *ea* when followed by the coda /d/—as in *head*—but not when followed by other coda consonants—as in *hen* or *help*. In the same study, children's spellings demonstrated reliable sensitivity already by 4th grade level to onset-vowel contexts, such as that the vowel /ɝ/ is usually spelled as *or* when preceded by the onset /w/—as in *word*—whereas several other spellings follow when the onset consonants are different, such as /b/ or /k/ in *bird*, *curd*, etc.

In the studies described above, surrounding context was manipulated through phonological cues. However, of particular relevance to the current study, Hayes et al. (2006)

investigated children's sensitivity to contextually conditioned spelling patterns that were better described in graphotactic rather than phonological terms. For example, an alternative account for the "doubling" effects in the context of long versus short vowel phonemes in Cassar and Treiman's study (1997) is that consonantal coda spellings are extended (e.g., doubled) when preceded by vowels spelled with a single letter (e.g., *Jeff*), while vowels spelled with two letters do not require a similar extension of the following consonant (e.g., *deaf*). Thus, the number of letters in the spelling of the short vowel phoneme /ɛ/ determines the number of letters in the coda phoneme—a graphotactic effect that is independent of phonological information. Conforming with the pattern described above, children as young as Grade 2 preferred spellings with coda doublets (e.g., *ff*) when preceded by single vowel spellings, but single-letter codas (e.g., *f*) when preceded by two-letter vowels. Even more interestingly, children's spelling of nonwords was found to be consistent with the graphotactic rather than the phonological rule. That is, spelling of codas depended on the number of preceding vowel letters (one vs. two) regardless of the length of the previous vowel phoneme.

In summary, the above literature provides evidence of children's sensitivity to certain statistical properties of the orthography in the early stages of literacy acquisition. This is in contrast to the long standing notion that nonphonological information, such as knowledge of acceptable letter sequences, has an influence on spelling productions "late" in development (e.g., Frith, 1985; Gentry, 1982). Undoubtedly, children's initial orthographic sensitivity is limited and imperfect (Pacton et al., 2002). Thus, an aim of the present study was to investigate differences in the learning of simpler, positionally conditioned versus more complex, contextually conditioned spelling patterns. Studies, such as those by Treiman and colleagues, demonstrate that knowledge of more complex patterns—embodied as context-based regularities in the orthography—takes longer to develop and reaches adult-like levels only in more advanced spellers (Hayes et al., 2006; Treiman & Kessler, 2006). However, none of these studies directly assesses how learning of simple or more complex orthographic patterns occurs. Are relatively few exposures to wordlike strings embedding the constraints sufficient for learning? Does pattern complexity moderate learnability?

3.2 Experiments 2.1 – 2.8

3.2.1 Rationale

Using the incidental graphotactic learning task introduced in Chapter 2 and a between-subject design manipulating learner group, i.e., skilled versus developing learners, and graphotactic constraint complexity, i.e., positional versus contextual constraints, we investigated whether novel graphotactic learning can be induced under similar brief incidental conditions in typically developing 7-year-olds and adults. We also examined the effect of the amount of exposure (short vs. long exposure) on learning performance. Previous studies have shown that longer exposure enhances incidental pattern learning in a wide range of domains (e.g., sequential structure learning: Gaillard, Destrebecqz, Michiels, & Cleeremans, 2009; harmonic pattern learning: Jonaitis & Saffran, 2009; visual statistical learning: Turk-Browne, Jungé, & Scholl, 2005). For example, Saffran et al. (1997) showed that longer exposure to a continuous speech stream had a direct benefit to participants' ability to detect word boundaries. Using the same task, Evans, Saffran, and Robe-Torres (2009) demonstrated that the statistical learning abilities of children with specific language impairment vary as a function of the amount of exposure given. Therefore, we also anticipated that increasing the amount of exposure would enhance learning.

As in Chapter 2, we also examined the relationship between learning performance and literacy skill in a group of skilled adult learners. Going beyond our previous study—where reading and spelling abilities were measured by the WRAT Reading and Spelling subtests—we considered adults' performance in three additional measures: two standardized measures of word fluency and nonword fluency (Test of Word Reading Efficiency; Torgesen, Wagner, & Rashotte, 1999), and an experimental measure of exception word reading (following Jared, 2002). Furthermore, given that our previous study showed only a few modest positive correlations between reading/spelling and *d*' scores, we extended our investigations to examine whether a stronger relationship is evident during development. To this aim, we assessed children's reading/spelling abilities by means of three experimental, age-appropriate literacy measures: (a) a one-minute word reading task, (b) a picture-word matching (silent reading) task and (c) a nonword reading task. Based on the findings of previous studies (reviewed in detail in section 2.1.4), we hypothesized that better incidental graphotactic learning performance would be associated with better reading/spelling performance. Moreover, we expected that correlations would be stronger among typically developing children, as participants in this population presumably represent a broader range of reading and spelling abilities.

3.2.2 Method

3.2.2.1 Participants

Adults. One-hundred-thirteen (29 male; 84 female) undergraduate students with normal or corrected-to-normal visual acuity were recruited through a Psychology department participant panel and received course credit or small monetary compensation for their participation. All were monolingual native English speakers and reported no history of dyslexia or other learning difficulties. They were randomly assigned to one of four experimental conditions formed by crossing the factors of type of constraint (positional vs. contextual) and exposure duration (short vs. long). Sample sizes are reported in Table 3.1. All participants gave their informed consent before the start of the experimental session.

Children. A group of 137 typically developing children at the end of Year 2/beginning of Year 3 (71 male; 66 female) were drawn from a longitudinal cohort of 189 monolingual English speaking children (Caravolas et al., 2012). The sample was recruited from seven classrooms of six primary schools in three counties of Northern England. At the time of the current study, the mean age of the cohort was 88.75 months (7; 5 years; range = 6; 10 – 8; 2). Children were randomly assigned to one of four groups (sample sizes are reported in Table 3.1). Across schools, all children had received the same amount of formal tuition (almost three years) and were being taught by similar literacy methods, according to governmental guidelines. An earlier study (Caravolas et al., 2012) revealed no significant effects of school on the cohort's development of literacy skills.

3.2.2.2 Material

Experiments 2.1 – 2.8 replicated the material used in the 2-vowel positional and contextual constraints learning tasks with Latin letters (see Chapter 2, section: 2.3.2.2.1; stimuli are shown in Appendix A and B), except that only unseen items were shown at test[7]. As explained in Chapter 2, 7 out of the 32 C_1VC_2 sequences used were real words. To investigate whether the pattern of results obtained in the main analyses was unduly influenced by the presence of real words, follow-up analyses, which excluded real words were also carried out.

3.2.2.3 Measures

Adults. All adult participants were assessed on standardized literacy ability measures in addition to the experimental task. To measure reading and spelling achievement, the relevant subtests of the Wide Range Achievement Test-Forth Edition (WRAT-IV; Wilkinson & Robertson, 2006) were used. Participants were also administered a nonspeeded exception word reading task (stimuli from Jared, 2002; Appendix D) and the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). The last comprises of two speeded reading tests, one of which assesses recognition and reading aloud of real words (Sight-Word Efficiency subtest), while the other one assesses participants' ability to use grapheme-to-phoneme correspondences (i.e., phonemically decode) and read nonsense words (Phonemic Decoding Efficiency subtest). In both subtests, participants are instructed to read aloud the words/nonwords presented as fast as possible, skipping any items they cannot read. Performance on each subtest is typically measured by the total number of words read correctly within 45 seconds.

Children. Child participants were administered an extensive test battery, including measures of general ability, reading ability and other related cognitive skills (see Caravolas et al., 2012). The subset of the measures used for the purpose of the present study, their method of administration and scoring is described below.

---

[7]Unfortunately, due to experimenter's error, lists 3/4 were not used as exposure or legal unseen materials with adult participants; only list 1 and list 2 were used for this purpose and were counterbalanced within experimental groups (list 3 always served as illegal test items). List was counterbalanced with child participants such that half of the children in each experimental condition were presented with list 1/2 items during exposure and list 2/1 and 3 items at test; the other half were presented with list 3/4 items during exposure and list 4/3 and 1 items at test.

One minute reading. One hundred and forty four high-frequency words were used to create a graded one-minute reading task. The test comprised 98 monosyllabic, 44 disyllabic, and 2 trisyllabic words selected from among high frequency words in a national lexical corpus. Items were presented in 16-point Helvetica font and were arranged into three columns printed on both sides of an A4 size sheet. This list began with one-letter words and increased in word length and syllabic complexity. The administrator asked each child to read aloud as many words as s/he could in one minute, using his/her index finger to follow along. The task was timed by stopwatch. No feedback was provided. Task performance was measured in terms of the number of words correctly read within 60 seconds.

Picture-word matching. Children were presented with a picture and four possible words next to it, and were required to choose the one that corresponded to the depicted item. Stimuli consisted of 51 coloured drawings of highly imageable objects. Four words were assigned to each picture: (a) the word corresponding to the drawing (*target word*; e.g., cat), (b) a phonological distractor (e.g., hat), (c) a semantic distractor (e.g., dog), and (d) an unrelated distractor (e.g., sea). Children were told that they were going to play a matching game during which they had to determine what they saw in the picture and then pick the word that matched it by making a tick in a box next to the word. There was one demonstration and two practice items. Correct identifications were scored with 1 point, while incorrect choices and responses children left blank were scored with 0. Children were given three minutes to complete the task. Performance was measured by the total number of correct responses.

Nonword reading. Children were asked to read aloud for one minute as quickly and accurately as possible from a list of 144 nonwords. The items were derived from the list of words used in the analogous word reading task. The list began with one-letter nonwords and increased in length and syllable complexity. Items were arranged into three columns printed on both sides of an A4 size sheet using Helvetica 16-point font. The administrator told each child that s/he would show him/her some funny pretend words to read as quickly as possible, using his/her index finger to follow along. The task was timed by stopwatch. No feedback was provided. As in the one minute reading task, performance was measured in terms of the number of nonwords correctly read within 60 seconds.

3.2.2.4 Procedure

Adult participants were tested individually in a quiet, dimly illuminated room. All participants were first administered the incidental graphotactic learning task followed by the literacy-related measures. Testing took place in a single experimental session which took approximately 30 minutes to complete.

Children were tested in a quiet area of the school. All cognitive and literacy-related measures were administered individually in one session with the exception of the picture-word matching task which was administered in a separate group session. The experimental task was administered in a separate session, prior to or after the administration of the cognitive and literacy-related measures.

Incidental graphotactic learning task. The procedure was the same as in experiments 1.1 – 1.3, with the following exceptions: (a) the number of exposure blocks was doubled in the long learning variants. That is, participants were presented with 6 blocks of 48 trials each, for a total of 288 trials (18 repetitions/string); (b) there were no legal seen items at test. Participants were presented with legal unseen items (n = 16) that embedded the constraints or illegal items (n = 16) which violated the constraints; (c) in order to avoid extreme biases at test, all participants were informed that half of the stimuli were compatible with the previously seen set of strings, whereas the other half were not.

The task procedure was almost identical for child participants, with the following exceptions: (a) a brief "pre-exposure" phase was introduced, during which children were asked to practice a few button sequences in order to become familiar with the keyboard; (b) the number of practice trials was increased from 6 to 12; (c) the distraction phase consisted of two simple counting tasks (from 1 to 10 [forward]; from 20 to 1 [backwards]) and an orally presented calculation task involving performing five single-digit oral calculations of the form $x + y = z$ with $1 \leq x, y \leq 3$ and $2 \leq z \leq 6$.

3.2.3 Results

3.2.3.1 Incidental Graphotactic Learning task

Data were visually examined to identify potential participant outliers. One child participant who was at chance in the exposure phase of the task was excluded from the analyses, and data from four additional child participants were dropped from all analyses on the grounds that they repeatedly pressed the same button during the test phase. The final sample consisted of 132 children (positional constraints short, $n = 33$; positional constraints long, $n = 34$; contextual constraints short, $n = 34$; contextual constraints long, $n = 31$). No adult participant was dropped from the incidental graphotactic learning task analyses.

Exposure phase. It was anticipated that, if the colour letter detection task was a valid and effective means of exposing participants to the letter strings, children and adults would perform with high accuracy, and reaction times would remain relatively stable or would decrease over blocks of trials; both measures would thus demonstrate participants' sustained attention on the task. Inspection of all groups' accuracy scores confirmed that performance in the colour detection task was at ceiling in every block (> .98 for adults and > .93 for children). Thus, accuracy data were not further analyzed. In order to investigate whether performance was also stable in terms of reaction times, mean correct trimmed RTs (Figure 3.1) were subjected to two separate mixed ANOVAs with block (for short exposure: 1, 2, 3; for long exposure: 1, 2, 3, 4, 5, 6) as a within-subject variable and group (adults, children) and type of constraint (positonal, contextual) as between-subject variables. Performance did not vary as a function of counterbalanced list; we therefore omit further discussion of this factor.

Reaction times were stable in the short exposure variants, confirmed by a nonsignificant main effect of block, $F(1.89, 222.98) = 1.95$, $p > .05$, $\eta^2 = .02$. Block did not interact with group, $F(1.89, 222.98) = 1.57$, $p > .05$, $\eta^2 = .01$, or type of constraint, $F(1.89, 222.98) = 2.11$, $p > .05$, $\eta^2 = .02$. The three-way interaction between block, group and type of constraint was not significant, $F(1.89, 222.98) = 2.32$, $p > .05$, $\eta^2 = .02$. The main effect of type of constraint was not significant, $F(1, 118) < 1$, neither was the group by type of constraint interaction, $F(1, 118) < 1$. On the other hand, the analyses revealed an unsurprising significant main effect of group, F(1,

118) = 353.99, $p < .001$, $\eta^2 = .75$, with longer RTs for children ($M = 1152.72$, $SE = 25.44$) than adults ($M = 439.88$, $SE = 28.08$).

Some instability in RTs was observed in the long exposure variants, confirmed by a significant main effect of block, $F(3.92, 466.63) = 7.21$, $p < .001$, $\eta^2 = .05$, order 5 trend, $F(1, 119) = 30.12$, $p < .001$. The interaction between group and block was significant, $F(3.92, 466.63) = 7.00$, $p < .001$, $\eta^2 = .05$, due to a much smaller level of variation in adults, $F(3.52, 200.59) = 4.09$, $p = .005$, $\eta^2 = .07$, cubic trend, $F(1, 57) = 6.30$, $p = .015$, than in children, $F(3.91, 250.12) = 8.07$, $p < .001$, $\eta^2 = .11$, order 5 trend, $F(1, 64) = 38.39$, $p < .001$. For example, post hoc paired *t* tests revealed a significant difference between block 2 and block 3 RTs in children, $t(64) = 4.29$, $p < .001$, $d = 0.53$, but not in adults, $t(57) = 0.37$, $p > .05$, $d = 0.05$. Critically, neither group showed a systematic increase in RT; this, in combination with the very high accuracy scores suggested that both participant groups remained focused on task. The main effect of type of constraint was also significant, $F(1, 119) = 9.17$, $p = .003$, $\eta^2 = .02$, and was qualified by a significant group by type of constraint interaction, $F(1 ,119) = 8.90$, $p = .003$, $\eta^2 = .02$. This revealed that child participants were faster in the exposure phase of the positional ($M = 956.43$, $SE = 39.13$) than contextual constraints variant ($M = 1162.40$, $SE = 50.98$), $t(63) = 3.24$, $p = .002$, $d = 0.80$, whereas no significant difference was found for adults, $t(56) = 0.11$, $p > .05$, $d = 0.03$. Importantly, as the task demands for the colour detection task were identical across conditions, we interpret this significant difference to reflect random child-group differences in RTs. As in the previous analyses, there was a large significant main effect of group, $F(1, 119) = 335.25$, $p < .001$, $\eta^2 = .71$, such that children ($M = 1058.95$, $SE = 23.56$) were slower than adults ($M = 431.05$, $SE = 24.92$). No other interaction was significant, all *F*s < 1. In sum, adults performed the letter detection task more quickly and with greater stability in RT; however, all groups perfomed with very high levels of accuracy, confirming their sustained attention during the exposure manipulation.

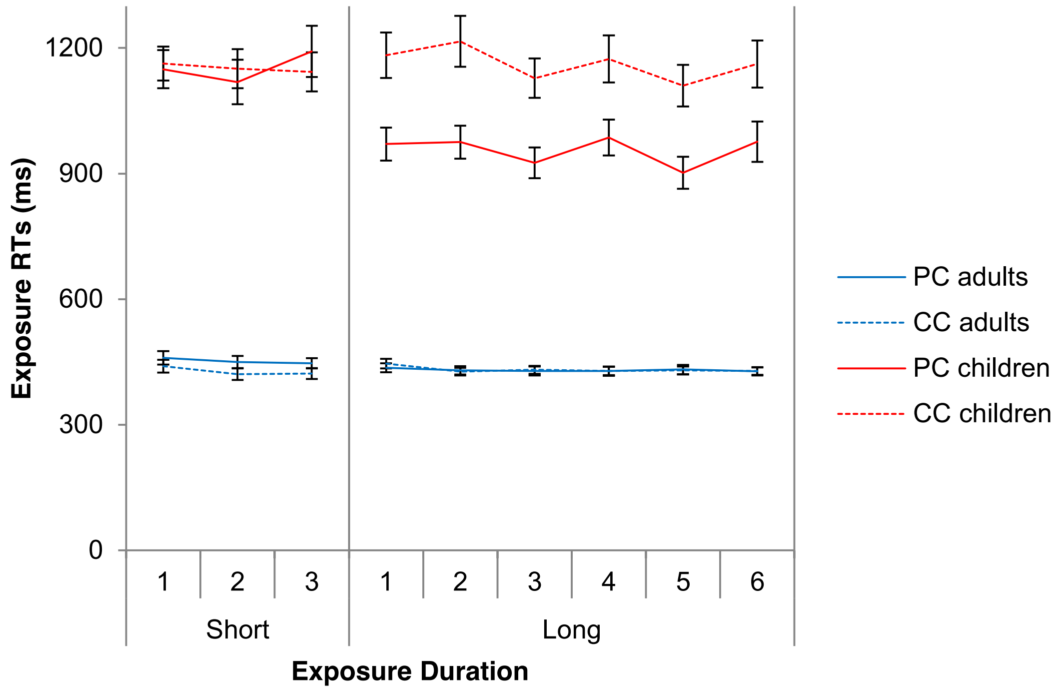


Figure 3.1. Mean correct exposure RTs (ms) as a function of block plotted separately for each experimental group in the short and long exposure variants of the incidental graphotactic learning task (Experiments 2.1 – 2.8). Error bars represent standard error of the mean.

Discrimination accuracy. All data points were included for analyses of discrimination accuracy in adults' data. For the child data, responses associated with latencies below 300ms were identified as extreme outliers and discarded from discrimination accuracy analyses. This represented only a small number of cases (positional constraints short, $n = 4$; positional constraints long, $n = 6$; contextual constraints short, $n = 4$; contextual constraints long, $n = 4$), which did not affect the qualitative pattern of results. Table 3.1 presents the mean proportion of items endorsed as legal as a function of stimulus type (illegal vs. legal unseen items) for adults and children in each experimental condition. The resulting signal detection theory measures (*d*' and *c* values) are shown in the right panel of Table 3.1.

Table 3.1. Proportion of Items Endorsed as Legal (*SD*s) as a Function of Stimulus Type in Each Experimental Condition and the Resulting Signal Detection Theory Measures (Experiments 2.1 – 2.8).

Condition	N (m; f)	Endorsement rates[a]		SDT measures	
		Illegal (FA)	Legal Unseen (Hits)	*d*'	bias *c*
			Adults		
PC short	27 (8; 19)	.32 (0.13)	.70 (0.13)	1.05 (0.57)	-0.03 (0.26)
CC short	28 (7; 21)	.46 (0.12)	.58 (0.11)	0.30 (0.37)	-0.06 (0.25)
PC long	29 (6; 23)	.26 (0.18)	.65 (0.18)	1.23 (0.80)	0.17 (0.45)
CC long	29 (8; 21)	.49 (0.16)	.57 (0.13)	0.21 (0.37)	-0.07 (0.35)
			Children		
PC short	33 (18; 15)	.32 (0.16)	.55 (0.18)	0.66 (0.62)	0.18 (0.41)
CC short	34 (13; 21)	.40 (0.20)	.44 (0.17)	0.14 (0.47)	0.24 (0.51)
PC long	34 (20; 14)	.28 (0.17)	.59 (0.19)	0.95 (0.77)	0.22 (0.43)
CC long	31 (16; 15)	.44 (0.18)	.48 (0.15)	0.12 (0.45)	0.11 (0.39)

Note. m = male; f= female; FA = False Alarms; SDT = Signal Detection Theory. PC = Positional Constraints. CC = Contextual Constraints.
[a] Adults' endorsement rates were calculated based on all test phase data. Children's endorsement rates were calculated based on test phase data above 300ms.

Mean *d'* values were subjected to a 3-way ANOVA with group (adults vs. children), type of constraint (positional vs. contextual), and exposure duration (short vs. long) as between-subject variables. Performance did not vary across lists, and thus this factor was not considered further. This analysis revealed a significant main effect of group, $F(1, 237) = 9.53$, $p = .002$, $\eta^2 = .03$, with higher *d'* values for adults ($M = 0.70$, $SE = 0.05$) than children ($M = 0.47$, $SE = 0.05$), and a highly significant main effect of type of constraint, $F(1, 237) = 110.96$, $p < .001$, $\eta^2 = .30$, with higher *d'* values for positional constraints ($M = 0.97$, $SE = 0.05$) than contextual constraints learning ($M = 0.19$, $SE = 0.05$). Group and type of constraint did not interact, $F(1, 237) = 2.09$, $p > .05$, $\eta^2 = .01$. There was no effect of exposure duration, $F(1, 237) = 1.56$, $p > .05$, $\eta^2 = .00$, exposure duration by group interaction, $F(1, 237) < 1$, or group by type of constraint by exposure duration interaction, $F(1, 237) < 1$. The type of constraint by exposure duration interaction was marginally significant, $F(1, 237) = 4.01$, $p = .046$, $\eta^2 = .01$. Simple effect analyses revealed a trend for better performance in the positional constraints long exposure ($M = 1.08$, $SE = 0.10$) than the short exposure condition ($M = 0.83$, $SE = 0.08$); however, the Bonferroni corrected independent t test failed to reach significance, $t(121) = 1.93$, $p = .056$, $d = 0.35$. For contextual constraints learning, there was no difference between learning following short ($M = 0.21$, $SE = 0.05$) and long exposure ($M = 0.16$, $SE = 0.05$), $t(120) = 0.65$, $p > .05$, $d = 0.12$.

One-sample t tests confirmed that children and adults reliably classified more strings correctly than would be expected by chance in the positional constraints condition (children: $d' = 0.81$, $SE = 0.09$; $d' = 1.14$, $SE = 0.09$), $t_1(66) = 9.32$, $p < .001$; $t_2(55) = 12.21$, $p < .001$, as well as in the contextual constraints condition (children: $d' = 0.13$, $SE = 0.06$; adults: $d' = 0.25$, $SE = 0.05$), $t_1(64) = 2.35$, $p = .022$; $t_2(56) = 5.15$, $p < .001$. The corresponding measures of effect size indicated that the difference was large for positional constraints learning ($d_1 = 1.14$, $d_2 = 1.63$) and small to medium for contextual constraints learning ($d_1 = 0.29$, $d_2 = 0.68$).

Similar to d' values, c values were subjected to a 3-way ANOVA with group (adults vs. children), type of constraint (positional vs. contextual), and duration of exposure (short vs. long) as between-subject variables. There was a main effect of group, $F(1, 237) = 13.24$, $p < .001$, $\eta^2 = .05$, such that children ($M = 0.19$, $SE = 0.04$) tended to show a larger response bias than adults ($M = 0.00$, $SE = 0.04$), but no effect of type of constraint, $F(1, 237) = 2.34$, $p > .05$, $\eta^2 = .01$, or exposure duration, $F(1, 237) < 1$. There was no interaction of group by type of constraint, $F(1, 237) = 1.06$, $p > .05$, $\eta^2 = .00$, group by exposure duration, $F(1, 237) = 2.02$, $p > .05$, $\eta^2 = .01$, type of constraint by exposure duration, $F(1, 237) = 3.58$, $p > .05$, $\eta^2 = .01$, or type of constraint by exposure duration by group, $F(1, 237) < 1$. The detection theory analysis revealed a significant bias on the part of child participants towards responding “no”, $t(131) = 4.99$, $p < .001$, $d = 0.43$, and no response bias among adult participants, $t(112) = 0.13$, $p > .05$, $d = 0.01$.

Latency analyses. Trimmed data accounted for less than 6 per cent of all correct test-phase responses in adult participants (positional constraints short, 3.38%; contextual constraints short, 4.45%; positional constraints long, 5.56%; contextual constraints long, 4.07%) and less than 7 per cent in child participants (positional constraints short, 5.09%; contextual constraints short, 6.17%; positional constraints long, 5.10%; contextual constraints long, 5.14%). Adults’ and children’s mean correct trimmed RTs for legal and illegal items are shown separately for each experimental condition in Table 3.2. These were subjected to a three-way ANOVA with group (children, adults), type of constraint (positional, contextual), duration of exposure (short, long) as between-subject variables, and legality (legal, illegal) as a within-subject variable.

Table 3.2. Adults' and Children's Mean Correct Test RTs in ms (*SD*s) for Legal and Illegal Items in All Experimental Conditions (Experiments 2.1 – 2.8).

Condition	Illegal	Legal Unseen
	Adults	
PC short	1508.49 (470.91)	1303.04 (459.46)
CC short	1701.55 (546.64)	1591.24 (534.99)
PC long	1154.37 (431.19)	1028.42 (323.57)
CC long	1547.31 (596.77)	1404.74 (668.11)
	Children	
PC short	1600.39 (568.04)	1695.62 (854.11)
CC short	1710.58 (581.17)	1677.01 (579.09)
PC long	1452.75 (386.30)	1470.09 (486.50)
CC long	1465.54 (528.89)	1409.13 (433.72)

Note. PC = Positional Constraints; CC = Contextual Constraints.

All four main effects were significant. There was a main effect of legality, $F(1, 236) = 7.17$, $p = .008$, $\eta^2 = .03$, a main effect of group, $F(1, 236) = 5.75$, $p = .017$, $\eta^2 = .02$, and a main effect of type of constraint, $F(1, 236) = 6.24$, $p = .013$, $\eta^2 =.02$. Effects of exposure duration were also observed, $F(1, 236) = 12.83$, $p < .001$, $\eta^2 = .05$, with longer RTs for test strings in the short ($M = 1598.49$, $SE = 46.02$) than long exposure condition ($M = 1366.54$, $SE= 45.55$). The group by legality interaction was significant (Figure 3.2), $F(1, 236) = 8.37$, $p = .004$, $\eta^2 = .03$, such that an RT advantage for legal over illegal items was found for adults, $t(112) = 4.75$, $p < .001$, $d = 0.45$, but not children, $t(130) = 0.16$, $p > .05$, $d = 0.01$. Furthermore, the interaction indicated that children's RTs for legal items were significantly slower than adults' legal RTs, $t(242) = 3.12$, $p = .002$, $d = 0.40$. RTs for illegal items did not differ between groups, $t(242) = 1.20$, $p > .05$, $d = 0.15$.

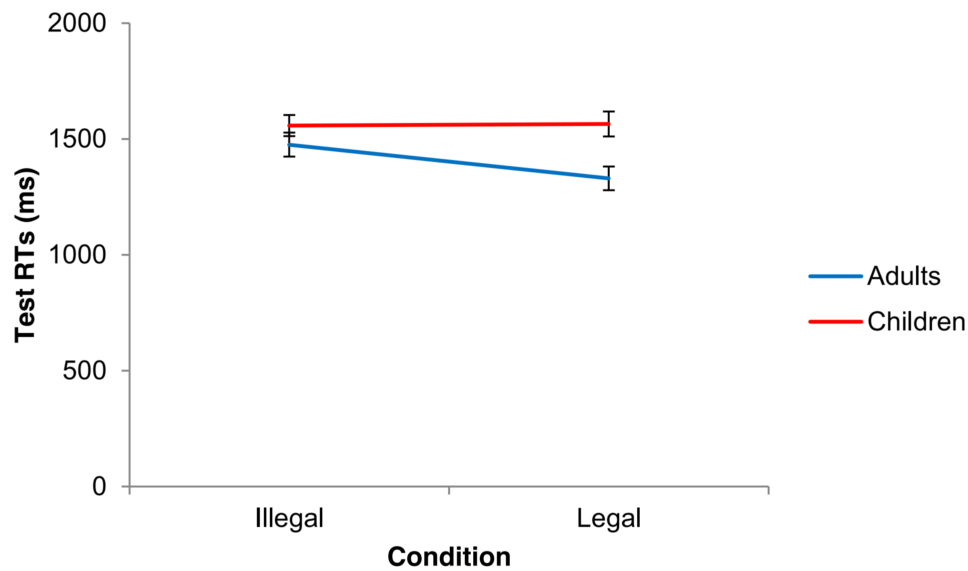


Figure 3.2. Group x Legality interaction (Experiments 2.1 – 2.8). Mean correct RTs (ms) as a function of legality plotted separately for adult and child participants. Error bars represent standard error of the mean.

A group by type of constraint interaction was also observed, $F(1, 236) = 5.43$, $p = .021$, $\eta^2 = .02$. As illustrated in Figure 3.3, adults responded significantly faster to items in the positional constraints than the contextual constraints learning condition, $t(105.02) = 3.27$, $p = .001$, $d = 0.62$, whereas children's RTs did not differ reliably in the two conditions, $t(129) = 0.39$, $p > .05$, $d = 0.07$. What is more, adults were faster than children in the positional constraints learning condition, $t(121) = 3.30$, $p = .001$, $d = 0.60$, but not in the contextual constraints learning condition, $t(119) = 0.13$, $p > .05$, $d = 0.02$. The three-way interaction between group, type of constraint and legality was not significant, $F(1, 236) = 1.79$, $p > .05$, $\eta^2 = .01$, neither was any other interaction in the omnibus analyses (all Fs < 1).

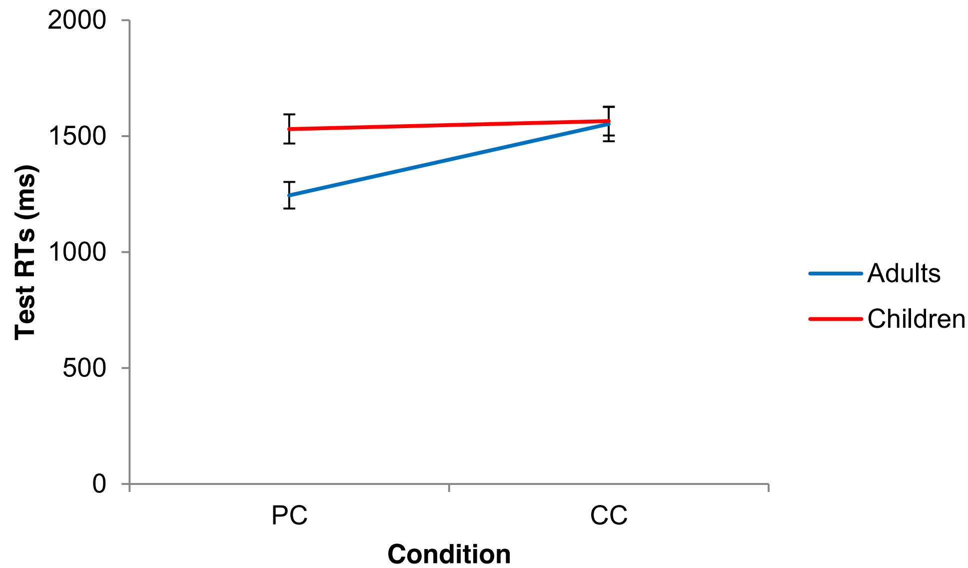


Figure 3.3. Group x Type of Constraint interaction (Experiments 2.1 – 2.8). Mean correct RTs (ms) as a function of type of constraint (positional constraints, contextual constraints) plotted separately for adult and child participants. Error bars represent standard error of the mean.

Analyses excluding lexical items. As mentioned in the Method, although we aimed to construct lists of nonword stimuli for the learning tasks, this was not fully possible while observing the other critical stimulus constraints; consequently, a few real words were included in some of the learning lists. To explore whether the patterns of results obtained in the previous analyses were driven by the presence of real words (ranging from zero to four words per list), all lexical items were removed from the lists, and the discrimination accuracy, bias *c*, and RT analyses were repeated. Removing real words from the lists resulted in lists comprising 12, 13, 15, 16 stimuli (for lists 1, 2, 3, and 4, respectively) for positional constraints learning, and 14, 15, 15 and 12 stimuli (for lists 1, 2, 3, and 4, respectively) for contextual constraints learning. To clarify, bearing in mind that participants were only ever tested on two lists (see Appendices A and B), participants who were exposed to list 1 in positional constraints learning, for example, were tested on 13 legal nonword items (list 2) and 15 illegal nonword items (list 3).These follow-up analyses yielded the same patterns of results in all but one outcome on discrimination accuracy.

Table 3.3 shows the proportion of items endorsed as legal in each experimental condition and the resulting signal detection theory measures. As in the previous analyses, neither the main effect of list nor any interaction with this factor was significant (all *p*s > .05).

Table 3.3. Proportion of Items Endorsed as Legal (*SD*s) as a Function of Stimulus Type in Each Experimental Condition and the Resulting Signal Detection Theory Measures (Analyses Excluding Lexical Items, Experiments 2.1 – 2.8).

Condition	N (m; f)	Endorsement rates[a]		SDT measures	
		Illegal (FA)	Legal Unseen (Hits)	*d*’	bias *c*
		Adults			
PC short	27 (8; 19)	.32 (0.13)	.73 (0.15)	1.18 (0.63)	-0.09 (0.28)
CC short	28 (7; 21)	.48 (0.12)	.58 (0.12)	0.26 (0.41)	-0.08 (0.26)
PC long	29 (6; 23)	.24 (0.18)	.66 (0.21)	1.31 (0.82)	0.17 (0.51)
CC long	29 (8; 21)	.51 (0.17)	.58 (0.14)	0.18 (0.39)	-0.10 (0.36)
		Children			
PC short	33 (18; 15)	.33 (0.17)	.56 (0.19)	0.70 (0.64)	0.17 (0.47)
CC short	34 (13; 21)	.40 (0.20)	.43 (0.17)	0.12 (0.49)	0.26 (0.53)
PC long	34 (20; 14)	.28 (0.17)	.61 (0.20)	0.99 (0.81)	0.19 (0.46)
CC long	31 (16; 15)	.44 (0.19)	.48 (0.17)	0.14 (0.48)	0.12 (0.43)

Note. m = male; f= female; FA = False Alarms; SDT = Signal Detection Theory. PC = Positional Constraints. CC = Contextual Constraints.
[a] Adults’ endorsement rates were calculated based on all test phase data. Children’s endorsement rates were calculated based on test phase data above 300ms.

A comparison of the data in Table 3.1 and Table 3.3 reveals broadly similar patterns between groups and conditions, suggesting that, overall, the inclusion of real words did not qualitatively affect performance. However, a small systematic difference can be observed in the corresponding *d'* values, such that exclusion of real words seemed to strengthen legality sensitivity for positional constraint learning, while reducing it slightly for contextual constraints learning (except between the contextual constraints long condition estimates in children). Statistical analysis of discrimination accuracy performance with real words removed confirmed these observations, replicating the previous finding of a main effect of group, $F(1, 237) = 9.82$, $p = .002$, $\eta^2 = .03$, and a main effect of type of constraint, $F(1, 237) = 123.57$, $p < .001$, $\eta^2 = .33$. However, a small significant group by type of constraint interaction, $F(1, 237) = 3.97$, $p = .047$, $\eta^2 = .01$, was now obtained. Simple effect analyses demonstrated that, relative to children ($M = 0.84$, $SE = 0.09$), adults ($M = 1.25$, $SE = 0.10$) showed an advantage in discrimination accuracy for positional learning, $t(121) = 3.00$, $p = .003$, $d = 0.54$, but not for contextual constraint learning, $t(120) = 1.10$, $p > .05$, $d = 0.20$ (adults: $M = 0.22$, $SE = 0.05$; children: $M = 0.13$, $SE = 0.06$). As in the main analyses, discrimination accuracy was not significantly affected by duration of exposure, $F(1, 237) = 1.48$, $p > .05$, $\eta^2 = .00$, and exposure duration did not interact with group , $F(1, 237) <1$. Also, the group by type of constraint by exposure interaction was not significant, $F(1, 237) <1$. Furthermore, the marginal trend for type of constraint by exposure interaction was no longer significant, $F(1, 237) = 2.21$, $p > .05$, $\eta^2 = .01$.

One-sample *t* tests confirmed both child and adult participants reliably classified more strings correctly than would be expected by chance in the positional constraints learning condition, $t_1(66) = 9.34$, $p < .001$; $t_2(55) = 12.70$, $p < .001$. Performance was also above chance in the contextual constraints learning condition, for child participants, $t(64) = 2.19$, $p = .033$, and adult participants, $t(56) = 4.16$, $p < .001$. The corresponding measures of effect size were large for both children's and adults' performance in the positional constraints learning condition ($d_1 = 1.14$, $d_2 = 1.70$) and small to medium in contextual constraints learning ($d_1 = 0.27$, $d_2 = 0.55$).

The mean *c* values (Table 3.3) were subjected to a 3-way ANOVA as was done for the full data set. We observed a significant main effect of group, $F(1, 237) = 14.74$, $p < .001$, $\eta^2 = .06$, such that children ($M = 0.19$, $SE = 0.04$) tended to show a larger response bias than adults ($M = -0.03$, $SE = 0.04$). As in the previous analysis, there was no significant effect of type of constraint, $F(1, 237) = 1.26$, $p > .05$, $\eta^2 = .00$, or exposure duration, $F(1, 237) <1$, on mean *c* values. There was no interaction of group by type of constraint, $F(1, 237) = 1.74$, $p > .05$, $\eta^2 = .01$, group by exposure duration, $F(1, 237) = 2.78$, $p > .05$, $\eta^2 = .01$, or type of constraint by exposure by group, $F(1, 237) < 1$. The ANOVA revealed a significant interaction of type of constraint by exposure, $F(1, 237) = 4.01$, $p = .046$, $\eta^2 = .02$. The difference between *c* values for positional constraints ($M = 0.18$, $SE = 0.05$) versus contextual constraints learning following long exposure ($M = 0.01$, $SE = 0.06$), $t(121) = 2.09$, $p = .039$, $d = 0.38$, was not significant after Bonferroni correction. There was no difference between *c* values in the two conditions following short exposure, $t(120) = 0.66$, $p > .05$, $d = 0.12$. In summary, consistent with the analysis which included lexical items, *c* values only differed as a function of group. One sample *t* tests replicated the finding of children's—but not adults'—bias towards responding "no", $t_1(131) = 4.55$, $p < .001$, $d = 0.40$; $t_2(112) = 0.68$, $p > .05$, $d = 0.06$. Adults and children's RTs as a function of legality are shown separately for each experimental condition in Table 3.4.

Table 3.4. Adults' and Children's Mean Correct Test RTs in ms (*SD*s) for Legal and Illegal Items in all Experimental Conditions (Analyses Excluding Lexical Items, Experiments 2.1 – 2.8).

Condition	Illegal	Legal Unseen
	Adults	
PC short	1508.03 (470.70)	1314.76 (452.94)
CC short	1751.10 (587.17)	1592.10 (548.49)
PC long	1157.96 (433.37)	1046.08 (329.18)
CC long	1557.96 (621.61)	1417.76 (677.28)
	Children	
PC short	1616.70 (577.45)	1723.38 (867.82)
CC short	1712.68 (577.83)	1717.36 (607.11)
PC long	1418.53 (379.97)	1426.10 (467.07)
CC long	1477.01 (553.11)	1430.11 (436.99)

Note. PC = Positional Constraints; CC = Contextual Constraints.

Replicating the previous set of analyses, we obtained a significant main effect of legality, $F(1, 236) = 6.18$, $p = .014$, $\eta^2 = .02$, a significant main effect of group, $F(1, 236) = 4.98$, $p =$

.027, $\eta^2 = .02$, a significant main effect of type of constraint, $F(1, 236) = 7.52$, $p = .007$, $\eta^2 = .03$, and a significant main effect of exposure duration, $F(1, 236) = 14.47$, $p < .001$, $\eta^2 = .05$, with faster RTs for items in the long ($M = 1366.44$, $SE = 46.33$) than short exposure variants ($M = 1617.01$, $SE = 46.81$). As in the analyses carried out on all test phase items, the group by legality interaction was significant, $F(1, 236) = 9.98$, $p = .002$, $\eta^2 = .04$, showing that adults were faster with legal ($M = 1340.96$, $SE = 51.69$) than illegal items ($M = 1491.23$, $SE = 53.63$), $t(112) = 4.57$, $p < .001$, $d = 0.43$, but not children (legal RTs: $M = 1575.31$, $SE = 55.06$; illegal RTs: $M = 1556.39$, $SE = 46.65$), $t_2(130) = 0.47$, $p > .05$, $d = 0.04$. Children's RTs for legal items were significantly slower than those of adults, $t(242) = 3.07$, $p = .002$, $d = 0.40$, whereas no group difference was observed with regards to RTs for illegal items, $t(242) = 0.92$, $p > .05$, $d = 0.12$. The significant group by type of constraint interaction also replicated, $F(1, 236) = 4.68$, $p = .032$, $\eta^2 = .02$, showing adults' faster RTs to items in the positional ($M = 1260.98$, $SE = 58.33$) versus contextual constraints condition ($M = 1569.88$, $SE = 75.86$), $t_1(104.58) = 3.23$, $p = .002$, $d = 0.61$, but not children's (positional learning: $M = 1522.93$, $SE = 64.07$; contextual learning: $M = 1580.80$, $SE = 64.37$), $t_2(129) = 0.64$, $p > .05$, $d = 0.11$. Children were slower than adults in the positional constraints learning condition, $t(121) = 2.97$, $p = .004$, $d = 0.54$, but not in the contextual constraints learning condition, $t(119) = 0.11$, $p > .05$, $d = 0.02$. None of the remaining interactions was significant, *p*s > .05.

To summarize, analyses based on responses to nonword stimuli only yielded essentially the same pattern of results as those in which some real words were included. The one discrepant result revealed that adults' performance worsened somewhat in contextual learning when real words were removed, bringing their sensitivity scores in line with those of children.

3.2.3.2 Measures

Adults. Data from one participant tested in the positional constraints learning variant and one participant in the contextual constraints learning variant were excluded due to bad quality recordings. The remaining sample consisted of 132 participants (positional constraints learning, $n = 55$; contextual constraints learning, $n = 56$). Preliminary screening of the data revealed a negative skew in WRAT Reading performance (skew = -1.11, $SE = .32$) for participants in the positional constraints learning condition. Scores were log transformed to correct for normality before carrying out further analyses. Accuracy (proportion of words/nonwords read correctly) on the TOWRE subtests was consistently close to ceiling across experimental groups (positional constraints learning: Sight Word Efficiency: $M = .91$, $SD = 0.08$; Phoneme Decoding: $M = .87$, $SD = 0.10$; contextual constraints learning: Sight Word Efficiency: $M = .91$, $SD = 0.08$; Phoneme Decoding: $M = .86$, $SD = 0.09$). Thus, we calculated the number of words/nonwords read correctly per second, a measure of reading rate, which was normally distributed across TOWRE subtests and experimental groups. Exception word reading and WRAT Spelling performance were also normally distributed under both experimental conditions.

Descriptive statistics and the correlations obtained for all measures are shown separately for each experimental group in Table 3.5. Overall, the majority of the inter correlations between the different measures of reading and spelling ability were in the expected direction. For participants in the positional constraints learning condition, WRAT Reading performance was significantly positively correlated with performance in the WRAT Spelling test and nonword reading skill, measured by the TOWRE Phonemic Decoding subtest. Spelling was significantly positively correlated with performance in the exception word reading test and the TOWRE Phonemic Decoding subtest. Higher performance in the exception reading task was significantly positively associated with higher scores in both TOWRE subtests. Finally, there was a significant positive correlation between performance in TOWRE-Phonemic Decoding and Sight-Word Efficiency subtest. On the other hand, *d*' scores for positional constraints learning did not correlate with any measure of reading or spelling ability. Similarly, despite several significant intercorrelations between the different reading and spelling measures, no correlations were observed between *d*' scores in the contextual constraints learning condition and the reading/spelling measures.

Children. Preliminary screening of children's data revealed no major departures from normality. Descriptive statistics and the bivariate correlations obtained for all measures are reported separately for each experimental group in Table 3.6. Similar to the adult correlational analyses, bivariate correlations were computed between all measures separately for each experimental group (Table 3.6). Notable are the high positive inter correlations between literacy measures among children in the positional constraints learning condition, as well as children in the contextual constraints learning condition. However, none of these measures showed an association with performance in the incidental graphotactic learning task.

Table 3.5. Mean Raw Scores (*SD*s) and Inter correlations of the Literacy Measures in Experiments 2.1 – 2.4 (Adult Participants); Correlations with Positional and Contextual Constraints Learning Performance (*d*' Scores) Collapsed across Exposure.

Measure	Range	Mean (*SD*)	1	2	3	4	5	6
		PC learning (*n* = 55)						
1. IGL	-0.16 – 2.75	1.15 (0.70)	—	-.14	.07	.12	.04	-.12
2. WRAT Reading[a]	49.00 – 66.00	60.47 (3.94)		—	.59**	.46**	.01	.33*
3. WRAT Spelling	38.00 – 57.00	47.00 (3.21)			—	.56**	.06	.46**
4. Exception Words[b]	69.00 – 79.00	74.89 (2.42)				—	.32*	.46**
5. TOWRE-SWE[c]	1.49 – 2.68	2.13 (0.22)					—	.50**
6. TOWRE-PDE[c]	0.73 – 1.75	1.28 (0.20)						—
		CC learning (*n* = 56)						
1. IGL	-0.36 – 1.16	0.27 (0.36)	—	-.05	-.16	.13	.18	.01
2. WRAT Reading	51.00 – 67.00	60.23 (3.54)		—	.63**	.72**	.35	.64**
3. WRAT Spelling	37.00 – 55.00	46.57 (3.88)			—	.62**	.13	.48*
4. Exception Words[b,d]	71.00 – 79.00	75.74 (1.91)				—	.38	.65**
5. TOWRE-SWE[c,d]	1.58 – 2.55	2.12 (0.21)					—	.52*
6. TOWRE-PDE[c,d]	0.76 – 1.77	1.26 (0.20)						—

Note. PC = Positional Constraint; IGL = Incidental Graphotactic Learning; WRAT = Wide Range Achievement Test; TOWRE = Test of Word Reading Efficiency; SWE = Sight-Word Efficiency; PDE = Phonemic Decoding Efficiency.
[a]log transformed (raw scores are reported for ease of interpretations). [b]out of 79. [c]words/nonwords read correctly per second. [d]n = 27; The measure was not administered to participants in the contextual constraints long condition.
*p < .05. **p < .01.

Table 3.6. Mean Raw Scores (*SD*s) and Inter correlations of the Literacy Measures in Experiments 2.5 – 2.8 (Child Participants); Correlations with Positional and Contextual Constraints Learning Performance (*d*' Scores) Collapsed across Exposure.

Measure	Range	Mean (*SD*)	1	2	3	4
		PC learning (n = 60)				
1. IGL	-0.81 – 2.68	0.83 (0.73)	—	.19	.10	.13
2. Reading[a]	6.00 – 119.00	73.55 (22.93)		—	.85**	.77**
3. NW Reading[a]	9.00 – 70.00	35.45 (14.68)			—	.71**
4. PWM[b]	6.00 – 61.00	29.13 (10.25)				—
		CC learning (n = 62)				
1. IGL	-0.81 – 1.47	0.15 (0.46)	—	-.11	-.09	-.09
2. Reading[a]	6.00 – 123.00	78.06 (20.77)		—	.83**	.72**
3. NW Reading[a]	2.00 – 72.00	39.94 (18.23)			—	.67**
4. PWM[b]	5.00 – 52.00	30.19 (9.98)				—

Note. PC = Positional Constraint; IGL = Incidental Graphotactic Learning; CC = Contextual Constraint; PWM = Picture Word Matching; NW = Nonword.
[a]Number of words/nonwords (out of 144) read correctly within 60 seconds. [b]Total number of correct responses (out of 51).
**p < .01.

3.3 General Discussion

Experiments 2.1- 2.8 sought to determine whether statistical learning could be induced in the acquisition of novel orthographic constraints in 7-year-old typically developing children and skilled adult learners. Our results are easily summarized in three main findings. First, following a successful exposure phase, where ceiling performance by all participant groups suggested sustained attention to the training stimuli throughout this phase, we obtained clear evidence of learning. Even though the red letter exposure task did not explicitly invite participants to learn the statistical patterns embedded in the wordlike strings, adults and 7-year-old children demonstrated above-chance discrimination (*d*') ability at test. Learning was further demonstrated on reaction times among adults who made correct classification responses more quickly for pattern-conforming than nonconforming test strings, a result that was consistent with the latency pattern reported in Chapter 1 (Greek-letter positional constraints learning; Latin-letter contextual constraints learning) and the Onishi et al. (2002; Chambers et al., 2010) studies. RT effects were not found for children, a finding which may be attributed to greater variability in this dataset; it is possible that response RTs are not sensitive enough to capture learning at this early age. In any case, the lack of significant RT effects should not be taken to contradict the children's reliable discrimination accuracy results.

Second, as anticipated, children and adults learned the zero-order positional constraints more easily than the first-order contextual constraints, suggesting that incidental learning is moderated by the complexity of the orthographic pattern being learned, and that learners are sensitive to these stimulus characteristics from at least 7 years of age. Our findings on pattern complexity are consistent with the developmental trends observed in the acquisition of orthographic patterns in tasks more closely related to reading and spelling. Simple positional patterns (e.g., deterministic patterns on the allowable position of doublets) are readily acquired and used by children who have just begun to receive formal literacy instruction (Cassar & Treiman, 1997; Pacton et al., 2002; Treiman, 1993). Context-based effects, such as the coda-to-vowel contingencies studied by Treiman and Kessler (2006), take significantly longer to acquire and reliably influence children's choice of alternative spellings only after several years of schooling (Caravolas, Kessler, Hulme, & Snowling, 2005). We return to this issue below to consider in more detail how the complexity effect might have arisen in the present study, and how it might operate in statistical learning more generally.

Third, we observed a developmental effect in learning performance. This manifested in children responding more slowly and somewhat more variably than adults during the exposure task, and attaining slower discrimination RTs in the positional constraints condition. Moreover, our main discrimination accuracy analysis indicated that adults learn the informational structure embedded in the exposure strings more easily than 7-year-old children. However, while both groups learned both positional and contextual patterns reliably to above chance levels, the adults' advantage over children proved less robust for learning contextual constraints in that they were not faster than children in responding to test items, nor were they more accurate in this condition when lexical items were removed from the discrimination analyses. That is, in a follow-up analysis which excluded all lexical items (up to four in any given list), the groups no longer differed on the contextually constrained stimuli. The latter result tentatively seems to suggest that adults especially benefited from the presence of real words in performance on the contextual constraints condition. However, bearing in mind that our tasks included a relatively small pool of items and that removing items from analyses reduces statistical power, these latter analyses need to be considered with appropriate caution. Regardless the fluctuations in *d*' scores across the full and follow-up analyses, the RT and discrimination accuracy results for positional constraints learning are not fully consistent with the claim that implicit learning is age invariant, as proposed by Reber (1993). Instead, our results suggest that while such learning is certainly possible across the age span (as demonstrated by above-chance *d*' scores for both groups in both learning conditions), *what* and *how much* can be learned seems to be influenced by the amount of experience accrued in the learning domain, a variable typically correlated with age, and by the complexity of the patterns being abstracted. Age-related differences in the domain of orthographic learning have been reported in numerous previous studies (e.g. Cassar & Treiman, 1997; Hayes et al., 2006; Treiman & Kessler, 2006) and are consistent with the view that pattern sensitivity strengthens over time, as children progress towards more complete orthographic learning (Hayes et al., 2006).

Contrary to expectations, our manipulation of length of exposure did not have a robust influence on learning. Doubling the amount of exposure led to faster correct responses at test; however, it did not result in generally higher discrimination ability. At a first glance, this result is inconsistent with the idea that graphotactic sensitivity strengthens as a result of spellers' increasing exposure to print (Hayes et al., 2006). Furthermore, it is in disagreement with previous studies showing that longer exposure enhances incidental learning of statistical regularities in a wide range of domains (e.g., Gaillard et al., 2009) and learner groups (e.g., Saffran et al., 1997). There are a number of possible reasons for our participants' "failure" to benefit from longer exposure to the wordlike strings. Perhaps the difference in length between our short versus long exposure condition was not sufficiently large to induce better learning in the long condition. Alternatively, learning of complex patterns may be enhanced through the experience of a wider variety of tokens that embed the constraint, or by the "little but often" approach of spaced learning (e.g., Karpicke & Roediger, 2007). More research is needed to determine the appropriate threshold in the frequency of repetitions and/or variety of tokens experienced during the exposure phase, which would lead to reliable improvements in learning. It will also be of interest to examine specific characteristics of the learning process involved in this paradigm, such as its time course and durability. At minimum, it is clear that learning of statistically predictable patterns and contingencies develops fast, after only a few exposures to critical stimuli (e.g., Dell et al., 2000; Onishi et al., 2002). Some experimental work in the speech production domain suggests that learning effects are not long-lasting (e.g., Taylor & Houghton, 2005).

We now turn to considerations of what factors over and above the complexity of the constraints (zero-order vs. first-order constraints) may have influenced performance in the present study. For example, it is conceivable that the task demands in our exposure manipulation inadvertently, preferentially enhanced positional constraints learning. While the exposure task was designed to ensure that participants attended to the whole string in seeking the red letter, and the letters flanking the red target will have been in their effective visual field (e.g., Rayner, Well, & Pollatsek, 1980), attending explicitly to the red letter may have augmented learning of the single-letter positional constraints, while attenuating that of the two-letter contextual constraints (see Pacton & Perruchet, 2008). Also, it is possible that differences in the number of exposures to specific tokens in the positional constraints (four repetitions per token, e.g., d as onset) versus the contextual constraints (two repetitions per token, e.g., do body) condition (a necessary consequence of controlling for the length of the overall exposure phase) may have biased learning in favour of positional constraints. Mitigating against this hypothesis, however, and corroborating the suggestion that pattern complexity is a factor in statistical learning are the results of the long exposure manipulation. Here, we found that even when doubling the number of exposures in the contextual constraints condition, such that participants saw each target instance (e.g., do) four times, performance did not reach the levels obtained in the positional constraints short exposure condition, even though the number of instance exposures was now equated and the number of set-wide exposures was doubled.

We further considered what type of knowledge is formed during the incidental graphotactic learning task. Although our task involved only a visual presentation of letter strings, it cannot be ruled out that participants did not also covertly verbalise these and coincidentally also learned the constraints as phonotactic. However, as verbalization was neither encouraged nor necessary here, learning of graphotactics presumably took precedence over any additional incidental phonotactic learning. The extent to which graphotactic and phonotactic constraints are separable during orthographic learning is an interesting question for future research.

We also explored which stimulus characteristics may have provided additional, redundant cues and possibly contributed to learners' above chance classification performance. In our stimuli, both positional and contextual constraints were effectively anchored at the outer edges of the letter strings, where learning may have benefited from their perceptual salience. Statistical learning might have been less pronounced if both types of constraints were embedded in less salient positions of longer letter strings. Along these lines, it would be interesting to tease apart the relative importance of the complexity versus the position of the pattern being learned. For example, the advantage observed for positional learning might have been weaker if positional constraint patterns were middle-based and contextual constraint patterns were edged-based. Another source of redundancy arose from the design of the stimuli, in which the embedded constraints were reflected in more than one position across strings. For example, our positional constraints learning condition embedded constraints on single consonants occurring in the onset and coda positions. It is therefore possible that participants were learning the constraints from both units. Similarly, in the contextual learning experiments, we presented CV as well as VC first-order contingencies equally frequently, and it is possible that participants relied on the constraints embedded in either the body or the rime or both units for legal/illegal string discrimination. Also, learning about the legal positions of single letters may have additionally benefitted from the presence of statistical information governing the legal position of body/rime units, which occurred, although only half as frequently, in the positional constraints learning experiment. Spoken as well as written language are replete with redundancy of cues, and in this sense, our stimuli may have provided a very favourable learning context. Teasing apart correlated cues in the input and evaluating their respective importance and optimal thresholds for learning is an exciting challenge for this area of orthographic learning.

A considerable number of studies in the artificial grammar learning literature have demonstrated that task performance is driven by several, often highly correlated test item properties. Sensitivity to chunk information (i.e., co-occurring elements in a string)—usually quantified in terms of a string's associative chunk strength (e.g., Knowlton & Squire, 1994, 1996) or the amount of chunk novelty carried by a test phase string (e.g., Meulemans & Van der Linder, 1997)—and similarity to specific training items (e.g., Brooks & Vokey, 1991) are only a few alternatives to Reber's original claim (1969) that people form an abstract representation of rules. On this issue, we suggest that above-chance classification performance on our incidental graphotactic learning task does not necessarily entail forming an abstract representation of our graphotactic rules governing the position and context of letter distributions (see Pacton et al., 2002 for a similar point). Performance on the positional constraints may be interpreted as sensitivity to locations where single letters (as in our conceptualization) occurred, but also where digraphs occurred; whereas in contextual constraint learning performance, it can be argued that participants learned allowable letter co-occurrences (as in our conceptualization) or digraph constraints. The present study was not designed to disentangle the role of fragment and training/test phase item similarity, and investigations of their relative importance for learning are the focus of our ongoing research.

To conclude, the current study offers novel insights into the learning mechanisms contributing to the acquisition of graphotactic knowledge, i.e., knowledge about the legal combinations of letters in script. Although several authors have argued that statistical learning processes underlie children's appreciation of simple orthographic conventions, to our knowledge, this is the first study to directly assess how this learning can occur. Our data clearly demonstrate that a few minutes of incidental exposure is sufficient to induce learning of graphotactic patterns that are commonly found in most alphabetic orthographies. Learning effects in our study appear to emerge as early as end of Year 2, a finding that contrasts with the widely held position that orthographic knowledge is unavailable to beginner spellers (e.g., Frith, 1985; Gentry, 1982). Admittedly, our study is limited by the artificial nature of our deterministic underlying statistics. For example, our significant contextual learning effect does not necessarily imply that Year 2 children are already sensitive to the probabilistic, context-based patterns they encounter in real language settings. However, above chance learning of some simple effects of context provides support to the notion that statistical learning processes operate early on and they may provide the mechanism for the acquisition of at least some patterns in written language.

3.3.1 Correlational Analyses

Further to examining the learning mechanisms underlying the acquisition of novel graphotactic constraints, our study followed up our investigation of the relationship between statistical learning skill and literacy skill. A few previous studies (reviewed in Chapter 2) have empirically demonstrated that variations in lab-based statistical or implicit learning skill are related to literacy performance among typically developing children (e.g., Arciuli & Simpson, 2012; Steffler, 2004) and dyslexic adults (e.g., Howard et al., 2006). Our pilot study revealed only a few weak positive correlations between task performance and adult learners' performance on the Reading subtest of the WRAT; however, our sample sizes were relatively small, and the possibility that our study was inappropriately powered to reveal associations needed to be considered. We also explored the hypothesis that performance on WRAT Reading, a standardized measure of reading accuracy, may be relatively insensitive to individual-to-individual-variation among highly proficient readers. Thus, our battery was complemented by the Test of Word Reading Efficiency, a measure or real word/nonword reading fluency (Torgesen et al., 1999) and an exception word reading task. Finally, we explored the relationship between learning and literacy skill among 7-year-old children, anticipating stronger links, possibly due to—but not limited to—a broader range of variance in literacy skill in a developing population. None of these provisions proved sufficient to reveal a systematic relationship between individual differences in incidental learning performance and reading or spelling skill. In fact, in contrast to our hypotheses, we obtained virtually no correlations among either group of learners and regardless of the type of measure used to assess literacy skill. The possibility that our incidental graphotactic learning measure may not have been sufficiently internally consistent for correlations to emerge needs to be taken into account in the interpretation of our findings. Partly in response to this concern, the relationship between skill learning and literacy was assessed using a more popular measure of learning in Chapters 4 and 5: Reber's (1967) artificial grammar learning task.

Chapter 4

Artificial Grammar Learning in Skilled and Dyslexic Adult Readers

4.1 Introduction

Implicit learning processes enable the acquisition of complex pattern structure in domains as diverse as categorical learning, music, and written language acquisition (Cleeremans et al., 1998; Steffler, 2001; Tillmann & McAdams, 2004). A widely used paradigm for the study of this phenomenon is the artificial grammar learning task (Reber, 1967). As reviewed in Chapter 1, different theories have proposed that participants' ability to discriminate between grammar-conforming and nonconforming stimuli in the task may be driven by their sensitivity to abstract rules (e.g., Reber, 1989); the amount of bigram/trigram overlap between test and training phase stimuli (e.g., Perruchet & Pacteau, 1990); the similarity between test and specific training phase stimuli (e.g., Brooks & Vokey, 1991); or sensitivity to more than one of these stimulus properties (e.g., Meulemans & Van der Linden, 1997). While all of these factors may be of some relative importance to participants' judgements of grammaticality (e.g., see Pothos & Bailey, 2000), they are frequently confounded with each other, at least within finite-state grammar material, making it difficult to determine what drives above chance discrimination ability (Johnstone & Shanks, 1999).

In a carefully designed experiment, Kinder (2000; Kinder & Lotz, 2009) investigated participants' ability to categorize test-phase letter strings as grammatical/ungrammatical on the basis of three types of information: (a) the extent to which they contained allowable/familiar training chunks (chunk strength), (b) their adherence to positional constraints set by the grammar, and (c) their degree of similarity to specific letter strings presented during the acquisition (training) phase of the experiment. As in previous studies (e.g., Knowlton & Squire, 1994, 1996; Meulemans & van der Linden, 1997), chunk strength was captured by the amount of bigram/trigram overlap between a test and a training phase string. Participants' sensitivity to this source of information was investigated by comparing the endorsement of strings made of allowable/frequent training chunks relative to strings that shared, on average, fewer chunks with the full set of training items and contained some impossible letter combinations/triplets. Sensitivity to positional knowledge over and above participants' sensitivity to allowable/frequent letter pairs/triplets was assessed by comparing the endorsement of low chunk strength items (i.e., false alarm rate) containing letters in admissible absolute positions (but resulting in an illegal bigram/trigram) with participants' endorsement of illegal items embedding positional violations on top of the same amount of bigram/trigram violation. A significantly lower proportion of FAs for the latter set of items relative to those violating position and bigram/trigram information would suggest that positional constraints were reliably learnt. Finally, test/training phase string similarity was operationalized along the lines of Brooks' early studies (e.g., Brooks & Vokey, 1991), as the number of letters by which a letter string presented at test differed from its closest training item (specific similarity or edit distance): strings varying by one letter where thought to be similar to a specific training item, whereas items differing from their closest training item by 2 letters or more, where thought to dissimilar to them. The standard well-formedness task assessing learning was preceded by a memorization to criterion phase, during which participants were presented with 12 training strings repeated 4 times each in one study (Kinder, 2000) or 8 times each in another study (Kinder & Lotz, 2009) provided that participants were able to reproduce (type) all of them correctly. Kinder's (2000) study provided evidence of chunk frequency sensitivity, as well as sensitivity to the positional constraints governing the allowable position of letters. Specific similarity did not modulate participants' endorsement of similar/dissimilar items in the "shorter" (at least 4 repetitions/trial) memorization phase; however, Kinder and Lotz's (2009) follow-up study using a longer training phase (at least 8 repetitions/trial) gave rise to a significant effect of specific similarity on performance, as well as bigram/trigram and positional knowledge sensitivity. A simple recurrent network model "confronted" with the same experimental situation replicated these behavioral findings, showing reliable sensitivity to all three sources of knowledge (Kinder, 2000; Kinder & Lotz, 2009).

Kinder and Lotz's (2009) study is, to the best of our knowledge, the first systematic investigation of participants' sensitivity to the allowable position of single letters within the artificial grammar learning task. A few previous studies, though, have shown that participants' grammaticality judgments reflect, to some extent, knowledge of *where* bigrams and trigrams occur within training strings (Johnstone & Shanks, 1999; Kinder & Assmann, 2000; Perruchet & Pacteau, 1990; Dienes, Broadbent & Berry, 1991). That is, participants may be encoding that VX, a frequently occurring training bigram was usually located in stimulus beginnings (i.e., in the most leftward, position 1) or middles (e.g., position 3), but never in other positions. Perruchet and Pacteau (1990; experiment 2) were among the first to show that artificial grammar learners are weakly but reliably sensitive, relative to untrained controls, to structural violations stemming from placing an allowable bigram in an unexpected (wrong) position. Using a sequential letter dependency task (e.g., *can the stem VXR_ be followed by M? X? R?*), Dienes et al. (1991) further demonstrated that participants' willingness to endorse an allowable bigram (e.g., RM) as grammatical was systematically affected by whether or not it was allowable in a given position. However, diverting attention to an unrelated (random number generation) task during acquisition had a detrimental effect on participants' sensitivity to positional constraints on the allowable position of bigrams. As explained in Chapter 1, information governing permissible bigram and trigram locations was also identified by Johnstone and Shanks (1999) as a confound in Meulemans and van der Linden's (1997) manipulations of associative chunk strength (Experiment 2a) and grammaticality (Experiment 2b) (see section 1.3.1 of the Literature Review). More convincingly, sensitivity to chunk locations was shown to be a significant predictor of participants' classification responses in a re-analysis of Meulemans and van der Linden's (1997) data (Johnstone & Shanks, 1999) and in a more recent study using Meulemans and van der Linden's (1997) material (Kinder & Assmann, 2000).

4.1.1 Artificial Grammar Learning in Dyslexia

Deficits in unintentional implicit skill learning have been recently put forward as a potential explanation for several developmental disorders involving language, such as dyslexia (e.g., Nicolson et al., 2001) or specific language impairment (e.g., Hsu & Bishop, 2010). An often claimed advantage of this hypothesis is that general learning deficits may parsimoniously account for children's language-related difficulties, as well as their difficulties in other highly automatized domains of knowledge. In this context, several studies have examined implicit skill learning in child and adult developmental dyslexic populations by means of the serial reaction time task (Nissen & Bullemer, 1987). The mixed findings of this relatively new literature were discussed in Chapter 1. In this chapter, we discuss findings from the only four—to the best of our knowledge—studies which have employed different variants of Reber's artificial grammar learning task.

Rüsseler et al. (2006) compared skilled (*n* = 12) and dyslexic adults' (*n* = 12) ability to classify grammar-conforming/nonconforming letter strings following training (memorization to criterion) with twenty 4-letter-long strings generated from an artificial grammar. Both groups required a similar number of memorization trials to reach criterion, showed similar levels of grammaticality sensitivity and outperformed a control group trained with random, uninformative letter strings. Thus, there was no evidence of an implicit learning deficit among dyslexic adults. Pothos and Kirk (2004) assessed learning in dyslexic and nondyslexic university students using a "sequential" and an "embedded shapes" variant of the task originally used by Knowlton and Squire (1996). Half of the grammatical stumuli were made of bigrams and trigrams which appeared frequently during the memorization phase of the experiment (grammatical/high chunk strength items), whereas the other half shared, on average, fewer bigrams/trigrams with the training sequences (grammatical/low chunk strength items). Similarly, half of the ungrammatical sequences had a high amount of bigram/trigram overlap with the sequences presented during memorization (ungrammatical/high chunk strength items), whereas the other half shared, on average, fewer bigrams/trigrams with the training sequences (ungrammatical/low chunk strength items). Under the "sequential" condition, the stimuli consisted of simple geometrical shapes arranged sequentially (e.g., circle-rectangle-diamond-circle-diamond). Stimuli in the embedded condition adhered to the same underlying structure as the "sequential" stimuli but were converted to nested figures embedding the sequence of shapes (e.g,. the circle was embedded within a rectangle, which was in turn embedded within a diamond etc). Following a brief acquisition phase, during which skilled and dyslexic readers were instructed to attend to the stimuli (presented one at a time for a fixed amount of time), participants were informed about the rule-governed nature of the stimuli and were asked to discriminate between grammatical and ungrammatical novel items. Typical and dyslexic participants performed very similarly in the embedded shapes condition (mean accuracy in both groups was 54%). Surprisingly, dyslexics were significantly more accurate (55%) than nondyslexic individuals in the sequential variant, the latter performing at guessing levels.

Using a 2 by 2 factorial design, Ise, Arnoldi, Bartling, and Schulte-Körne (2012) investigated chunk strength sensitivity in 9-year-old typical versus poor spellers following exposure to pronounceable (e.g., XABOZ) versus unpronounceable (e.g., FTGCZ) strings. Although performance was reliably better than guessing under all four experimental conditions, good spellers' proportion of correct responses was significantly higher than that of poor spellers. Thus, although sensitivity to chunk strength was relatively spared in the group of poor spellers, it was concluded that implicit learning is less efficient in this population. Another recent study has assessed learning in a manner that provides some insights into dyslexics' sensitivity to chunk frequency versus abstract (rule-based) information. Following Knowlton and Squire's design (1996) as Pothos and Kirk (2004) did, Pavlidou, Kelly, and Williams (2010; see also Pavlidou, Williams, & Kelly, 2009) developed a shapes variant of the task where chunk strength and grammaticality were manipulated in an orthogonal fashion. Artificial grammar learning performance of sixteen dyslexic children was compared to that of a group of typically developing children matched for age, verbal- and performance-IQ. In discussing their findings, Pavlidou et al. (2009) highlight two group differences. First, a significant difference was observed between typically developing and dyslexic children's ability to judge the grammaticality of the test phase sequences. That is, when grammatical and ungrammatical items were matched for chunk strength, only typically developing children were able to classify the sequences correctly at above chance levels. Similarly, it was argued that, in contrast to typically developing children, dyslexic children's sensitivity to chunk strength was at chance.

There are several reasons for caution in the interpretation of the above findings on artificial grammar learning in dyslexia. Any conclusion based on Pothos and Kirk's (2004) finding that dyslexics are superior to skilled readers in some aspects of implicit learning is at best tentative, given that learning was not reliably demonstrated in the case of controls. What is more, from a study design point of view, both studies with adult dyslexics (Pothos & Kirk, 2004; Rüsseler et al., 2006) failed to ensure that groups were matched in terms of general ability, a widely recognized methodological pitfall in the interpretation of dyslexia data (Goswami, 2003). Group matching was not an issue in the developmental studies of Pavlidou et al. (2010) or Ise et al. (2012). However, there are some statistical analysis issues in the former study. For example, chunk strength sensitivity in Pavlidou et al. (2010) was analysed by means of two separate one-sample *t* tests (one on the proportion of "yes" responses for high associative chunk strength items, and one the proportion of "yes" responses for low associative chunk strength items) againsts chance (i.e., 50%) performance. This type of analysis is open to criticism along the lines of signal detection theory (Macmillan & Kaplan, 1985), given that hit rates may confound participants' sensitivity and response criterion (i.e., their overall tendency to respond "yes" across the full set of test phase items). *d*' sensitivity measures, as the ones used throughout this thesis, offer an elegant solution to the contamination of participants' discrimination ability by their response bias and comprise an adequate way of comparing learning against chance and/or between groups.

4.1.2 Correlates of Implicit Skill Learning

Chapter 2 reviewed evidence from experimental studies investigating the relationship between individual differences in literacy skill and the ability to learn frequency-based information without engaging deliberate effort. Although this is a relatively new area of research, there is some evidence of a positive relationship between performance on different implicit/statistical learning measures and reading/spelling ability (however, for nonreplications of a significant association between literacy and implicit/statistical learning performance, see Chapters 2 and 3, and Waber et al., 2003). These associations are taken to corroborate the hypothesis that the ability to detect statistical structure is fundamentally linked to literacy acquisition and highlight the need for longitudinal studies to gain further insights into the relationship between these two skills (Arciuli & von Koss Torkildsen, 2012).

Empirical interest in the relationship between implicit learning and psychometric intelligence, frequently defined and quantified in terms of performance on standardized measures of IQ, has a longer history in the field. Reber's (1993) theoretical assertion that learning under implicit conditions, being evolutionarily older than explicit learning, should be unrelated to intellectual level has been supported by a substantial body of research. For example, Reber et al. (1991) showed that adults' performance on an artificial grammar learning task was only weakly ($r = .25$) and nonsignificantly correlated with a standardized overall measure of IQ (Wechsler Adult Intelligence Scale-Revised: WAIS-R; Wechsler, 1981). On the other hand, explicit skill learning, measured by performance on a 2-alternative forced choice task requiring explicit problem solving strategies (e.g., which one is the next letter in the sequence ABCBCDCDE, *D* or C?), was significantly correlated with IQ ($r = .69$, $p < .01$). Replicating Reber et al.'s (1991) findings in a larger sample of participants aged between 18 and 77 years, McGeorge, Crawford, and Kelly (1997) obtained (a) a nonsignificant and much weaker pattern of correlation between artificial grammar learning performance and full-scale IQ measured by the full WAIS-R (r = .04), and (b) a significant positive association between full-scale IQ and performance on a "series completion" learning task similar to that used in Reber et al.'s (1991) study (r = .58).

More compelling evidence for different patterns of association between psychometric intelligence and implicit versus explicit learning is provided by Gebauer and Macintosh (2007). In this study, both forms of learning were measured using the same artificial grammar learning task—adapted from Vokey and Brooks (1992)—under implicit and explicit (i.e., rule discovery) instructions, thereby, eliminating the methodological confound of possible idiosyncratic task differences. What is more, high internal consistency reliability (Cronbach's alpha ≈ .75) was documented for the final version of the task employed in this study. Using an impressive sample of 605 participants between 11 and 32 years of age, Gebauer and Macintosh (2007) examined the relationship between artificial grammar learning performance—among other implicit learning tasks—and three different components of intelligence: Crystallized intelligence (g_c), fluid intelligence (g_f) and memory (g_y). Performance in the implicit instruction group (percentage of correct test phase responses to the question: "is the string new/old") showed a very weak significant positive correlations with memory (r = .10) and no relationship with individual differences in crystallized or fluid intelligence. On the other hand, variations in $g_{f,}$ and g_y were significantly positively correlated with performance in the explicit instruction group (percentage of correct test phase responses to the question "is the string derived from the rule system you tried to uncover in the previous phase"), even though both associations were modest (fluid intelligence: r = .25; memory: r = .27). Importantly, these findings were replicated using an additional within-subject repeated measures design, whereby, participants who performed under implicit learning conditions were administered an explicit learning version of the task.

Reasoning along similar lines against the use of a single IQ score, Kaufman et al. (2010) investigated the relationship between implicit learning and psychometrical intelligence with a large battery of elementary tasks that are associated with intelligence (e.g., paired-associates learning, speed of processing skill; see Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009) and higher-order cognitive tasks that typically load to a single aggregated measure of intelligence, "g", (e.g., Raven's Advanced Progressive Matrices test). Although it should be noted that the task used by Kaufman et al. (2010) was a serial reaction time task, and thus their findings are not directly comparable to the studies described earlier, the study goes beyond previous research in showing that, while unrelated to "g" and working memory, implicit learning is positively associated with more elementary cognitive abilities such as speed of processing skill. As an interesting aside, implicit skill learning in Kaufrman et al.'s (2010) study was positively associated with participants' scores on two academic achievement tests tapping on second language learning (GCSEs German and French). This relationship, which remained significant after controlling for IQ, is not only consistent with the proposed link between implicit skill learning and second language acquisition (e.g., Robinson, 2005), but also demonstrates a relationship between this type of learning and language learning skill outside the laboratory.

4.2 Experiments 3.1 – 3.2

4.2.1 Rationale

The proposed link between implicit skill learning and literacy impairment is consistent with Nicolson and Fawcett's (1990, 1994, 1995; Nicolson et al., 2001) automaticity (cerebellum) deficit hypothesis of dyslexia, according to which reading fails to automatize as a result of a general learning deficit. For example, it has been suggested that impaired implicit skill learning could undermine the development of well-specified phonological representations or have a detrimental effect on dyslexic children's ability to form grapheme-to-phoneme associations (Gombert, 2003; Howard et al., 2006, Sperling et al., 2004). As discussed in the Literature Review, automatization deficits are thought to be particularly prominent when conscious compensation strategies are prevented from operating. Therefore, learning under implicit conditions is a good candidate for assessing the validity of this theory: conscious compensation cannot presumably occur in the absence of "knowledge that you are learning". To date, the small literature investigating implicit learning in dyslexia has yielded equivocal results. Furthermore, most research to date has employed the serial reaction time task, a motor-reliant-processing measure of implicit skill learning. It is possible that dyslexic readers' performance may be spared once the need to execute sequential motor movements is attenuated. Thus, one of the aims of the present study was to shed light on these inconsistent findings by investigating dyslexic adults' performance on an artificial grammar learning task. If implicit learning deficits contribute to the literacy deficits experienced by dyslexic individuals, one would expect them to perform at chance and/or significantly worse than a group of skilled readers. What is more, one would expect significant positive correlations between implicit learning and literacy performance scores.

On the other hand, spared performance would suggest that implicit learning deficits are not causally related to dyslexia.

As has been demonstrated by several studies in the domain of artificial grammar learning, implicit learning may involve learning of different types of information, upon which rests successful test phase discrimination ability. Thus, a more pertinent question than asking whether learning is spared or impaired is whether dyslexic individuals are less sensitive to certain forms of knowledge that can be acquired during the task. However, the findings of both studies using a dyslexic adult population (Pothos & Kirk, 2004; Rüsseler et al., 2006) are silent to this question. Rüsseler et al. (2006) reported spared learning only in terms of grammatical sensitivity. Pothos and Kirk's (2004) employed a design pitting grammaticality against chunk strength; nevertheless, analyses of participants' sensitivity to the latter were not reported. Pavlidou et al. (2010) showed that dyslexic children are impaired in the discrimination of grammatical versus ungrammatical items when these were matched for associative chunk strength; however, the study did not provide appropriate analysis of dyslexic children's chunk strength sensitivity. It is clear that a well-controlled study investigating sensitivity to chunk strength information—that is, frequency-based learning—among typical and dyslexic readers is still required. Also, there are no previous studies assessing dyslexic adults' ability to learn positional constraints or whole-training exemplars controlling for chunk frequency information. Both of these properties are ecologically relevant constructs to reading and spelling skill. The allowable position of letters/digrams within one's written language is among the first types of statistical information acquired by developing spellers (Cassar & Treiman, 1997; Pacton et al., 2001; Treiman, 1993) and can be learnt, at least to some extent, under minimal exposure/incidental conditions (Chapter 3). Sensitivity to specific similarity may also map—although perhaps more indirectly—onto real orthographic learning. Reading and spelling performance can benefit from an analogical inference strategy (e.g., an unknown word such as *peak* may be read in analogy to the known word *beak*), as suggested by a body of literature (e.g., Goswami, 1988; Ehri, 1997). To this end, the present study sought to examine whether sensitivity to these factors, which have been previously shown to influence artificial grammar learning performance in typical populations also forms the basis of dyslexic students' classification performance.

The artificial grammar learning task used in this study replicated all aspects of the material, design and procedure of Kinder and Lotz's (2009; Experiment 2). There are several reasons why this particular grammar was chosen. From a methodological point of view, there are two advantages, in that, the grammar (a) generates strings of equal length and (b) does not create salient patterns of repetitions (i.e., immediate repetitions of letters such as MT*RRR*). Thus, it is no longer necessary to control for the effect of these variables in classification accuracy. Second, unlike most standard finite-state grammars, Kinder and Lotz's (2009) grammar allows for a systematic manipulation of the position of letters, a type of information that is statistically constrained in written language and is highly relevant to literacy skills (e.g., Treiman, 1993), yet is relatively overlooked/hard to assess in artificial grammar learning studies. Sensitivity to chunk frequency information and specific similarity, another two sources of information manipulated within Kinder and Lotz's (2009) design, are similarly important for orthographic learning (Ehri, 1994, 1995; Frith, 1985; Goswami, 1988, Nation, 1996). In this respect, Kinder and Lotz's design is well-suited for a detailed investigation of dyslexic readers' implicit learning abilities.

4.2.2 Method

4.2.2.1 Participants

Thirty adults with developmental dyslexia (13 male; mean age = 24.17 years, range = 18.50 years – 54.42 years) and 32 skilled adult readers (6 male; mean age = 20.36 years, range = 18.50 years – 35.75 years) took part in the study. Skilled readers were recruited through the Bangor University participant panel and received course credit for their participation. Inclusion criteria were: (a) no documented history of dyslexia or other learning difficulties, (b) WRAT Reading/Spelling performance above 95 (checked a posteriori), and (c) full-scale, verbal and performance IQ above 85 (checked a posteriori). On these grounds, data from 8 participants were excluded from any further analyses reducing the sample to 24 participants (6 male; mean age = 20.49 years, range = 18.50 years – 35.75 years). Dyslexic readers were recruited through the Miles Dyslexia Center at Bangor University and were compensated monetarily for their participation. Half of the participants had received a formal diagnosis of dyslexia via the Miles Dyslexia Center's assessment service; the other half were diagnosed by a qualified educational psychologist during secondary school or later. The inclusion criteria for dyslexic participants were: (a) no known co-occurrence of another developmental disorder and, (b) full-scale, verbal and performance IQ above 85 (checked a posteriori). To assess for co morbidity with other developmental disorders, dyslexic students' assessment reports were examined following written consent. Data from 2 participants were excluded on these grounds. Three age outliers were further excluded to achieve age matching with the group of skilled readers. The final sample consisted of 25 dyslexic students (9 male; mean age = 21.63 years, range = 18.50 years – 27.67 years). All participants were monolingual native English speakers and reported having normal or corrected-to-normal visual acuity. Ethical clearance for the study was granted by the School of Psychology, Bangor University.

4.2.2.2 Background measures

Participants received an extensive battery of standardized tests that assess cognitive, literacy and literacy-related skills. As in our previous studies, single-word reading and spelling were assessed using the Word Reading and Spelling subtests of the WRAT (Wilkinson & Robertson, 2006). General cognitive ability was assessed using all four subtests (Matrices, Verbal Analogies, Diamonds, and Vocabulary) of the Wide Range Intelligence Test (WRIT; Glutting, Adams, & Sheslow, 2000). The test allowed full-scale, verbal and nonverbal IQ to be derived. Verbal short-term memory was measured using the Digit Span subtest of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1998). Speed of processing was measured using the Symbol Search subtest of the WAIS-III (Wechsler, 1998). In addition to Symbol Search, participants were also administered the Symbol Digit Modalities Test (SDMT; Smith, 1991), a variant of the Digit Symbol-Coding subtest of the WAIS-III (Wechsler, 1998) assessing visual-motor coordination/speed of processing skill. Finally, a nonword phoneme deletion task and two rapid automatized naming tasks were used as an index of participants' phonological processing skill and naming speed. The administration of such an extensive battery served three purposes. It allowed us to (a) establish the presence of literacy difficulties in our dyslexic sample relative to our skilled reader sample; (b) assess the typicality of dyslexic individuals' profile regarding the presence of a phonological processing deficit; (c) explore associations between artificial grammar learning performance and reading/spelling skill, *as well as* more indirect oral language abilities (phoneme awareness, rapid automatized naming) that are robust predictors of reading and spelling ability (Caravolas et al., 2012; Lervåg, Bråten, & Hulme, 2009).

Rapid automatized naming. Two versions of a Rapid Automatized Naming (RAN) task (digits; objects) taken from Caravolas et al. (2012) were administered. Participants were presented with an array of high frequency items (digits condition: 2, 3, 6, 7, 9; objects condition: dog; eye; hat; lion; table) and were asked to name them sequentially (left to right, across and down the page) as fast as they could. Items were repeated eight times over five lines of an A4 card (see Appendix E and F). Two blocks were administered in each condition, with items arranged in a quasi-random order. Performance on each block was timed from the onset of participants' first response to the offset of their last response. Response latencies from the two blocks were combined for each participant. The reliability (that is, the correlation between performance in the first and second block) was $r = .96$ for RAN Digits and $r = .88$ for RAN Objects.

Nonword phoneme deletion. Phonological awareness was measured by a nonword phoneme deletion task developed by Caravolas for use with dyslexic adults (e.g., Judge, Caravolas, & Knox, 2006, 2007). The task required participants to identify a specified phoneme of a CVCC nonword and correctly reproduce the nonword without the phoneme as quickly as possible. During block 1 ($n = 12$) participants were presented with nonwords such as /stɛk/ and were required to delete the second phoneme (i.e., produce the nonword /sɛk/). During block 2 ($n = 12$), the items required the deletion of the penultimate phoneme (e.g., /fɛsp/ → /fɛp/). Two practice trials were given prior to the administration of each block to familiarize participants with the block requirements and procedure. To ensure that each nonword was heard correctly, participants were asked to repeat it before removing the relevant phoneme. Items that were misheard or incorrectly repeated (e.g., /tɹib/ instead of /tɹip/) were given a score of 1 provided that the relevant phoneme was correctly identified and the item was repeated consistently with and without the relevant phoneme. The maximum accuracy score averaged across blocks was 12. The Cronbach's Alpha reliability was .92. Response latencies—recorded separately for each block from the onset of participants' repetition of the first stimulus to the offset of their last response—were also averaged across block. The reliability was $r = .95$.

4.2.2.3 Material

The material was identical to that used by Kinder and Lotz (2009, Experiment 2). Twenty-four 7-letter strings were generated by traversing the paths of the artificial grammar shown in Figure 4.1. These were divided into two sets of 12 items each (list 1 and 2) and were used as training exemplars during the memorization phase of the experiment. List was counterbalanced across participants, so that half of them were exposed to list 1 strings, and the other half were exposed to list 2 strings. The two sets of training strings are shown in Tables G1 and G2 in the Appendix.

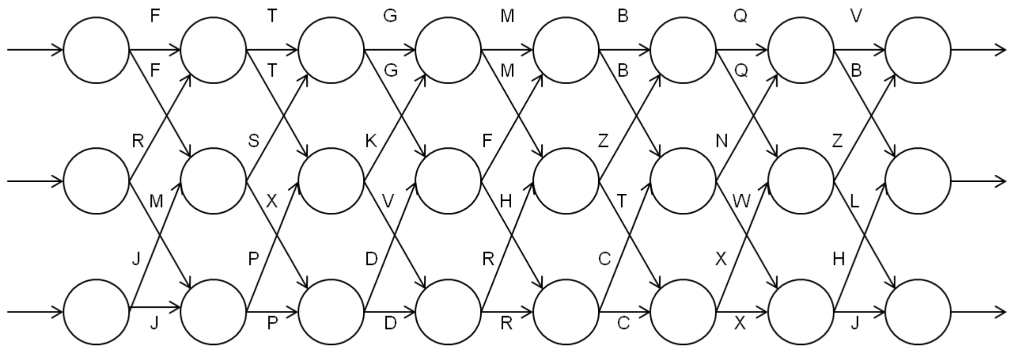


Figure 4.1. Finite state grammar used in Experiments 3.1 - 4.4. Adapted from Kinder and Lotz (2009)

The same 96 items used by Kinder and Lotz (2009) served as test phase items. These were assigned to two counterbalanced lists shown in Tables H1 and H2 in the Appendix. Twenty-four (i.e., half) of the items in each list consisted solely of *permissible* chunks, the majority of which appeared *frequently* during the memorization phase of the experiment (*high chunk strength* items). The remaining 24 items (*low chunk strength* items) contained some permissible chunks shown during the memorization phase, but in addition, they all contained *illegal chunks* (e.g., HF is not allowed by the grammar shown Figure 4.1). To quantify differences between the two 24-item sets in each list, we calculated their mean global associative chunk strength, a frequently used measure of memorization/test phase string chunk overlap (e.g., Knowlton & Squire, 1994; Meulemans & van der Linden, 1997). As in the Knowlton and Squire (1994) study, this was computed by (a) partitioning each test phase item (e.g., list 1 test phase item FSGMBWJ) into its constituent bigrams (FS, SG, GM, MB, BW, WJ) and trigrams (FSG, SGM, GMB, MBW, BWJ), (b) summing their frequency of occurrence across the 12 training items (1+2+2+2+1+1+1+1+1+1+1) and (c) averaging the sum across the 11 chunks which comprised *each* test phase string in this experiment (14/11 = 1.27). As anticipated, an independent *t* tests confirmed that the mean global associative chunk strength of low strength items (list 1: M = 0.64, SD = 0.23, list 2: M = 0.64, SD = 0.22) was significantly lower than the mean global associative chunk strength of high strength items (list 1: M = 1.27, SD = 0.10, list 2: M = 1.27, SD = 0.13), $t_{\text{list1}}(32.03) = 12.11$, $p < .001$, d =

3.50, $t_{\text{list2}}(37.95) = 11.81$, $p < .001$, $d = 3.43$. Several authors have highlighted the psychological salience of information embedded in word beginnings and endings (e.g., Endress & Mehler, 2010) and the strong influence chunks embedded in such positions exert in artificial grammar learning performance (e.g., Redington & Chater, 1996). Thus, we also sought to ensure that the two sets of test strings differed significantly in terms of their chunk strength at anchor positions. Following Knowlton and Squire (1994), the anchor associative chunk strength of a test phase item (e.g., FSGMBWJ) was the average frequency of its initial/final bigrams (FS, WJ) and trigrams (FSG, BWJ) at the beginnings and ends of all 12 training items ([1+1+1+1]/4 = 1). Once again, independent *t* tests on the mean anchor associative chunk strength of high versus low strength strings confirmed that the difference was significant, $t_{\text{list1}}(46) = 6.40$, $p < .001$, $d = 1.85$, $t_{\text{list2}}(46) = 6.09$, $p < .001$, $d = 1.76$ and in the right direction (low strength items: $M = 0.60$, $SD = 0.22$, list 2: $M = 0.58$, $SD = 0.19$; high strength items: list 1: $M = 1.00$, $SD = 0.21$, list 2: $M = 1.00$, $SD = 0.28$).

In sum, both analyses clearly demonstrate that chunk frequency information reliably discriminated between the high and low chunk strength test phase items. However, it should be noted that grammaticality was entirely confounded with chunk frequency information in the current design (Kinder & Lotz, 2009). That is, high chunk strength items always adhered to the rules set by the grammar (i.e., they were grammatical in nature) whereas low chunk strength items were always also ungrammatical in nature. Thus, although the two sets of strings are referred to as high and low chunk strength strings, respectively, throughout this chapter, they not only differ in terms of their mean chunk frequency, but also in terms of their grammaticality.

Sensitivity to letter position knowledge. Kinder and Lotz's (2009) low chunk strength items consisted of two subtests of 12 items each (in each counterbalanced list), which were used to assess participants' sensitivity to positional information. Type I items violated the grammar in terms of the distribution of adjacent elements, as well as the absolute allowable position of letters. For example, the illegal bigram FM used in the string FXDHFML also violates the grammar's constraint that F and M cannot occur in position 5 and 6, respectively. The remaining 12 items (type II) always respected the absolute position of letters set by the grammar, yet contained an illegal bigram/trigram (e.g., the bigram RB of the string FSKRBWJ is illegal, even though R and B are allowed in position 4 and 5, respectively). Type I and type II items were matched for global associative chunk strength,

$t_{list1}(22) = 0.64$, $p > .05$, $d = 0.26$, $t_{list2}(22) = 0.08$, $p > .05$, $d = 0.03$ (list 1: $M_{type\ I} = 0.67$, $SD_{type\ I} = 0.25$ vs. $M_{type\ II}= 0.61$, $SD_{type\ II} = 0.21$; list 2: $M_{type\ I} = 0.64$, $SD_{type\ I} = 0.23$ vs. $M_{type\ II}= 0.64$, $SD_{type\ II} = 0.22$) and were also identical in terms of anchor associative chunk strength, (list 1: $M_{type\ I} = 0.60$, $SD_{type\ I} = 0.23$ vs. $M_{type\ II} = 0.60$, $SD_{type\ II} = 0.23$; list 2: $M_{type\ I} = 0.58$, $SD_{type\ I} = 0.19$ vs. $M_{type\ II} = 0.58$, $SD_{type\ II} = 0.19$). This matching ensured that both sets of items contained the same amount of bigram/trigram violation; thus, any differences in their endorsement could be only accounted for by the presence/absence of positional violations.

Sensitivity to specific similarity. High chunk strength items were similarity divided into two subsets of 12 items (in each counterbalanced list) which served to assess participants' sensitivity to specific similarity. In Brooks' terminology (e.g., Brooks & Vokey, 1991), test phase items that differ from their closest training item by a single letter are thought to be similar ("close") to them, whereas items that differ by 2 letters or more are thought to be dissimilar ("far") to them. Kinder and Lotz's (2009) type III items (e.g., FXDHCXJ) deviated from their closest training item (in this case, JXDHCWH) by three letters or more. Thus, they were dissimilar to training items. Type IV items (e.g., MP*D*RTXL) differed by exactly one letter from their closest training item (in this case MP*V*RTXL), thus, they were similar to one of the items presented during the memorization phase. Type III and type IV items were identical in terms of global associative chunk strength, (list 1: $M_{type\ III} = 1.27$, $SD_{type\ III} = 0.10$ vs. $M_{type\ IV} = 1.27$, $SD_{type\ IV} = 0.11$; list 2: $M_{type\ III} = 1.27$, $SD_{type\ III} = 0.13$ vs. $M_{type\ IV} = 1.27$, $SD_{type\ IV} = 0.15$) and anchor associative chunk strength (list 1: $M_{type\ III} = 1.00$, $SD_{type\ III} = 0.21$ vs. $M_{type\ IV}= 1.00$, $SD_{type\ IV} = 0.21$; list 2: $M_{type\ III} = 1.00$, $SD_{type\ III} = 0.30$ vs. $M_{type\ IV} = 1.00$, $SD_{type\ IV} = 0.26$)[8]. Thus, any difference in the endorsement of the two subsets could only be attributed to the factor of specific similarity.

---

[8] Our calculations of global/anchor associative chunk strength for type I – type IV items match those reported by Kinder (2000; Table 1, p. 99) but are different from those reported by Kinder and Lotz (2009; Table 2, p.667), even though the material was identical across studies. To the best of my knowledge, this discrepancy is due to a nontrivial difference in the way associative chunk strength metrics were calculated. That is, Kinder (2000; as well as ourselves) computed both metrics on the basis of a single block of training strings, whereas Kinder and Lotz (2009) appear to have calculated these metrics on the basis of 4 blocks of training strings (which is in fact, rather surprising, given that each training string occurred 8 times in Kinder and Lotz's [2009] study). This discrepancy has no effect whatsoever on the *t* tests ran to ensure matching between type I – type II items, as well as type III – type IV items.

4.2.2.4 Apparatus

Computer experiments were run on a Windows XP-based laptop with a 15 inch CRT color monitor. The experiment was designed and run using the E-prime software (Schneider et al., 2002).

4.2.2.5 Procedure

Participants were tested individually in a single session that took approximately 2 hours to complete. The artificial grammar learning task was administered first followed by the background measures given in a fixed order.

Artificial grammar learning task. Participants performed the memorization phase of the experiment (phase 1) which was followed by a surprise test phase (phase 2). A typical memorization/test phase trial is shown in Figure 4.2. In the beginning of phase 1, participants were told that they were taking part in a short-term memory task involving 7-letter strings. During each trial, a training string was displayed at the center of a white background for unlimited time. Participants were instructed to memorize it and press spacebar when ready. Following a 3000ms interval (black fixation on a white background), a blank screen appeared and participants were asked to reproduce the string on the keyboard. Self-corrections were allowed using the backspace key. To submit their response, participants were asked to press spacebar when ready. If the response was correct a new string was presented. If the response was incorrect, the same string was presented again. The cycle repeated until all 96 memorization trials (8 repetitions/training string) were correctly reproduced. Training strings were presented in a single block; however, participants were prompted to use breaks as often as they wished between trials by pressing Cntrl + Shift. The order of trials was randomized for each participant.

Following completion of the memorization phase, the well-formedness task was administered. Participants were informed that all previous letter strings were formed according to a complex set of rules which was hard to unravel. They were asked to classify a new set of strings as grammatical or ungrammatical, based on whether they thought that they followed the rules used to generate the strings they had just memorized. To avoid extreme bias in responding “yes”/“no”, participants were informed that approximately half of the strings conformed to the rules, whereas the other half violated the rules. Test strings were presented one at time in a randomized order. Each test string was displayed at the center of a white background and remained on the screen until a response was collected. Each string presentation was followed by a 1000 ms fixation (black cross appearing at the center of the computer screen). Stimuli were presented in a single block and no feedback was given.

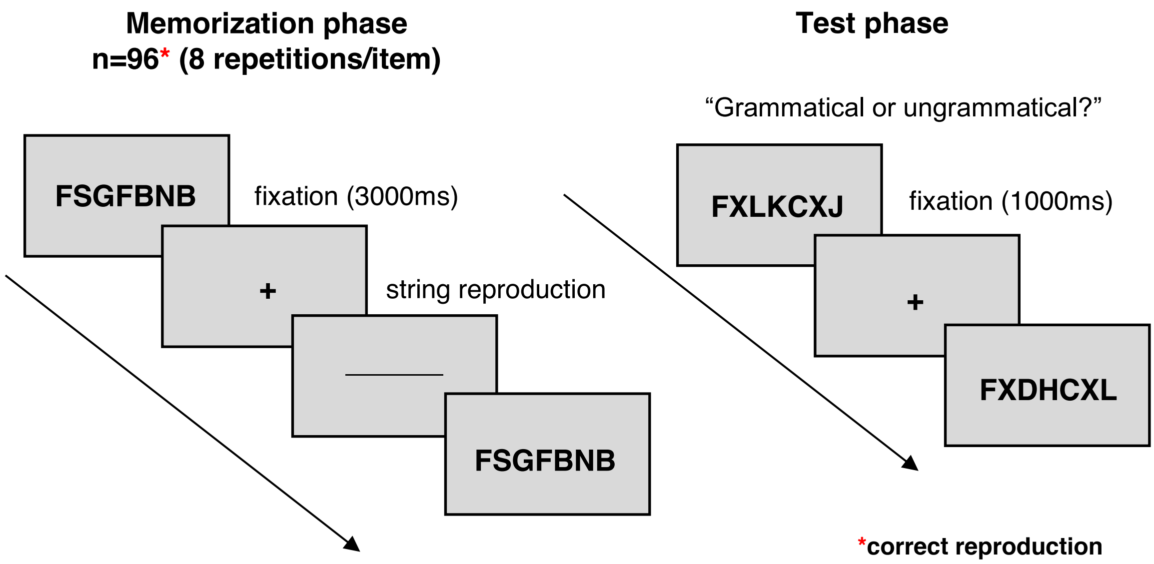


Figure 4.2. The artificial grammar learning task. Trial outline.

4.2.3 Data Analyses

Performance during the memorization phase of the experiment was measured by (a) the proportion of strings reproduced correctly within a single attempt, (b) the mean number of trials needed for correct reproduction of all 96 letter strings (i.e., number of trials to criterion), and (c) mean correct memorization RTs. All three variables were subjected to independent *t* tests with group (skilled readers, dyslexic readers) as a between-subject factor. To investigate participants' ability to discriminate between test strings on the basis of chunk strength, mean endorsement rates were subjected to 2-way mixed factorial ANOVAs with group (skilled readers, dyslexic readers) as a between-subject variable and chunk strength (high, low) as a within-subject variable. A chunk strength sensitivity measure, *d*', and the associated measure of response bias, *c*, was also computed—separately for skilled and dyslexic readers—on the basis of the proportion of high versus low chunk strength strings classified as grammatical (hits; FAs). Finally, separate analyses were carried out on participants' endorsement of low strength items and high strength items to investigate their sensitivity to positional information and specific similarity, respectively. To analyze sensitivity to the former, endorsement rates for low chunk strength items were subjected to a mixed 2 by 2 factorial ANOVA with group (skilled readers, dyslexic readers) as a between-subject variable and type of violation (chunk and positional violations [type I items], chunk violations [type II items]) as a within-subject variable. For the effect of specific similarity, endorsement rates for high chunk strength items were subjected to a mixed 2 by 2 factorial ANOVA with group (skilled readers, dyslexic readers) as a between-subject variable and specific similarity (similar [type IV items], dissimilar [type III items]) as a within-subject variable. In all of the analyses, list (1, 2)—which was counterbalanced across subjects— was initially included to ensure that this manipulation did not result in systematic bias. None of the main effects or interactions involving list as a factor was statistically significant in the memorization phase data analyses (all ps > .05). Accordingly, data were collapsed across lists. List interacted with a number of factors in the test phase analyses. Analyses where significant interactions emerged are reported including this factor.

4.2.4 Results

4.2.4.1 Background measures

The means and *SD*s for skilled and dyslexic readers' performance on each of the cognitive and literacy measures are shown in the left panel of Table 4.1. Group differences were analyzed by means of independent t tests, the results of which are shown in the right panel of Table 4.1. Skilled and dyslexic readers were matched for age, verbal, nonverbal and full-scale IQ. However, it is noteworthy that dyslexic readers had slightly higher (above average) nonverbal IQ performance when compared to skilled readers, who scored in the upper limit of the average range of the WRIT. Although this seems a somewhat unexpected finding in the general population, it may simply reflect the fact that many individuals with dyslexia in higher education are able to compensate for their language-based difficulties on the basis of relatively high nonverbal abilities.

Skilled readers performed within the average and above average range of the WRAT Reading and Spelling test, respectively. Dyslexic participants performed within the average range on both subtests of the WRAT, a finding which suggests that our sample indeed consisted of relatively well-compensated dyslexic adults. However, as expected, both literacy scores were significantly worse relative to those of skilled readers. Dyslexic individuals additionally demonstrated lower scores relative to typical readers—although their means were well within the average range—in verbal short-term memory (Digit Span performance) and speed of processing skill (Symbol Search; Symbol Digit Modalities Test performance). Finally, significant between-group differences emerged in terms of performance on the phonological processing measures. Dyslexic participants were significantly slower and more error prone on the nonword phoneme deletion task. Significantly longer response latencies were also observed in the rapid automatized naming of digits and objects.

Table 4.1. Mean scores (*SD*s) and Ranges on the Cognitive and Literacy Related Measures as a Function of Group and Results from Independent *t* tests (Experiments 3.1 – 3.2).

Variable	Skilled readers (n = 24)		Dyslexic readers (n = 25)		t	sig.	Cohen's d
	Mean (*SD*)	Range	Mean (*SD*)	Range			
Handedness (r; l)	23; 1		23; 2				
Age (years)	20.49 (3.38)	18.50 – 35.75	21.63 (2.39)	18.50 – 27.67	1.36	.179	0.39
IQ tests (WRIT)							
Full-scale IQ[a]	107.96 (7.78)	92.00 – 120.00	109.80 (10.01)	97.00 – 137.00	0.72	.477	0.21
Verbal IQ[a]	105.13 (8.51)	86.00 –119.00	104.08 (10.09)	90.00 – 128.00	0.39	.697	0.11
Nonverbal IQ[a]	108.83 (11.52)	87.00 – 128.00	113.40 (11.49)	95.00 – 137.00	1.39	.171	0.40
WRAT Reading[a]	105.96 (6.43)	92.00 –120.00	96.68 (7.71)	85.00 – 114.00	4.56	***	1.31
WRAT Spelling[a]	115.50 (11.75)	95.00 – 136.00	93.08 (11.55)	60.00 – 114.00	6.73	***	1.92
WAIS Digit Span[b]	10.17 (2.33)	7.00 – 17.00	8.64 (2.23)	4.00 – 16.00	2.34	*	0.67
WAIS Symbol Search[b]	13.67 (3.16)	7.00 – 19.00	11.64 (2.94)	4.00 – 16.00	2.33	*	0.67
SDMT[c]	61.08 (7.73)	48.00 – 75.00	49.88 (11.08)	21.00 – 68.00	4.09	***	1.17
RAN digits mean time[d]	14.14 (2.52)	8.99 – 17.54	18.71 (6.25)	12.00 – 40.90	3.33	**	0.96
RAN objects mean time[d]	20.94 (2.44)	16.99 – 26.05	25.93 (7.08)	16.91 – 50.62	3.32[e]	**	0.94
NWPD accuracy[f,g]	11.35 (0.71)	9.00 – 12.00	10.76 (1.02)	8.50 – 12.00	2.32[e]	*	0.67
NWPD latencies[d,f,g]	39.86 (10.55)	24.85 - 78.88	58.36 (20.96)	29.95 – 107.77	3.91[e]	***	1.11

Note. l = left-handed. r = right-handed. RAN= Rapid Automatized Naming. SDMT = Symbol Digit Modality Test. WAIS = Wechsler Adult Intelligence Scale; WRAT = Wide Range Achievement Test; WRIT = Wide Range Intelligence Test.
[a]Standard scores. [b]Scaled scores. [c]Raw scores. [d]In seconds. [e]Correction for unequal variances applied. [f]Max = 12. [g]Skilled readers: n = 23 due to missing data from one participant. [g]Averaged across blocks.
* $p < .05$. ** $p < .01$. *** $p < .001$.

4.2.4.2 Artificial Grammar Learning

Memorization phase. To investigate potential group differences in memorization phase performance, three independent *t* tests with group as a between-subject variable were carried out. First, we examined differences in participants' ability to correctly reproduce the strings on their first attempt. Dyslexic readers were slightly less successful (77.50% correct, $SD = 0.13$) than skilled readers (82.42 % correct, $SD = 0.08$); however, the difference did not reach statistical significance, $t(38.81) = 1.57$, $p = .125$, $d = 0.45$. Next, an independent *t* test was carried out to compare group differences in the mean number of trials needed for the accurate reproduction of each of the 96 strings. The distribution of this measure was slightly positively skewed (skilled readers: skew = 1.13, $SE = 0.47$; dyslexic readers: skew = 1.67, $SE = 0.46$), partly due to the presence of three extreme participant outliers (skilled readers: $n = 2$; dyslexic readers, $n = 1$). To reduce the influence of such outliers, their scores were replaced with the values of the sample mean + 2.5 *SD*s. A trend for more trials/string in the dyslexic ($M = 1.34$, $SD = 0.27$) than skilled readers group ($M = 1.24$, $SD = 0.14$) failed to reach significance, $t(35.94) = 1.82$, $p = .077$, but was associated with a medium effect size, $d = 0.52$. A final set of analyses was carried out to explore group differences in terms of mean correct memorization RTs (trimmed to the values of 2 *SD*s from each group mean). An independent *t* test showed that dyslexic participants were significantly slower in memorizing the strings ($M = 7673.59$, $SD = 2478.51$) when compared to skilled readers ($M = 5142.54$, $SD = 1950.12$), $t(47) = 3.96$, $p < .001$, $d = 1.13$.

Taken together, the above analyses indicate some group differences in memorization phase performance. Although differences were marginal in terms of performance accuracy, dyslexic participants were significantly slower, which suggests that the task posed a greater challenge to them. The implications of this finding for the comparison of skilled versus dyslexic readers' test phase performance are considered later on.

Test phase. Accuracy data were initially examined for extreme outliers. With the exception of two responses associated with extreme RTs (< 100 ms), all other data points were included in the analyses. Mean endorsement rates for high and low chunk strength strings are shown separately for skilled and dyslexic readers in the left of panel of Table 4.2. These were subjected to a three-way ANOVA with group (skilled readers, dyslexic readers) and list (1, 2) as between-subject variables and chunk strength (high, low) as a within-subject variable. The AVOVA revealed a significant main effect of chunk strength, with higher endorsement rates for high chunk strength (M = .70, SE = 0.02) than low chunk strength items (M = .70, SE = 0.02), $F(1, 45) = 238.92$, $p < .001$, $\eta^2 = .82$. There was no group by chunk strength interaction, $F(1, 47) < 1$, indicating a similar effect of chunk strength on skilled and dyslexic readers' responses. The effect of list was not significant, $F(1, 45) <1$, neither was the chunk strength by list, $F(1, 45) <1$, or list by group interaction, F(1, 45) = 1.83, p > .05, $\eta^2 = .03$. A significant main effect of group was observed, $F(1, 45) = 5.40$, $p = .025$, $\eta^2 = .10$, due to a larger overall rate of endorsements ("yes" responses) among dyslexic individuals relative to controls; however, this was qualified by a significant three-way interaction between group, list and chunk strength, $F(1, 45) = 5.49$, $p = .024$, $\eta^2 = .16$. Simple effect analysis revealed a group by list interaction for items in the high strength set, $F(1, 45) = 6.54$, $p = .014$, $\eta^2 = .12$, but no such interaction for items in the low chunk strength set, $F(1, 45) < 1$. We return to this result in the analyses of endorsement rates for high chunk strength items (see section: sensitivity to specific similarity).

Chunk strength sensitivity (d') and criterion c values are shown separately for each group in the right panel of Table 4.2. To explore differences in discrimination ability, d' values were analyzed by means of a 2-way ANOVA with list (1, 2) and group (skilled readers, dyslexic readers) as between-subject variables. There was no effect of list, $F(1, 45) < 1$, or group, $F(1, 45) = 1.21$, $p > .05$, $\eta^2 = 0.02$. However, the list by group interaction was significant, $F(1, 45) = 5.17$, $p = .028$, $\eta^2 = 0.02$. Simple effect analyses showed a trend for stronger discrimination accuracy among skilled readers ($n = 9$: $M = 1.26$, $SE = 0.50$) when compared to dyslexic readers ($n = 13$: $M = 0.82$, $SE = 0.47$) in list 2, which was not significant after Bonferonni correction, $t(20) = 2.13$, $p = .046$, $d = 0.92$. There was no difference between skilled ($n = 15$: $M = 0.93$, $SE = 0.13$) and dyslexic participants' discrimination ability ($n = 12$: $M = 1.08$, $SE = 0.09$) in list 1, $t(25) = 0.93$, $p > .05$, $d = 0.37$. d' values for skilled and dyslexic readers were subjected to separate one sample t tests, which confirmed that performance in both groups reliably exceeded chance levels, skilled readers: $t(23) = 9.87$, $p < .001$, $d = 2.01$, dyslexic readers, $t(24) = 11.31$, $p < .001$, $d = 2.26$.

Preliminary analyses including counterbalanced list as a factor revealed no significant main effect of list, or group by list interaction in the criterion bias c analyses (all ps > .05). Thus, mean *c* values were subjected to an independent *t* test. The analysis revealed a significant difference as a function of group, $t(47) = 2.33$, $p = .024$, $d = 0.67$ (dyslexics: $M = -0.16$, $SE = 0.05$, skilled readers: $M = 0.07$, $SE = 0.09$). A one sample *t* test indicated that dyslexic participants exhibited a reliable bias towards saying "yes", $t(24) = 3.19$, $p < .01$, $d = 0.63$, whereas skilled readers showed no evidence of response bias, $t(23) = 0.82$, $p > .05$, $d = 0.17$.

Sensitivity to letter position knowledge. Table 4.2 shows mean endorsement rates for each subset of low chunk strength items. To investigate whether participants were sensitive to the positional information manipulation, mean endorsement rates for low chunk strength items were entered into a 2 by 2 mixed factorial ANOVA with type of violation (chunk and positional violations [type I items], chunk violations [type II items]) as a within-subject variable and group (skilled readers, dyslexic readers) as a between-subject subject variable. The ANOVA revealed no main effect of type of violation, $F(1, 47) < 1$, or group by type of violation interaction, $F(1, 47) < 1$. That is, unlike Kinder and Lotz (2009), participants committed the same proportion of false alarms, regardless of whether items respected or violated positional information. The only significant effect was the main effect of group, due to a higher overall rate of false alarms in the dyslexic group ($M = .39$, $SE = 0.03$) when compared to skilled reader group ($M = .29$, $SE = 0.03$), $F(1, 47) = 5.60$, $p = .022$, $\eta^2 = .11$, indicating—consistent with the bias *c* analyses—that dyslexic participants endorsed more items of low associative chunk strength.

Sensitivity to specific similarity. To assess participants' sensitivity to specific similarity, mean endorsement rates for high chunk strength items (Table 4.2) were subjected to a 3-way list (list 1, list 2) by similarity (similar [type III items], dissimilar [type IV items]) by group (skilled readers, dyslexic readers) ANOVA. Consistent with our hypothesis, there was a significant main effect of similarity, $F(1, 45) = 9.27$, $p < .01$, $\eta^2 = .16$, indicating that participants endorsed type IV (similar) items ($M = .74$, $SE = 0.02$) more frequently than type III (dissimilar) items ($M = .66$, $SE = 0.02$). A significant similarity by group interaction would have indicated differences in skilled readers' and dyslexic participants' sensitivity to specific similarity. However, this interaction was not significant, $F(1, 45) = 3.35$, $p > .05$, $\eta^2 = .06$. The similarity by list, and, similarity by group by list interactions were not significant, $F(1, 45) < 1$, and $F(1, 45) = 1.80$, $p > .05$, $\eta^2 = .03$, respectively, indicating that our key experimental manipulations were unaffected by the manipulation of counterbalanced list. The main effects of group, *F*1, 45) = 2.11, $p > .05$,

η^2 = .04, and list, $F(1, 45) < 1$, were not significant; however, consistent with the three-way interaction obtained in the mean endorsement rate analyses, there was a significant list by group interaction, $F(1, 45) = 6.56$, $p = .014$, $\eta^2 = .12$. The interaction occurred because the main effect of group was significant in list 2, $t(17.91) = 3.11$, $p = .006$, $d = 1.15$, with higher proportions of "yes" responses in dyslexic ($M = .77$, $SE = 0.02$) when compared to skilled readers ($M = .62$, $SE = 0.05$), but not list 1, $t(20) = 0.81$, $p > .05$, $d = 0.34$. Thus, dyslexic participants showed a general tendency towards responding "yes", which held true across both types of items (high strength, low strength) and lists (1, 2) except for responses to the subset of high strength items in list 1. This interaction does not affect the conclusion that both groups exhibited reliable sensitivity to items that were similar to specific training items; however, it is taken into account in the discussion of dyslexic participants' tendency to endorse items as grammatical.

Table 4.2. Proportion of Items Endorsed as Grammatical (*SD*s) by Skilled and Dyslexic Readers as a Function of Stimulus Type and the Resulting Signal Detection Theory Measures (Experiments 3.1 – 3.2).

Group	Endorsement rates						SDT measures	
	Low Chunk Strength (FAs)			High Chunk Strength (Hits)			*d'* ± ME	Bias *c* ± ME
	Type I	Type II	Mean	Type III	Type IV	Mean		
Skilled readers	.27 (0.14)	.32 (0.17)	.30 (0.14)	.60 (0.17)	.72 (0.19)	.66 (0.17)	1.05 ± 0.22	0.07 ± 0.18
Dyslexic readers	.39 (0.19)	.38 (0.14)	.39 (0.12)	.71 (0.13)	.74 (0.16)	.73 (0.09)	0.94 ± 0.17	-0.16 ± 0.10

Note. Skilled adult readers: $n = 24$. Dyslexic readers: $n = 25$. Type I items are low chunk strength items which contain bigram/trigram violations as well as positional violations. Type II items are low chunk strength items which contain bigram/trigram violations only. Type III items are high chunk strength items that differ from any training item by 3 letters or more, i.e., they are relatively dissimilar to the items used in the memorization phase. Type IV items are high chunk strength items that differ from the closest training item by one letter, i.e., they are similar to a specific training item.
FA= False Alarms. ME = Margin of Error. SDT = Signal Detection Theory.

4.2.4.3 Correlational Analyses

To examine the relationship between artificial grammar learning and the various cognitive and literacy measures, Pearson's correlations were calculated between the measures of memorization/test phase task performance and the remaining variables (inter correlations among the measures are shown in Appendix I [skilled readers] and J [dyslexic readers]). Separate analyses were carried out for skilled and dyslexic readers. We considered pooling across the two groups of participants; however, this was deemed statistically inappropriate given the differences that emerged between skilled and dyslexic readers' performance on some aspects of the artificial grammar learning task (e.g., memorization RTs). What is more, pooling revealed stronger associations very occasionally, and in fact, sometimes obscured correlations between variables that were significant for one of the two groups of participants. For these reasons, we do not report results from the analyses across groups.

An initial examination of the distributional properties of all variables revealed a few participant outliers and some departures from normality. To reduce the influence of outliers, scores that deviated from the sample mean by 2.5 *SD*s were replaced by the values of the sample mean ± 2.5 *SD*s. In all cases, this provided a better approximation to a normal distribution. However, performance on some of the measures was still somewhat not normally distributed. The distribution of skilled readers' response latencies in the nonword phoneme deletion task was positively skewed (skew = 1.35, *SE* = 0.48) and leptokurtic (kurtosis = 3.65, *SE* = 0.93). With regards to dyslexic readers' performance, positive skews and leptokurtosis were observed in the distribution of response latencies in the RAN digits (skew = 1.84, *SE* = 0.46; kurtosis = 3.32, *SE* = 0.90) and RAN objects task (skew = 1.30, *SE* = 0.46; kurtosis = 2.05, *SE* = 0.90), whereas performance on the Symbol Digit Modalities Test was negatively skewed (skew = -1.29, *SE* = 0.46). Variables that exhibited positive skews were log transformed prior to correlational analyses. To normalize the distribution of dyslexic readers' performance on the Symbol Digit Modalities Test, scores were first reflected and then transformed by replacing raw scores with their square roots.

Relationship between *d'* scores and memorization performance. The analyses of phase 1 (training) performance revealed that dyslexic participants devoted significantly more time, relative to skilled readers, to the memorization of strings. What is more, there were marginal differences in the number of training trials skilled versus dyslexic participants received. These findings motivated us to take a closer look at memorization phase performance and its relation to test phase performance in each group of participants.

An initial analysis of the relationship between the two indexes of memorization accuracy (proportion of strings reproduced correctly within a single attempt, number of trials to criterion) revealed that the two measures were extremely highly negatively correlated among skilled readers, $r(22) = -.95$, $p < .001$, and among dyslexic readers, $r(23) = -.96$, $p < .001$, such that the higher the proportion of 1^{st} attempt correct responses, the fewer the number of trials to criterion. Given the high correlation between the two accuracy indexes, a composite memorization accuracy score was derived by summing the *z* scores for the proportion accuracy/attempt and the mean number of trials to criterion (whose direction was reversed so that higher values indicated higher accuracy). Pearson's correlations between the composite score of memorization accuracy, mean correct memorization RTs and *d*' values are reported in Table 4.3. Skilled readers' reaction times tended to show a relationship with memorization accuracy, such that longer RTs were associated with higher accuracy (i.e., there was a trend for a classic speed-accuracy trade-off). On the other hand, there was no evidence of a relationship between correct memorization RTs and memorization accuracy in dyslexic participants' performance.

Table 4.3. Correlations Between *d*' scores in Experiments 3.1 – 3.2, the Composite Measure of Memorization Accuracy, and Memorization RTs

Variable	1	2	3
Skilled readers (n = 24)			
1. *d*' scores	—	.30	.26
2. Memorization ACC		—	.39
3. Memorization RTs			—
Dyslexic readers (n = 25)			
1. *d*' scores	—	.26	.13
2. Memorization ACC		—	.06
3. Memorization RTs			—

There are no previous studies investigating the extent to which performance during the memorization phase of the experiment is linked to participants' ability to judge the well-formedness of the stimuli presented at test. However, it is reasonable to assume a link between performance in the two phases, such that, for example, better memorization performance would be associated with an enhanced ability to discriminate between high strength/grammatical and low strength/ungrammatical test phase items. Some indirect evidence in favour of our hypothesis is provided by Karpicke and Pisoni (2004). Using a continuous acquisition/test phase design throughout which participants were simply asked to reproduce 4- to 10-item long sequences (e.g., visuospatial sequences of colored lights), Karpicke and Pisoni (2004) demonstrated that, following a period of "exposure" (i.e., reproduction) to grammatical sequences, participants were better at reproducing novel grammatical sequences than novel ungrammatical sequences. Critically, individual differences in learning ability (calculated by partialling out accuracy scores for ungrammatical sequences from accuracy scores for grammatical sequences; *immediate memory span score*) showed a significant positive relationship with participants' ability to reproduce the sequences in the acquisition phase.

This was not the case in our study whereby learning skill was assessed by means of the standard well-formedness task. As can be seen from Table 4.3, the composite score of memorization accuracy was weakly and nonsignificantly associated with *d*' scores in the skilled reader group analysis and was unrelated to dyslexic readers' discrimination ability. We also sought to examine the relationship between memorization RTs and test phase performance. For example, were longer memorization RTs positively associated with better test phase performance? That was not the case, either. Mean correct memorization RTs were unrelated to *d*' scores across both analyses. Thus, at least in an artificial grammar variant where stimuli were memorized to criterion, there was no evidence of a positive relationship between memorization performance and participants' ability to discriminate between well- and ill-formed test phase items. It is worth speculating whether such an association would have emerged if participants were exposed to a fixed number of trials, which would presumably increase the variability of participants' discrimination ability.

Relationship between *d*' scores and the cognitive/literacy measures. Of particular interest to the study were any associations between performance on the well-formedness task and (a) the literacy and phonological processing measures, (b) the IQ measures, and (c) the verbal short-term memory measure. The full set of correlations between *d*' scores and the cognitive and literacy measures are shown in Table 4.4.

There was no evidence of a relationship between *d*' scores and WRAT Reading or Spelling performance in any of the analyses; however, a significant negative correlation (shown in Figure 4.3) emerged between dyslexic readers' response latencies in the nonword phoneme deletion task and *d'* values, such that slower phoneme deletion performance was associated with better discrimination ability. Similarly, a trend for a positive relationship between (longer) RAN Digits latencies and higher discrimination ability was observed among the same group of participants.

Consistent with previous studies, full-scale IQ was unrelated to *d*' scores in the dyslexic reader group analyses. Among skilled readers, a trend for a negative relationship between general psychometric intelligence and test phase performance emerged (r = -.31, p = .144). A clearer association along the lines of a speed-accuracy trade off was observed between skilled readers' *d*' scores and Symbol Search performance: Slower processing speed was related to higher discrimination ability in the experimental task (scatter plot shown in Figure 4.3).

Digit Span performance did not correlate with *d*' scores in the skilled readers sample analyses. A trend for an unexpected negative relationship emerged in the dyslexic sample analyses, such that lower Digit Span performance was associated with higher discrimination ability in the experimental task. There were no other significant correlations or trends between *d*' values and performance on the cognitive and literacy measures.

Table 4.4. Correlations between Discrimination Accuracy, *d*' scores, in Experiments 3.1 – 3.2 and Performance on the Cognitive and Literacy Measures.

Variable	Skilled readers (*n* = 24)	Dyslexic readers (*n* = 25)
WRIT Full-scale IQ[a]	-.31	-.06
WRIT Verbal IQ[a]	-.14	-.06
Verbal Analogies[b]	.19	-.10
Vocabulary[b]	-.23	-.22
WRIT Nonverbal IQ[a]	-.27	-.03
Matrices[b]	-.27	.11
Diamonds[b]	-.12	-.16
WRAT Reading[b]	.01	-.23
WRAT Spelling[b]	.06	-.25
WAIS Digit span[b]	.07	-.33
WAIS Symbol Search[b]	**-.50***	-.11
SDMT	.15	-.10
RAN digits mean time[c]	.13	.35
RAN objects mean time[c]	.11	.32
NWPD latencies[c,d]	.14	**.40***

Note. RAN= Rapid Automatized Naming. SDMT = Symbol Digit Modality Test. WAIS = Wechsler Adult Intelligence Scale; WRAT = Wide Range Achievement Test; WRIT = Wide Range Intelligence Test.
[a]Standard scores. [b]Raw scores. [c]In seconds. [e]skilled readers: n = 23 due to missing data from one participant
* $p < .05$

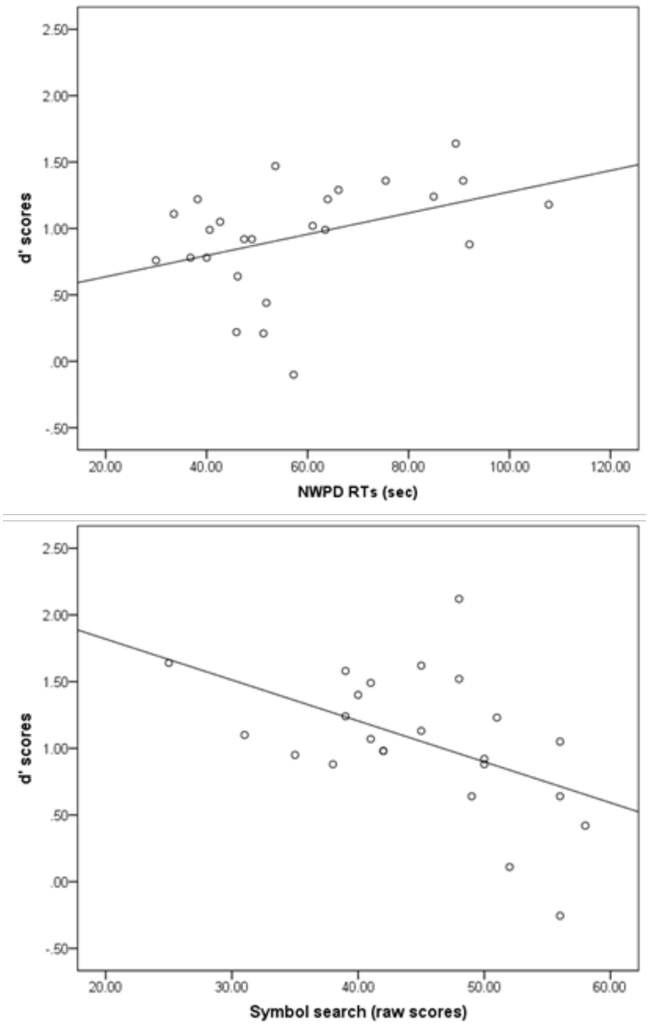


Figure 4.3. Scatter plot showing (a) a significant positive relationship between dyslexic readers' performance on the artificial grammar learning task (*d*' scores) and nonword phoneme deletion latencies; (b) a significant negative relationship between skilled readers' performance on the artificial grammar learning task (*d*' scores) and performance on the Symbol Search task

Relationship between memorization performance and the cognitive/literacy measures. A final series of correlations examined the relationship between performance in the memorization phase of the experiment and the cognitive and literacy measures. Given that Kinder and Lotz's (2009) task assessed rote memorization skill, we anticipated a different pattern of correlations with memorization performance relative to those observed with performance on the well-formedness task (see Reber et al., 1991 for a similar point). The full set of correlations between the composite memorization accuracy score and the various cognitive and literacy measures is shown in Table 4.5.

In skilled readers, the analyses revealed (a) a significant positive correlation with performance on the Verbal Analogies subtest of the WRIT, (b) a trend for a positive relationship with short-term memory ability ($r = .40$), which just failed to reach significance, $p = .054$, and (c) a trend for a positive relationship with performance on the Diamonds subtest of the WRIT, $r = .36$, $p = .085$. There were no trends or significant correlations between dyslexic students' cognitive, literacy or literacy-related scores and memorization accuracy.

Table 4.5. Correlations between the Composite Score for Memorization Accuracy in Experiments 3.1 – 3.2 and Performance on the Cognitive and Literacy Measures.

Variable	Skilled readers (*n* = 24)	Dyslexic readers (*n* = 25)
WRIT Full-scale IQ[a]	.25	-.03
WRIT Verbal IQ[a]	.17	-.03
Verbal Analogies[b]	**.51***	.01
Vocabulary[b]	.15	-.18
WRIT Nonverbal IQ[a]	.18	-.02
Matrices[b]	-.09	-.21
Diamonds[b]	.36	.13
WRAT Reading[b]	.09	.06
WRAT Spelling[b]	.03	.03
WAIS Digit Span[b]	.40	-.04
WAIS Symbol Search[b]	-.01	.16
SDMT	-.08	.21
RAN digits mean time[c]	.07	.11
RAN objects mean time[c]	-.19	.17
NWPD latencies[c,d]	.13	-.09

Note. RAN = Rapid Automatized Naming. SDMT = Symbol Digit Modality Test. WAIS = Wechsler Adult Intelligence Scale; WRAT = Wide Range Achievement Test; WRIT = Wide Range Intelligence Test.
[a]Standard scores. [b]Raw scores. [c]In seconds. [d]skilled readers: n = 23 due to missing data from one participant
* $p < .05$

Next, we examined the relationship between mean correct memorization RTs and the cognitive and literacy measures (Table 4.6). Both groups' memorization RTs were significantly negatively correlated with performance on the Symbol Digit Modalities test. That is, higher graphomotor speed of processing skill was associated with faster memorization RTs. Individual differences in dyslexic readers' nonverbal reasoning skill, measured in terms of performance on the Diamonds subtest of the WRIT, were negatively associated with memorization RTs. A similar negative association was found with full-scale IQ and with nonverbal IQ, indicating a positive link between psychometric intelligence and memorization speed; however, when Diamonds performance was partialled out, the association with nonverbal IQ and full-scale IQ no longer remained, $r_1 = -.06$, $p = .768$, $r_2 = -.22$, $p = .312$, suggesting that the association was driven by nonverbal ability rather than general ability performance.

Table 4.6. Correlations between Mean Correct Memorization RTs in Experiments 3.1 – 3.2 and Performance on the Cognitive and Literacy Measures.

Variable	Skilled readers (n = 24)	Dyslexic readers (n = 25)
WRIT full-scale IQ[a]	.04	**-.43***
WRIT verbal IQ[a]	-.15	-.17
Verbal Analogies[b]	.12	-.25
Vocabulary[b]	-.07	-.22
WRIT nonverbal IQ[a]	.18	**-.48***
Matrices[b]	.01	-.13
Diamonds[b]	.26	**-.61****
WRAT Reading[b]	-.19	-.26
WRAT Spelling[b]	-.31	.18
WAIS Digit Span[b]	-.08	-.27
WAIS Symbol Search[b]	-.19	-.32
SDMT	**-.47***	**-.40***
RAN digits mean time[c]	.07	.34
RAN objects mean time[c]	.32	.27
NWPD latencies[c,d]	.28	.32

Note. RAN = Rapid Automatized Naming. SDMT = Symbol Digit Modality Test. WAIS = Wechsler Adult Intelligence Scale; WRAT = Wide Range Achievement Test; WRIT = Wide Range Intelligence Test.
[a]Raw scores. [b]Scaled scores. [c]In seconds. [d]skilled readers: n = 23 due to missing data from one participant
* $p < .05$. ** $p < .01$.

4.3 General Discussion

Previous research investigating implicit skill learning in dyslexia has been inconclusive, reporting spared (e.g., Kelly et al., 2002; Rüsseler et al., 2006) as well as impaired (e.g., Howard et al., 2006; Pavlidou et al., 2010) performance on different, though well-established, implicit learning tasks. The main aim of the present study was to revisit implicit knowledge acquisition in dyslexic adults (university undergraduate students) by means of an artificial grammar learning task. Our study replicated Kinder and Lotz's (2009) material, design and procedure, which manipulated three test item properties, namely (a) their associative chunk strength, (b) their adherence to positional constraints on individual letters set by the grammar, and (c) their similarity to specific training items. Following memorization and correct reproduction of a small set of training letter strings, participants were presented with a well-formedness task assessing sensitivity to such aspects of stimulus structure. Learning of permissible and frequent string fragments (bigrams, trigrams) was assessed by comparing participants' endorsement of low versus high associative chunk strength strings. Sensitivity to positional constraints was assessed by comparing sensitivity to two subsets of low chunk strength strings, only half of which introduced violations on the absolute allowable position of letters. Finally, participants' sensitivity to whole-training strings was assessed by comparing sensitivity to two subsets of high chunk strength strings, only half of which deviated from their closest training string by one letter and were, therefore, thought to be similar to them (specific similarity, as originally studied by Brooks & Vokey, 1991).

Following previous studies (e.g., Pavlidou et al., 2010; Rüsseler et al., 2006), we began by considering differences between skilled and dyslexic readers' performance in the memorization phase of the experiment. Mean accuracy of string reproduction within participants' first attempt, serving as an index of successful memorization performance, was reasonably high in both of the skilled and dyslexic reader group analyses; however, we observed a trend for group differences in the analyses of the number of trials to criterion, such that dyslexic participants tended to need more trials to correctly reproduce each training string. The main effect of group was significant (and the effect size was large) in the analyses of mean correct memorization RTs, showing that longer exposure was necessary for dyslexic relative to skilled readers for the purpose of memorizing and reproducing the strings correctly. Our findings are inconsistent with Rüsseler et al. (2006), who reported no differences in memorization ability between skilled and dyslexic participants drawn from a similar population to ours. However, Rüsseler et al.'s (2006) participants were trained with 4-letter long strings, and memorization RT data were not collected. The observed trends and significant group differences in our study suggest that, even though dyslexic readers reached criterion with only a small number of additional trials, their trajectory to memorization phase completion was probably somewhat different from that of skilled readers.

In contrast to the above differences in memorization performance, there was no evidence to suggest that dyslexic readers differed from their nonimpaired counterparts in terms of chunk strength sensitivity. Our analysis of participants' *d*'scores—an *unbiased* measured of participants' sensitivity—confirmed that both skilled and dyslexic readers reliably discriminated between high and low chunk strength items with an overall level of performance of similar magnitude to that commonly found in the implicit learning literature. Dyslexic readers' performance was not only significantly above chance, but was also not statistically different from that of skilled readers. Thus, chunk sensitivity, at least via memorization to criterion, leads to comparable levels of discrimination ability in skilled and dyslexic adults. We should note, though, that a significant difference had emerged in the analysis of participants' response bias, such that only dyslexic participants showed a significant liberal bias towards accepting items (and consistently, they endorsed more items as "grammatical" than controls; this was not true, though, with regards to the endorsement of list 1 high strength items, but we do not have a ready explanation for this, probably spurious, group by list interaction operating among high strength items). The demonstration that implicit chunk strength sensitivity—that is, in effect, frequency-based sensitivity—is spared in the context of dyslexia is important for ruling out a stastical learning deficit interpretation of dyslexic readers' literacy difficulties. The ability to process words in terms of common letter chunks or spelling units (e.g., ending chunks such as _ell, shared between words such as b*ell*, s*ell*, t*ell*) is widely acknowledged as an important aspect of skilled word reading (e.g., Ehri, 1994, 1995; Frith, 1985) and spelling (Nation, 1997). What is more, developing readers' and spellers' increasing reliance on letter chunks—rather than individual graphemes—has been sometimes attributed to implicit skill learning (Nation, 1997). The present study demonstrates, that at least in compensated dyslexic adults and for learning under laboratory conditions, the ability to detect chunk frequency information is spared in dyslexia.

Beyond demonstrating whether learning is spared or impaired as predicted by the automaticity deficit hypothesis, we were interested in assessing how would an implicit learning deficit impact dyslexic individuals' sensitivity to frequency-based information that is widely embedded in written language. Thus, going further than previous studies, we considered group differences in skilled versus dyslexic readers' sensitivity to the additional sources of information provided by the grammar: constraints on where single letters occurred and specific similarity. Replicating Kinder and Lotz's (2009) study, our results revealed a significant main effect of similarity to specific training items, due to a higher proportion of "yes" responses for high chunk strength items that were similar to those presented during the memorization phase (e.g., RTGMTX*J* vs. RTGMTX*L*) than high chunk strength items that were dissimilar from their closest training item (e.g., *JS*GMTX*H* vs. *RT*GMTX*L*). Sensitivity to specific similarity—tapping indirectly, and perhaps only to some extent, on the ability to read/spell by an analogical inference strategy (e.g., Ehri, 1997; Goswami, 1988)—influenced skilled and dyslexic adult readers' classification performance to the same extent. However, regardless of group, Kinder and Lotz's (2009; Kinder, 2000) finding that positional information influences participants' endorsement of low chunk strength items was not replicated. That is, both skilled and dyslexic readers' ability to correctly reject items was not significantly enhanced by the presence of positional violations. Given that all aspects of our study were identical to Kinder and Lotz (2009), it is difficult to account for this finding. However, it is worth speculating whether sensitivity to positional information would emerge had this manipulation been made within a set of higher chunk strength items[9]. Increasing the mean chunk strength of a string embedding letters in illegal positions is, unfortunately, far less than straightforward, at least within Kinder and Lotz's (2009) grammar. To name one example, one way of doing so involes violating positional constraints in anchor locations (e.g., position 5 and 6), such that the global associative chunk strength of the string suffers relatively less than when illegal letters are inserted in medial positions. However, it is obvious that this type manipulation would result in anchor associative chunk strength differences between type I versus type II items, that is, the critical comparison would not be warranted. It is also very interesting to consider whether constraints on the allowable position

[9] Arguably, a more adequate test of participants' sensitivity to positional information over and above chunk strength sensitivity requires an orthogonal manipulation of this source of knowledge. However, inserting letters in illegal positions renders such test phase stimuli ungrammatical. Thus, even though it's possible to manipulate adherence to positional information within items of similar associative chunk strength, it is not possible to do so within high strength/grammatical and low strength/ungrammatical items, as was the case in Kinder and Lotz's (2009) design.

of individual letters would be extracted and used in the endorsement of low strength items following a modification of the memorization phase demands. An attempt along these lines is introduced in one of the experiments reported in Chapter 5.

Several factors need to be taken into account when considering the contrast between the results of the current and previous investigations of implicit skill learning in dyslexia. In virtually all of the studies reviewed earlier, discrepant findings (impaired learning, spared learning) have been partially attributed to differences in methodological study aspects (e.g., differences in the cognitive demands posed by different tasks) or heterogeneity of dyslexic samples. We agree that one source of such inconsistent results likely comes from methodological differences between different implicit learning tasks and highlight the need for learning to be assessed in nonmotor implicit learning tasks. In terms of differences in sample characteristics, our study is generally comparable with previous studies carried out with dyslexic adults (e.g., Howard et al., 2006). However, it is important to note that in our experimental investigations, participants with co morbid disorders were excluded to avoid moderation of implicit skill learning by attention, a factor that may be required for above chance artificial grammar learning (e.g., Tanaka, Kiyokawa, Yamada, Dienes, & Shigemasu, 2008; Eitam, Schul, & Hassin, 2009).There is little doubt that our dyslexic sample was relatively high functioning, as evident from their performance within the average range on a standardized measure of reading and spelling. However, both reading and spelling skill was significantly impaired relative to age-matched controls and phonological processing deficits—the hallmark of dyslexia (Snowling, 2000)—were observed. Furthermore, our dyslexic sample showed impairments relative to their skilled reader counterparts in speed of processing skill, and more importantly, in verbal short-term memory—another well-documented area of difficulty for dyslexic individuals (although their mean scores were well within the population average range) (e.g., Hulme & Snowling, 2009; Ramus et al., 2003). Thus, although further research is needed to establish whether implicit skill learning is spared in a representative sample of adult dyslexic individuals, our findings suggests that this is the case in relatively compensated high-functioning dyslexic individuals.

4.3.1 Correlational Analyses

Our design included a large battery of cognitive and literacy measures making it possible to investigate the association of artificial grammar learning performance with a number of factors that have been previously shown to be related or unrelated to implicit skill learning. As hypothesized by Reber et al. (1991), performance during the memorization phase of the experiment showed significant correlations with several cognitive measures. Analyses among skilled readers revealed a positive association between the composite measure of memorization accuracy and verbal reasoning skill, as well as a trend for a positive relationship ($r = .40$) with verbal short-term memory. Memorization speed among dyslexic individuals was positively associated with nonverbal reasoning skill, a finding which, if replicated, may indicate a compensatory role of nonverbal upon impaired verbal reasoning skill in dyslexic individuals. Furthermore, both groups show similar positive associations between graphomotor speed, measured in terms of performance on the Symbol Digit Modalities Test, and memorization speed. This last finding may suggest that this basic speed of processing ability underpins and contributes similarly among literacy impaired/unimpaired adults to the speed of memorization of letter strings generated from an artificial grammar learning. The significant correlation between graphomotor speed of processing and memorization speed may also reflect that some task variance is shared between the two measures, in that they both require the rapid processing of alphanumeric symbols.

In contrast to our hypothesis, our correlational analyses revealed that discrimination ability at test showed little relation to performance in the vast majority of our literacy and literacy-related measures (with the exception of a significant negative correlation between dyslexic participants' latencies on the nonword phoneme deletion task and *d*' scores; to anticipate, this finding was not replicated in the following experiments reported in Chapter 5). What is more, *d*' scores showed only a weak relation to memorization performance. The amount of time dyslexic participants spent memorizing the strings was unrelated to their performance in the well-formedness task, and there was only a trend for a positive link between skilled readers' ability to reproduce the strings and their test phase performance. Finally, we obtained no evidence of a relationship between skilled readers' performance on a verbal short-term memory task (digit span) and *d*' values. This finding is consistent with previous studies showing that artificial grammar learning is minimally related to individual differences in working memory ability (e.g., Gebauer & Mackintosh, 2007; Robinson, 2005; see also, Kaufman et al., 2010; Tagarelli, Borges-Mota, & Rebushat, 2011, for similar findings using other implicit learning tasks).

Turning to the relationship between artificial grammar learning performance and intelligence, recent studies appear to converge on the conclusion that implicit learning is, at best, weakly related to psychometric intelligence (e.g., McGeorge et al., 1997; Gebauer & Mackintosh, 2007; Reber et al., 1991). This is in line with Reber's (1993) original view that implicit learning is relatively immune to individual difference in psychometric intelligence. In our study, individual differences in dyslexic participants' full-scale, verbal and nonverbal IQ were uncorrelated with their ability to judge the grammaticality of the test phase items. However, we observed a small trend for a negative relationship with full-scale IQ in the skilled reader sample analyses. Our conclusions with regard to this finding are relatively modest. A similar negative correlation between artificial grammar learning test phase performance and IQ—measured by performance on the WAIS-R—was found by Robinson (2005), who tentatively attributed the relationship to a relatively unsuccesful analytic problem-solving approach adopted by participants with higher, but not lower, performance on the IQ measure. However, it is unclear whether this explanation fully accounts for our findings. For example, Robinson (2005) assessed learning by means of a change-letter-set (i.e., transfer) procedure requiring participants to judge the grammaticality of letter strings instantiated with a different set of letters than those used during training. As briefly discussed in the Literature Review, this is a more demanding version of the artificial grammar learning task, whereby discrimination ability is frequently lower than that observed in nontrasnfer tasks.

Motivated by previous studies arguing for a need to go beyond the use of a single IQ score (e.g., Kaufman et al., 2010), we also examined correlations between test phase performance and speed of processing skill among skilled and dyslexic readers. Our skilled reader sample analyses revealed a significant negative relationship between implicit skill learning and speed of processing measured by the Symbol Search subtest of the WAIS in skilled readers. That is, worse (i.e., slower) performance on the Symbol Search task was associated with higher discrimination ability. This finding is the opposite of Kaufman et al. (2010) demonstration that a latent variable of speed of processing skill (constructed from verbal, figural and numerical speed indexes) correlates positively with performance on a probabilistic serial reaction time task. Before drawing conclusions regarding this discrepancy (e.g., is it due to differences in the way learning is measured in the artificial grammar learning and the serial reaction time task?), a replication of our finding is required.

4.3.2 Concluding Summary

The results of the present study suggest that most aspects of performance on an implicit learning task (discrimination ability and sensitivity to chunk frequency information, sensitivity to specific similarity; but not memorization performance) are spared in dyslexic adults relative to skilled readers matched for age, verbal- , nonverbal- and full-scale IQ. This is one of the few studies investigating implicit learning in dyslexia, and one of the first studies assessing learning in terms of performance on an artificial grammar learning task. Among the strengths of the current study is the relatively large sample size of dyslexic individuals after controlling, to the best of our ability, for the presence of comorbid conditions such as attention deficit disorder. Even though we obtained no evidence of a profound implicit learning deficit, our finding of relatively spared learning among dyslexic adults is open to replication. There are several other interesting findings (e.g., the negative association between artificial grammar learning performance and speed of processing skill) that warrant further investigation. Finally, our results do not provide support for the theoretical assertion that variability in implicit learning is positively associated with variation in reading or spelling performance. We return to evaluate this hypothesis in experiments 4.3 – 4.4, which assess similar patterns of correlations in a nonlinguistic (shapes) variant of the artificial grammar learning task.

Chapter 5

Artificial Grammar Learning: Effects of Processing Task Demands and Stimulus Format

5.1 Introduction

Experiments 3.1 – 3.2 assessed skilled and dyslexic readers' sensitivity to three aspects of stimulus structure (chunk strength, letter positions, and specific similarity) that allow successful discrimination in the artificial grammar learning task. We partially replicated Kinder and Lotz (2009) by showing that skilled readers' classifications (as well as dyslexic readers' classifications in our study) favored (a) test strings that comprised frequent/allowable training bigrams and trigrams, (b) strings that were similar to those presented for training. However, we did not find evidence of learning about the allowable position of individual letters (e.g., G is allowed in position 3, but not in position 4) —at least not in low chunk strength stimuli—although this factor exerted a small but significant influence on Kinder and Lotz's (2009; Kinder, 2000) participants' judgments of grammaticality. Given that the hypothesized type I over type II endorsement advantage that would have indicated learning of positional constraints on individual letters was not supported in Experiments 3.1 – 3.2, in this study, we sought to induce such learning by manipulating the processing demands of the training phase procedure. We anticipated on the basis of the episodic processing framework of artificial grammar learning (e.g., Whittlesea & Dorken, 1993) that positional encoding would be enhanced in a task emphasizing letter order information and with less load upon working memory (drag and drop procedure), rather than a recall task emphasizing both letter order and letter identity information (typing procedure).

5.1.1 The Episodic-Processing Theory of Artificial Grammar Learning

For several decades, implicit learning has been regarded as a mode of passive and unselective learning, whereby, participants' intentions and mental operations during the training phase have little to do with their impressive learning abilities (e.g., Hayes & Broadbent, 1988; Lewicki & Hill, 1989: in Whittlesea & Wright, 1997). Under this widely held assumption, it is not surprising that the induction (i.e., training phase) task has been mainly used as a means of misleading participants as to the true purpose of the experiment, and very few studies have expressed an interest in manipulating some of its aspects. Perhaps the only exceptions are (a) early studies examining *how much* is learnt under incidental (e.g., memorization) versus deliberate rule-searching instructions (among which, Reber, 1976; Reber, Kassin, Lewis, & Cantor, 1980) and (b) and, only more recently, studies examining *what* is learnt (e.g., rules vs. training/test-phase-item similarity) under such instructional sets (e.g., Mathews et al., 1989; Shanks, Johnstone, & Staggs, 1997; Johnstone & Shanks, 2001). To quote Reber and Allen (1978), "*…other than this important factor, it does not seem to make much difference how the subject is exposed to the material, although a firm attentional focus is certainly required*" (p. 190).

There is little doubt that learning is, to a great extent, driven by structural stimulus properties. However, according to the episodic-processing theory of artificial grammar learning, test phase discrimination performance is also dependent on different training phase parameters: (a) the processing undertaken during the acquisition phase and whether it overlaps with the processing undertaken at test (Whittlesea & Dorken, 1993); (b) the demands and mode of processing encouraged or imposed by the experimental training phase instructions (Wright & Whittlesea, 1998), and even more simply, (c) participants' own decisions as to how to involve themselves with the training stimuli according to the latter's perceptual manifestation (Wright & Whittlesea, 1998). In the next section, we discuss two representative experiments from Whittlesea and colleagues' pioneering work demonstrating the non-neutral nature of the induction artificial grammar learning phase.

In their first of a series of papers, Whittlesea and Dorken (1993; experiment 1) examined the effect of processing context on test phase classification performance. Participants were presented with "roughly" pronounceable letter-strings (e.g., ENRIGAD) which, unbeknown to them, were generated from two finite-state grammars; during the acquisition phase, half of strings were to be pronounced from memory and the other half were to be spelled from memory. Following completion of the acquisition phase, participants were asked to classify novel strings as “Spell” or “Pronounce”, i.e., as consistent with the “rules” of Grammar A (e.g., “spell” in one of the counterbalanced conditions) or the rules of Grammar B (i.e., “pronounce” in the same counterbalanced condition). The critical manipulation was that (a) prior to responding, participants were asked to read out half the stimuli and spell the other half, and (b) this manipulation was nested within the acquisition phase processing manipulation. For example, only half of the novel items that were generated from the Spell Grammar were also spelt prior to classification at test. Performance was significantly above chance, that is, participants were able to distinguish with better than chance accuracy between Grammar A- and Grammar B- consistent strings under the matched acquisition/test phase processing conditions (e.g., items pronounced during training *and* at test) as well as under the mismatched processing conditions (e.g., items that were pronounced during acquisition were spelled out at test). However, there was a significant advantage in performance under matched relative to mismatched conditions that was proven, in a follow-up analysis, to be unexplained by a simple response bias towards classifying the items consistently with how they were processed at test. It was therefore concluded that the formation of knowledge representation in artificial grammar learning is not independent of the processing undertaken during training. Processing aspects of the acquisition phase of the experiment are preserved and facilitate performance when they overlap with the type of processing required at test (Whittlesea & Dorken, 1993).

The second appealing claim of the episodic-processing account is that varying the demands imposed on stimuli’s organization during training will have profound consequences on how well different aspects of structural knowledge will be encoded and subsequently serve as a basis for test phase discrimination. That is, in contrast to all previous theories of artificial grammar learning reviewed in section 1.1.3, there is no “default mode” for participants to learn rules or small/large string fragments. In a simple but elegant demonstration of the “instability in structural organization”, Wright and Whittlesea (1998) examined differences in the form of knowledge brought forward to a simple but ambiguous Arabic numeral comparison task under two presentation conditions. For participants in one group, the stimuli were aligned side-by-side (i.e., horizontal alignment), whereas for participants in the other group, stimuli were aligned vertically. In both conditions, the same triplets of 4-digit numerals were presented, and a decision had to be made as to which out of two numerals (e.g., 6271 and 2438) was more similar to the third numeral in the triplet (e.g., 6438). All of the comparisons were ambiguous in that participants could respond either (a) on the basis of items' abstract numerosity (i.e., 6271 is more similar to 6438 than 2438 if one is processing them as "wholes", that is, focusing on their absolute numerical values) or (b) respond by comparing them on a digit-by-digit basis (i.e., there is more a digit overlap between 2438 and 6348 rather than 2438 and 6271). Although participants were not given any particular instruction as to how to respond and were presented with the same stimuli under the two experimental conditions, the spatial arrangement manipulation had a significant influence on their responses. The higher ease of encoding stimuli as wholes under side-by-side presentation gave rise to responses based on stimuli's more abstract properties (i.e., their numerosity). On the other hand, the higher ease of comparing numbers on a digit-by-digit basis in the horizontal alignment condition, gave rise to responses based on individual units. This is an interesting demonstration suggesting that even simple manipulations of the training phase procedure may promote the emergence of knowledge of different stimulus aspects. While the above example focused on the effect of training phase characteristics on participants' ability to learn abstract versus stimulus specific elements, we wished to examine the effect that different training phase demands may have on participants' sensitivity to another stimulus-specific aspect: their ability to learn constraints on the allowable position of letters.

5.2 Experiments 4.1 – 4.2

5.2.1 Rationale

The current study addressed the question of whether sensitivity to positional information may be induced by modifying the requirements of the string reproduction task—a task used to ensure that participants were memorizing the training strings. Our approach was driven by Whittlesea and colleagues (e.g., Whittlesea & Dorken, 1993; Wright & Whittlesea, 1998) episodic-processing account of artificial grammar learning, the only theoretical framework which ascribes training phase demands a role in what participants learn under implicit conditions. We also sought to replicate our previous finding of spared test phase performance in dyslexic adults and investigate whether their memorization difficulties would persist in the new version of task.

Participants were presented with the same stimuli as in experiments 3.1 – 3.2 and were asked to memorize them without being aware of their structured, grammar-conforming nature. However, instead of typing them from memory, they were asked to reproduce them by dragging and dropping the letters comprising each string—presented together with 3 nonrelevant (distractor) letters—into 7 response boxes. This is perceivably an easier memorization task that relies less heavily on recall. Moreover, a similar drag and drop procedure has been used to measure short-term memory capacity in immediate/delayed serial recall experiments (i.e., in drag and drop variants of the order recognition task; e.g., Beaman & Röer, 2009; Bell, Röer, & Buchner, 2013), and is known to reduce the cognitive demands placed on individual item memory and place more emphasis on order memory (Tremblay, Macken, & Jones, 2000). Along similar lines, we anticipated that our manipulation would enhance participants' sensitivity to allowable letter positions by allowing them to emphasize the retrieval of letter order, rather than the retrieval of both letter identity information and letter order (typing variant).

5.2.2 Method

5.2.2.1 Participants

Thirty-eight skilled readers (15 male; mean age = 20.51 years, range = 18.17 years – 37.83 years) and 28 students diagnosed with dyslexia (13 male; mean age = 21.40 years, range = 18.33 years – 33.17 years) took part in the study. Skilled readers were recruited through the Bangor University participant panel and received course credit for their participation. Data from 5 participants who did not meet the study's language criteria were not included in the analyses. Two skilled readers with standard scores of less than 85 on the Matrices and/or Vocabulary subtests of the WRIT, and one participant with unexpectedly low performance on the WRAT subtests (standard scores < 90) were also excluded. Dyslexic participants were recruited through the Miles Dyslexia Center, Bangor University and were paid for their participation. Twenty of them had received a formal diagnosis of dyslexia via the Miles Dyslexia Center's assessment service; the remaining eight participants were diagnosed by a qualified educational psychologist during secondary school or later. Three participants failed to complete the memorization phase of the artificial grammar learning task (i.e., reproduce all 96 training strings correctly) within a reasonable time frame (~75 minutes) and were unable to proceed to the test phase. These participants were not considered in any further analyses. In addition, we excluded data from (a) one participant who failed to comply with the study's language criteria, (b) three participants whose assessment report confirmed the presence of another comorbid developmental disorder, and (c) two participants with extreme low verbal/nonverbal IQ scores. Data from two dyslexic participants with standard scores of 83 and 84 on the Matrices and Vocabulary WRIT subtest, respectively, were retained in order to avoid further loss of statistical power. The final sample consisted of 30 skilled readers with no history of learning difficulties (13 male; mean age = 19.91 years, range = 18.25 years – 24.00 years), and 19 dyslexic readers (7 male; mean age = 20.80 years, range = 18.33 years – 33.17 years). All participants were monolingual native English speakers and had normal or corrected-to-normal visual acuity.

5.2.2.2 Background measures

As in the previous study, all participants completed a screening assessment consisting of psychometric tests, cognitive and literacy measures. The battery of background measures was as described in Chapter 4, except that, in order to reduce the length of the experimental session, (a) the Symbol Digit Modalities Test was not administered and, (b) verbal and nonverbal IQ were estimated using the Vocabulary and Matrices subtests of the WRIT.

5.2.2.3 Material

The memorization phase and test phase stimuli were as in Experiments 3.1 – 3.2. During string reproduction, participants were presented with 10 individual letters (40 ppt; Arial Font), 7 of which matched the letters of the "memorized" letter string (*target* letters), and 3 of which served as distractors (*distractor* letters). For example, in one of the trials, participants were presented with the letters {L, M, Z, Q, T, R, B, Z, K, D} and were asked to arrange them in 7 response boxes in order to reproduce the stimulus RTKMZQZ. During each trial, distractor letters were selected randomly and without replacement from the pool of consonants which did not comprise the letter string.

5.2.2.4 Procedure

As in the previous study, the artificial grammar learning task was followed by the administration of the cognitive, literacy and literacy-related measures (given in a fixed order). Testing took place in one to two sessions, each of which lasted approximately 1.5 hr.

Artificial grammar learning task. Similar to experiments 3.1 – 3.2, participants were informed that they were taking part in a short-term memory experiment. They were told that, during each trial, they would see a 7-letter string to be memorized for as long as necessary until spacebar was pressed. As in experiments 3.1 – 3.2, string presentation was followed by a 3000 ms interval (black fixation appearing at the centre of the screen). At that stage, participants were presented with 7 response boxes and a set of 10 individual letters, in the upper and lower half of the screen, respectively; they were asked to arrange (drag and drop) only the 7 relevant letters in the boxes in order to match the string they had just memorized. The left-right order of the individual letters was randomly determined for each participant during each trial. All participants were informed that (a) there was no time limit or constraints in the order by which letters could be dragged and dropped into the boxes; (b) letters could be rearranged for an unlimited number of times until a "ready to submit" button on the lower half of the screen was clicked; (c) there would be no performance feedback during string reproduction; however, if their response was incorrect, the same string would be presented again.[10] Three practice trials were administered to familiarize participants with the task instructions and requirements. As in the previous experiment, to reach criterion, participants were required to reproduce all 12 memorization phase strings, presented 8 eight times each, correctly. Completion of the memorization phase was followed by the surprise well-formedness task which was identical to that of experiments 3.1 – 3.2.

5.2.3 Data Analyses

Data analyses were the same as in experiments 3.1 – 3.2. Counterbalancing list was initially included as a between-subject factor in all of the analyses; contrary to experiments 3.1 – 3.2, there were no reliable interactions or main effects involving list in any of the test phase analyses. Therefore, test phase data were collapsed over this factor. There was a small main effect of list (p = .042) on mean correct memorization RTs, which did not interact with group (skilled readers, dyslexic readers). The effect is reported in the relevant section of the results; however, it is not further discussed as it did not affect the interpretation of our findings regarding group differences; thus, it was considered to be of little theoretical importance.

[10]Due to programming constraints, incorrect responses were paired with identical letter/distractor sets and left-right order arrangements.

5.2.4 Results

5.2.4.1 Background Measures

The means and *SD*s for skilled and dyslexic readers' performance on each one of the cognitive and literacy measures are shown in the left panel of Table 5.1. Independent *t* tests shown in the right panel of the same table confirmed that the two groups were matched for age, as well as performance on the Vocabulary and Matrices subtests of the WRIT. As in the previous study, dyslexic participants' performance fell within the average range on both WRAT subtests; however, their reading and spelling skill was significantly impaired when compared to that of skilled readers, who scored within the average and above average range of the WRAT Reading and Spelling test, respectively. Furthermore, the group of dyslexic participants showed significant impairments, relative to skilled readers, in Digit Span and Symbol Search performance, as well as phonological processing and rapid naming skill.

Table 5.1. Mean scores (*SD*s) and Ranges on the Cognitive and Literacy Related Measures as a Function of Group and Results from Independent *t* tests (Experiments 4.1 – 4.2).

Variable	Skilled readers (n = 30)		Dyslexic readers (n = 19)		t	sig.	Cohen's d
	Mean (*SD*s)	Range	Mean (*SD*s)	Range			
Handedness (r; l)	25; 5		14; 5				
Age (years)	19.91 (1.31)	18.25 – 24.00	20.80 (3.41)	18.33 – 33.17	1.09[e]	.288	0.35
WRIT Vocabulary[a]	103.60 (8.73)	85.00 – 119.00	103.11 (10.08)	84.00 – 119.00	0.18	.856	0.05
WRIT Matrices[a]	106.33 (10.67)	86.00 – 130.00	111.11 (10.84)	83.00 – 125.00	1.52	.136	0.44
WRAT Reading[a]	106.30 (7.49)	96.00 – 130.00	92.05 (8.68)	77.00 – 105.00	6.10	***	1.76
WRAT Spelling[a]	110.97 (9.01)	95.00 – 129.00	93.63 (9.77)	78.00 – 113.00	6.35	***	1.85
WAIS Digit Span[b]	11.00 (2.86)	4.00 – 16.00	8.26 (2.58)	6.00 – 17.00	3.38	**	1.00
WAIS Symbol Search[b]	13.67 (2.60)	9.00 – 19.00	11.84 (2.87)	6.00 – 19.00	2.30	*	0.67
RAN digits mean time[c]	14.33 (2.49)	9.57 – 20.60	18.72 (4.68)	12.05 – 27.72	3.77[e]	**	1.17
RAN objects mean time[c]	20.78 (2.66)	15.72 – 26.65	25.30 (5.27)	17.47 – 37.77	3.47[e]	**	1.08
NWPD accuracy[d]	10.98 (0.86)	9.00 – 12.00	9.45 (2.08)	5.50 – 12.00	3.06[e]	**	0.97
NWPD latencies[c]	38.32 (8.07)	25.49 – 60.35	69.16 (19.55)	30.00 – 109.91	6.53[e]	***	2.06

Note. r = right-handed. l = left-handed. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = NonWord Phoneme Deletion.
[a]Standard Scores. [b]Scaled scores. [c]In seconds. [d]Out of 12. [e]Correction for unequal variances applied.
* $p < .05$. ** $p < .01$. *** $p < .001$

5.2.4.2 Artificial Grammar Learning Task

Memorization phase. As in experiments 3.1 – 3.2, group differences in memorization accuracy were explored by means of independent *t* tests on the proportion of items reproduced correctly in a single attempt and the mean number of trials to criterion—after replacement of extreme outliers with the values of the sample mean ± 2.5 *SD*s. Dyslexic readers were less accurate (85%, $SD = 0.10$) in correctly reproducing the training strings within a single attempt than skilled readers (91%, $SD = 0.07$), $t(28.84) = 2.44$, $p = .021$, $d = 0.74$, and required a small but significant number of additional trials to reach criterion ($M = 1.21$, $SD = 0.16$) when compared to skilled readers ($M = 1.12$, $SD = 0.10$), $t(26.65) = 2.21$, $p = .036$, $d = 0.68$. Thus, the trends seen in the previous study (higher memorization accuracy in skilled relative to dyslexic readers) reached statistical significance in the drag and drop artificial grammar learning variant. In contrast to experiments 3.1 – 3.2, a 2-way ANOVA on mean correct memorization RTs—trimmed to the values of 2 *SD*s from each group mean—showed no effect of group, $F(1, 45) = 1.24$, $p > .05$, $\eta^2 = .02$, indicating that the difference between skilled ($M = 5557.75$, $SE = 381.85$) and dyslexic ($M = 6239.67$, $SE = 480.48$) readers' memorization RTs was not significant. In addition, there was a small but significant main effect of list on memorization RTs, $F(1, 45) = 4.38$, $p = .042$, $\eta^2 = .09$, with longer RTs in list 1 ($M = 6541.24$, $SE = 440.92$) when compared to list 2 ($M = 5256.19$, $SE = 426.92$). There was no group by list interaction, $F(1, 45) < 1$, confirming that the lack of group difference was not dependent on list.

Test phase. All data points were included in the test phase analyses. Mean endorsement rates for high and low chunk strength items are shown separately for skilled and dyslexic readers in the left of panel of Table 5.2. A two-way ANOVA with group as a between-subject factor and chunk strength as a within-subject factor revealed a significant main effect of chunk strength, $F(1, 47) = 176.67$, $p < .001$, $\eta^2 = .79$, with higher endorsement rates for high chunk strength ($M = .70$, $SE = 0.02$) than low chunk strength items ($M = .35$, $SE = 0.02$). There was no group by chunk strength interaction, suggesting that chunk strength modulated endorsement rates to the same extent in skilled and dyslexic readers, $F(1, 47) < 1$, and no main effect of group, $F(1, 47) < 1$.

Chunk strength sensitivity (*d*' scores) and response bias (*c* values) are shown separately for each group in Table 5.2. An independent *t* test confirmed that discrimination ability was not statistically different between skilled and dyslexic readers, $t(47) = 0.34$, $p >$

.05, $d = 0.09$. Both groups performed significantly better than expected by chance, skilled readers: $t(29) = 10.31$, $p < .001$, $d = 1.88$; dyslexic readers: $t(18) = 7.46$, $p < .001$, $d = 1.71$. Bias c values did not vary as a function of group, $t(47) = 0.60$, $p > .05$, $d = 0.17$. There was no evidence of bias in skilled readers' responses, $t(29) = 1.46$, $p > .05$, $d = 0.27$, or dyslexic readers' responses, $t(18) = 1.67$, $p > .05$, $d = 0.38$.

Sensitivity to letter position knowledge. Mean endorsement rates for low chunk strength type I and type II items (shown in Table 5.2) were subjected to a 2 by 2 mixed factorial ANOVA with type of violation as a within-subject variable and group as a between-subject variable. The main effect of type of violation was significant, $F(1, 47) = 11.01$, $p = .002$, $\eta^2 = .19$, and in the predicted direction. That is, participants committed a higher proportion of false alarms for type II items that contained bigram/trigram violations ($M = .40$, $SE = 0.02$) than type I items that contained both bigram/trigram and positional violations ($M = .31$, $SE = 0.02$). There was no type of violation by group interaction and no effect of group, both Fs$(1, 47) < 1$.

Sensitivity to specific similarity. A 2 by 2 factorial ANOVA on mean endorsement rates for high chunk strength type III and type IV items (Table 5.2) revealed a significant main effect of similarity, $F(1, 47) = 9.36$, $p = .004$, $\eta^2 = .16$, with higher endorsement rates for type IV (similar) items ($M = .73$, $SE = 0.02$) than type III (dissimilar) items ($M = .66$, $SE = 0.02$). The similarity by group interaction was not significant, $F(1, 47) = 2.16$, $p > .05$, $\eta^2 = .04$, neither was the main effect of group, $F(1, 47) < 1$.

Table 5.2. Proportion of Items Endorsed as Grammatical (*SD*s) by Skilled and Dyslexic Readers as a Function of Stimulus Type and the Resulting Signal Detection Theory Measures (Experiments 4.1 – 4.2)

Group	Endorsement rates						SDT measures	
	Low Chunk Strength (FAs)			High Chunk Strength (Hits)			*d'* ± ME	Bias *c* ± ME
	Type I	Type II	Mean	Type III	Type IV	Mean		
Skilled readers	.29 (0.13)	.39 (0.15)	.34 (0.11)	.65 (0.14)	.75 (0.14)	.70 (0.12)	0.98 ± 0.19	-0.06 ± 0.08
Dyslexic readers	.33 (0.15)	.40 (0.13)	.36 (0.11)	.68 (0.16)	.71 (0.15)	.70 (0.14)	0.93 ± 0.26	-0.10 ± 0.12

Note. Skilled adult readers: n = 30. Dyslexic readers: n =19. Type I items are low chunk strength items which contain bigram/trigram violations as well as positional violations. Type II items are low chunk strength items which contain bigram/trigram violations only. Type III items are high chunk strength items that differ from any training item by 3 letters or more, i.e., they are relatively dissimilar to the items used in the memorization phase. Type IV items are high chunk strength items that differ from the closest training item by one letter, i.e., they are similar to a specific training item.
FA= False Alarms. ME = Margin of Error. SDT = Signal Detection Theory.

5.2.4.3 Correlational Analyses

Outlier values that deviated by 2.5 *SD*s from the sample mean were replaced by the values of the sample mean ± 2.5 *SD*s. Following replacement, data were examined for departures from normality. Skilled readers' performance on the cognitive and literacy measures was normally distributed, with the exception of response latencies in the nonword phoneme deletion task which were positively skewed (skew = 1.03, *SE* = 0.43). In the dyslexic readers' group, the distribution of digit span performance was positively skewed (skew = 1.43, *SE* = 0.52) and leptokurtic (kurtosis = 3.08, *SE* = 1.01). Both measures were log-transformed to normalize their distribution. Inter correlations among the cognitive and literacy measures are shown in Appendix K (skilled readers) and L (dyslexic readers).

Relationship between *d'* scores and memorization performance. Inter correlations among the three indexes of memorization phase performance (Table 5.3) replicated the high negative relationship between the proportion of strings reproduced correctly within a single attempt and the mean number of trials to criterion seen in the previous study, skilled readers: $r(28) = -.98$, $p < .001$; dyslexic readers: $r(17) = -.98$, $p < .001$. Scores were *z*-transformed and combined into a composite measure of memorization accuracy. The correlation between skilled readers' composite accuracy score and their mean correct memorization RTs was positive and significant, such that slower responses for correct trials were associated with higher memorization accuracy (speed-accuracy trade-off) in this group. Dyslexic readers' memorization accuracy, on the other hand, was unrelated to their reaction times. As in the previous study, discrimination ability was not related to memorization phase performance (accuracy or RTs) in either data set.

Table 5.3. Correlations Between *d*' scores, the Composite Measure of Memorization Accuracy, and Memorization RTs in Experiments 4.1 – 4.2.

Variable	1	2	3
Skilled readers (*n* = 30)			
1. *d*' scores	—	.07	-.22
2. Memorization ACC		—	.41*
3. Memorization RTs			—
Dyslexic readers (*n* = 19)			
1. *d*' scores	—	-.28	-.05
2. Memorization ACC		—	-.02
3. Memorization RTs			—

Note. ACC = Accuracy
* $p < .05$.

Relationship between *d*' scores and the cognitive/literacy measures. As in the previous study, Pearson's correlation coefficients were calculated to determine the extent of association between participants' discrimination ability and their performance on the cognitive and literacy measures (Table 5.4). Once again, WRAT reading and spelling performance was not related to skilled or dyslexic readers' discrimination ability. Individual differences in the verbal and nonverbal IQ estimates were also unrelated to skilled and dyslexic readers' *d*' scores (except, perhaps, for a small trend of a weak positive relationship between dyslexic readers' *d*' scores and Matrices performance, $r(17) = .34$, $p = .152$. In contrast to experiment 3.1, we did not observe a negative relationship between Symbol Search performance and *d*' scores in the skilled reader sample analyses; however, a trend towards the opposite (i.e., positive) relationship between *d*' scores and performance on the Symbol Search task was observed in the dyslexic reader sample analyses, $r(17) = .44$, $p = .060$. Digit span performance did not correlate with *d*' scores in either data set.

Table 5.4. Correlations between Discrimination Accuracy, *d*' scores in Experiments 4.1 – 4.2 and Performance on the Cognitive and Literacy Measures (Raw Scores).

Variable	Skilled readers (n = 30)	Dyslexic readers (n = 19)
WRIT Vocabulary	-.15	.32
WRIT Matrices	-.16	.34
WRAT Reading	-.26	.22
WRAT Spelling	-.04	.26
WAIS Digit Span	.15	-.16[b]
WAIS Symbol Search	.04	.44
RAN digits mean time[a]	-.07	.04
RAN objects mean time[a]	.11	.01
NWPD latencies	.15[b]	.01[a]

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = NonWord Phoneme Deletion.
[a]In seconds. [b]Log transformed.

Relationship between memorization performance and the cognitive/literacy measures. Individual performance differences in the cognitive and literacy measures were unrelated to discrimination ability in the drag and drop artificial grammar learning variant. A similar pattern of results was observed with the composite memorization accuracy measure (Table 5.5).

Table 5.5. Correlations between the Composite Measure of Memorization Accuracy in Experiments 4.1 – 4.2 and Performance on the Cognitive and Literacy Measures (Raw Scores).

Variable	Skilled readers (n = 30)	Dyslexic readers (n = 19)
WRIT Vocabulary	-.21	-.13
WRIT Matrices	-.02	-.28
WRAT Reading	.09	.14
WRAT Spelling	-.08	.16
WAIS Digit span	.12	.33[b]
WAIS Symbol search	-.07	.14
RAN digits mean time[a]	-.15	.04
RAN objects mean time[a]	.15	-.00
NWPD latencies	.05[b]	-.25[a]

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = NonWord Phoneme Deletion.
[a]In seconds. [b]Log transformed.

However, as shown in Table 5.6, skilled readers' mean correct memorization RTs showed a significant negative association with Symbol Search performance. That is, replicating the findings of Experiment 3.1, skilled readers with faster speed of processing skill were also faster in memorizing the letter strings. A trend for similar negative association, $r(17) = -.43$, in the dyslexic reader sample analyses failed to reach significance, $p = .068$.

Dyslexic readers' correct memorization RTs were significantly negatively correlated with verbal short-term memory skill and tended to be negatively associated with reading performance, $r(17) = -.41$, $p = .081$, and spelling performance, $r(17) = -.38$, $p = .109$. The last two correlations did not surpass significance levels, likely due to the lack of statistical power for analyses. Scatter plots for the two significant correlations with memorization speed (skilled readers: Symbol Search; dyslexic readers: Digit Span) are shown in Figure 5.1a and Figure 5.1b, respectively.

Table 5.6. Correlations between Mean Correct Memorization RTs in Experiments 4.1 – 4.2 and Performance on the Cognitive and Literacy Measures (Raw Scores).

Variable	Skilled readers (n = 30)	Dyslexic readers (n = 19)
WRIT Vocabulary	-.23	-.28
WRIT Matrices	-.05	-.20
WRAT Reading	.11	-.41
WRAT Spelling	-.14	-.38
WAIS Digit Span	-.14	**-.48***
WAIS Symbol Search	**-.42***	-.43
RAN digits mean time[a]	-.15	.25
RAN objects mean time[a]	-.03	.29
NWPD latencies	.07[b]	.06[a]

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = NonWord Phoneme Deletion.
[a]In seconds. [b]Log transformed.
* $p < .05$

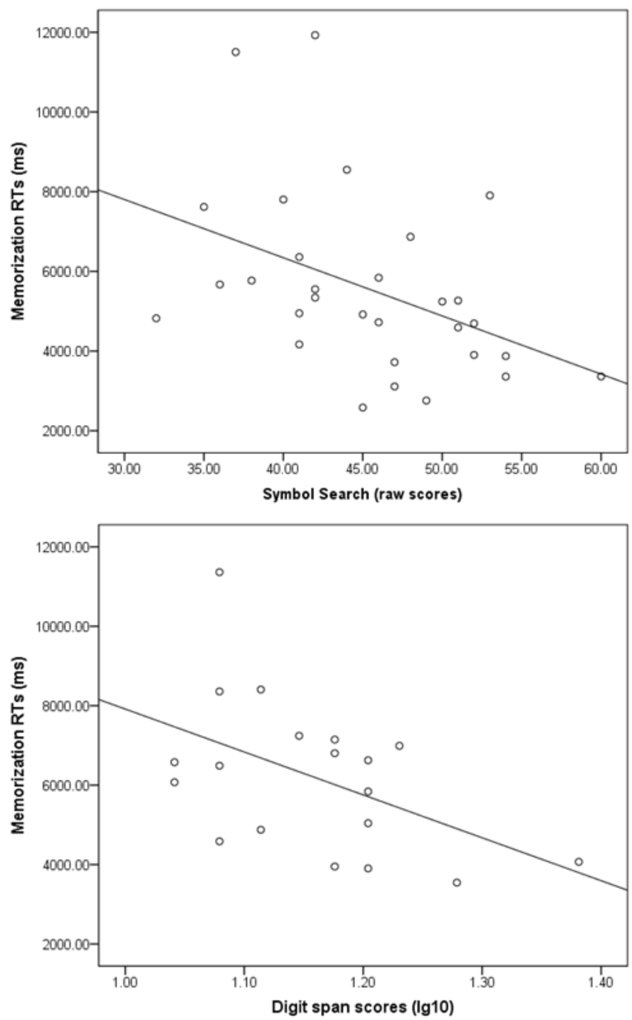


Figure 5.1. Scatter plot showing (a) a significant negative relationship between skilled readers' mean correct memorization RTs in the drag and drop letter variant and speed of processing skill (Symbol Search performance). The correlation remains significant after removal of two participants outliers with memorization RTs above 11,500 ms: $r(26) = -.37$, $p = .05$; (b) a significant negative relationship between dyslexic readers' mean correct memorization RTs in the drag and drop letter variant and performance on the digit span task (log transformed scores). The relationship remains negative and just fails to reach significance following removal of one outlier participant with mean correct memorization RTs above 11,000 ms: $r(16) = -.46$, $p = .056$.

5.2.4.4 Between-Experiment Comparisons: Experiments 3.1 – 4.2.

The present study replicated the material and procedure of experiments 3.1 – 3.2, while introducing a modification to the induction phase task demands. To compare performance between the typing and drag and drop variant, memorization and test phase data from both experiments were subjected to a series of ANOVAs similar to those reported earlier, with the additional factor Condition (typing variant, drag and drop variant). To ensure that effects were not due to (a) unexpected group differences regarding age, verbal and nonverbal IQ, (b) cognitive/literacy-related differences between the two dyslexic reader samples or (c) cognitive/literacy-related differences between the two skilled reader samples, separate one-way ANOVAs with group as a between-subject factor (skilled readers-typing variant, dyslexic readers-typing variant, skilled readers-drag and drop variant, dyslexic readers-drag and drop variant) were initially carried out.[11]

As shown in Table 5.7, the four groups of participants were matched for age, as well as performance on the Matrices and Vocabulary subtest of the WRIT. The effect of group was significant for each of the remaining variables**.** Levene's test for equality of variances was significant in the analyses of response latencies in the RAN digits and RAN objects task**,** and the nonword phoneme deletion task. Consequently, Welch’s ANOVAs were used; these analyses confirmed that the main effect of group was significant in each case: RAN digits: $F(3, 45.67) = 8.34$, $p < .001$, RAN objects: $F(3, 45.45) = 7.48$, $p = .001$, nonword phoneme deletion: $F(3, 43.68) = 19.10$, $p < .001$. Pairwise comparisons corrected for multiple tests using the Bonferonni or Games-Howell method (e.g., in the RAN digits latency analysis) showed significant differences in favor of skilled readers relative to dyslexic readers in the drag and drop variant (all measures: *p*s < .05), as well as in the typing variant (all measures: *p*s < .05 except for Digit Span[12], $p = .225$). The two groups of skilled readers were well matched in terms of performance on all measures (*p*s > .05) and so were the two groups of dyslexic participants (*p*s > .05). No

[11]For the purpose of this comparison, verbal and nonverbal IQ was estimated by means of performance on the Matrices and Vocabulary subtests of the WRIT, as full WRIT scores were only available in Experiments 3.1 – 3.2.

[12]Note that the difference between skilled and dyslexic readers’ Digit Span performance in the typing variant was significant in the original (i.e., uncorrected) independent *t*-test analyses reported in Chapter 4. This is the only inconsistency between the analyses reported here and those reported in section 4.2.4.1 (typing variant) or in section 5.2.4.1 (drag and drop variant).

Bonferonni-corrected pairwise comparison reached significance with regards to Symbol Search performance.

In sum, the above analyses revealed no unexpected cognitive and/or literacy-related performance differences between the two samples of skilled readers or the two samples of dyslexic readers. These findings verify that good sampling techniques and reliable background measures were used, and justify the validity of the between-experiment comparison reported in the following section.

Table 5.7. Results from one-way ANOVAs Comparing Cognitive and Literacy-Related Performance across the Four Experimental Groups (Experiments 3.1 – 4.2).

Variable	Levene's	*F*	*p*	η^2
Age (years)	.227	1.96	.125	.06
WRIT Vocabulary[a]	.825	0.77	.513	.02
WRIT Matrices[a]	.151	1.76	.160	.05
WRAT Reading[a]	.476	19.98	***	.39
WRAT Spelling[a]	.662	28.95	***	.48
WAIS Digit Span[b]	.401	6.40	**	.17
WAIS Symbol Search[b]	.774	3.67	*	.10
RAN digits mean time[c]	*	9.18	***	.23
RAN objects mean time[c]	**	8.66	***	.22
NWPD latencies[c]	***	21.62	***	.41

Note. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = Nonword Phoneme Deletion.
[a]Standard Scores. [b]Scaled scores. [c]In seconds.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Memorization phase.[13] We compared memorization performance across conditions (typing, drag and drop) and groups (skilled readers, dyslexic readers) by means of three separate ANOVAs on the proportion of strings reproduced within a single attempt, the number of trials to

[13]There was no effect of list or any interaction involving this factor in the between-experiment memorization phase analyses.

criterion and mean correct memorization RTs. Condition had a significant effect on the proportion of strings reproduced correctly/first attempt, $F(1, 94) = 15.54$, $p < .001$, $\eta^2 = .13$, with lower accuracy in the typing variant ($M = .80$, $SE = 0.01$) when compared to the drag and drop variant, ($M = .88$, $SE = 0.01$). The main effect of group was significant, such that dyslexic readers ($M = .81$, $SE = 0.01$) were overall less accurate than skilled readers ($M = .87$, $SE = 0.01$), $F(1, 94) = 7.98$, $p = .006$, $\eta^2 = .07$. There was no interaction between group and condition, $F(1, 94) < 1$. A similar finding emerged for the number of trials to criterion. The main effect of condition was significant, $F(1, 94) = 12.35$, $p = .001$, $\eta^2 = .11$, due to more trials per string in the typing variant ($M = 1.29$, $SE = 0.02$) when compared to the drag and drop variant ($M = 1.16$, $SE = 0.03$). There was also a significant main effect of group, $F(1, 94) = 7.79$, $p = .006$, $\eta^2 = .07$, dyslexic readers: $M = 1.28$, $SE = 0.03$; skilled readers: $M = 1.18$, $SE = 0.02$, and no interaction between group and condition, $F(1, 94) < 1$. The ANOVA on mean correct memorization RTs showed no effect of condition, $F(1, 94) = 1.38$, $p > .05$, $\eta^2 = .01$, a significant main effect of group, $F(1, 94) = 12.54$, $p < .001$, $\eta^2 = .11$, and a significant group by condition interaction, $F(1, 94) = 4.41$, $p = .038$, $\eta^2 = .04$. Decomposing the interaction pointed to significantly faster memorization RTs among skilled relative to dyslexic readers in the typing variant (chapter 4, section 4.2.4.2), but not in the drag and drop variant (see section 5.2.4.2 in this chapter).

Test phase.[14] Mean endorsement rates were subjected to a mixed ANOVA with condition (typing, drag and drop) and group (skilled readers, dyslexic readers) as between-subject variables and chunk strength (high, low) as a within-subject variable. The mean proportion of "yes" responses did not vary as a function of condition, $F(1, 94) < 1$, typing variant: $M = .52$, $SE = 0.01$; drag and drop variant: $M = .52$, $SE = 0.01$. There was a significant main effect of group, $F(1, 94) = 5.53$, $p = .021$, $\eta^2 = .05$, due to a higher proportion of "yes" responses by dyslexic readers ($M = .54$, $SE = 0.01$) when compared to skilled readers ($M = .50$, $SE = 0.01$) (see also the analyses of response bias *c*, below), and no condition by group interaction, $F(1, 94) = 2.98$, p > .05, $\eta^2 = .03$. None of the remaining interactions was significant, all *F*s < 1. The main effect of chunk strength was highly significant in the pooled data set, F(1,

[14]Given that list effects were of little theoretical value and/or not robust across analyses, this factor was not included in the between-experiment test phase analyses.

94) = 393.20, $p < .001$, $\eta^2 = .81$, with higher endorsement rates for high strength ($M = .70$, $SE = 0.01$) than low strength items ($M = .35$, $SE = 0.01$).

A two-way group by condition ANOVA on mean *d*' scores showed no effect of condition or group, *F*s < 1, or group by condition interaction, all $F(1, 94) < 1$. Skilled readers' ability to discriminate between high chunk strength/grammatical and low chunk strength/ungrammatical strings ($M = 1.01$, $SE = 0.07$) was significantly above chance across the two variants, $t(53) = 14.36$, $p < .001$, $d = 1.95$, and so was dyslexic readers' discrimination ability ($M = 0.94$, $SE = 0.07$), $t(43) = 13.24$, $p < .001$, $d = 2.00$. The measure of response bias, *c* values, did not vary as a function of condition, $F(1, 94) < 1$. Group, on the other hand, had a significant effect on mean *c* values, $F(1, 94) = 4.97$, $p = .028$, $\eta^2 = .05$, skilled readers: $M = -0.01$, $SE = 0.04$, dyslexic readers: $M = -0.13$, $SE = 0.04$. There was no group by condition interaction, $F(1, 93) = 2.40$, $p > .05$, $\eta^2 = .02$. Dyslexic participants' bias towards responding "yes" across the two variants was significant by a one sample *t* test, $t(43) = 3.49$, $p = .001$, $d = 0.53$. Skilled readers' response bias across variants was not significant, $t(53) = 0.03$, $p > .05$, $d = 0.01$. In sum, discrimination ability did not differ across task variants, and despite a significant positive bias on behalf of dyslexic participants, skilled and dyslexic learners showed comparable levels of chunk strength (i.e., frequency-based) *d*' sensitivity (i.e., after controlling for participants' response bias).

Sensitivity to letter position knowledge. Sensitivity to letter position knowledge was explored by means of a condition (typing, drag and drop) by group (skilled readers, dyslexic readers) by type of violation (chunk and positional violations [type I items], chunk violations [type II items]) ANOVA on the mean rate of endorsement of low chunk strength items. The analyses revealed no effect of condition, $F(1, 94) < 1$, and a significant main effect of group, $F(1, 94) = 5.21$, $p = .025$, $\eta^2 = .05$, due to a higher proportion of "yes" responses by dyslexic ($M = .37$, $SE = 0.02$) relative to skilled readers ($M = .32$, $SE = 0.02$). There was no interaction between group and condition, $F(1, 94) = 1.78$, p > .05, $\eta^2 = .02$, condition and type of violation, $F(1, 94) = 3.28$, $p > .05$, $\eta^2 = .03$, and condition by type of violation by group, $F(1, 94) < 1$. Type of violation did not interact with group, $F(1, 94) = 1.18$, $p > .05$, $\eta^2 = .01$. Type of violation had a significant effect on participants' mean endorsement of low chunk strength items in the pooled data set, $F(1, 94) = 8.46$, $p = .005$, $\eta^2 = .08$, such that type II items ($M = .37$, $SE = 0.02$) were endorsed more frequently than type I items ($M = .32$, $SE = 0.02$). To sum up, the analysis suggests that (a) letter position knowledge was reliably acquired—and rather unexpectedly, there was no interaction with condition, (b) the mean proportion of "yes" responses did not differ across conditions, and (c) the mean proportion of "yes" responses was significantly higher for dyslexic relative to skilled readers.

Sensitivity to specific similarity. Sensitivity to specific similarity was explored by subjecting mean endorsement rates for high chunk strength items to a 2 (condition: typing, drag and drop) by 2 (group: skilled readers, dyslexic readers) by 2 (specific similarity: dissimilar [type III items], similar [type IV items]) ANOVA. The analysis revealed no effect of condition, $F(1, 94) < 1$, no effect of group, $F(1, 94) = 1.59$, $p > .05$, $\eta^2 = .02$, or group by condition interaction, $F(1, 94) = 1.58$, $p > .05$, $\eta^2 = .02$. There was a main effect of specific similarity, $F(1, 94) = 17.84$, $p < .001$, $\eta^2 = .15$; however, the group by specific similarity interaction was also significant, $F(1, 94) = 5.38$, $p = .023$, $\eta^2 = .05$, such that skilled readers endorsed more type IV items ($M = .74$, $SE = 0.02$) than type III items ($M = .63$, $SE = 0.02$), $t(53) = 5.87$, $p < .001$, $d = 0.80$, whereas, for dyslexic readers, there was no difference between the proportion of type III items ($M = .70$, $SE = 0.02$) and type IV items ($M = .73$, $SE = 0.02$) endorsed, $t(43) = 1.10$, $p > .05$, $d = 0.17$. Condition did not interact with specific similarity, and there was no interaction between condition, group and specific similarity, both *Fs* < 1. In sum, the aggregate analysis of skilled versus dyslexic readers' sensitivity to specific similarity did not corroborate the findings of the analyses reported in sections 4.2.4.2 and 5.2.4.2. When endorsement rates for high chunk strength items were collapsed across conditions, it became obvious that only skilled readers' responses were affected by the similarity between test phase and specific training items.

5.2.5 Discussion

As in Experiments 3.1 – 3.2, a memorization to criterion task was used as a means of exposing participants to grammar-conforming letter strings and ensuring their sustained attention during the training phase of the experiment. However, participants were not instructed to type the strings on the keyboard from memory, as in the previous study; instead, 10 individual letters (7 target letters and 3 irrelevant distractors) were presented during each trial and, following a retention period of 3 seconds, participants were invited to drag and drop them in order to match the memorized letter strings. It was reasoned that while the typing task emphasizes memory for individual letter identity and letter order information to the same extent, performance during the drag and drop task would rely to a lesser extent[15] on memory for individual letters and place more emphasis on their order, i.e., V is the third letter within the string (e.g., Tremblay et al., 2000). Furthermore, unlike typing from memory which requires recall—as well as good control of the keyboard, another task difference worth at least some consideration—the drag and drop task benefited from a lesser memory load and no need to possess keyboard skills at a high level.

This modified drag and drop task was designed to assess the following questions: First, is implicit skill learning (i.e., discrimination ability at test) unimpaired in dyslexics relative to skilled adult readers (i.e., consistent with our previous study)? And second, will this modification of the training phase processing characteristics enhance skilled and dyslexic readers' sensitivity to constraints on the allowable position of letters? In a nutshell, memorization accuracy varied significantly between groups, but discrimination ability on the basis of chunk strength information was at similar levels for skilled and dyslexic readers. Letter position sensitivity was, as hypothesized, reliable in the drag and drop variant and did not differ as a function of group. However, some analyses-dependent differences were observed with regards to skilled and dyslexic readers' sensitivity to specific similarity. These key findings and other results from our study are further discussed in the sections below.

As in the previous chapter, we began by assessing group differences in performance on the memorization drag and drop phase of the experiment. Independent *t* tests revealed a significant advantage for skilled relative to dyslexic readers in terms of the proportion of strings reproduced correctly within a single attempt and the mean number of trials to criterion, but no difference in terms of mean correct memorization RTs. Given that dyslexic participants in this study were less accurate than skilled readers (and that they were slower but not less accurate in the preceding experiment), we investigated whether their relatively unimpaired reaction time performance could be explained by a speed-accuracy trade-off. If dyslexic readers were comparably fast to skilled readers by adopting a strategy along these lines, (i.e., by trading accuracy for speed), a significant negative relationship should be observed between their accuracy and speed of responding. Opposite to this expectation, the correlation between dyslexic

[15]It is acknowledged that the task poses some demands on item identity retrieval given that a few distractors were presented together with the letters that are necessary for stimulus reproduction.

readers' memorization accuracy and RTs was close to zero (r = -.02), indicating no systematic relationship between speed and accuracy of responding. Thus, dyslexic participants' similar RT performance to that of skilled readers was not at the expense of accuracy. On the other hand, a classic speed-accuracy trade-off appeared to be operating among skilled readers, such that the memorization accuracy decreased as a function of decreased (i.e., faster) memorization RTs, r = .41, p < .05.[16]

In spite of the above group differences with regards to memorization performance, we observed no reliable differences between skilled and dyslexic readers' test phase classification performance in Experiments 4.1 – 4.2. Skilled and dyslexic readers' endorsement rates were similarly affected by chunk strength, positional information and training/test item similarity. The analysis of participants' *d*' scores—estimating their ability to discriminate between high chunk strength/grammatical and low chunk strength/ungrammatical test phase item—confirmed that responding was better than chance in both groups and not reliably better for skilled relative to dyslexic readers, replicating the findings of Experiments 3.1 - 3.2. Furthermore, no difference was observed with regards to participants' response bias, which, unlike the previous study, was not statistically different from zero both for skilled and dyslexic learners. Finally, consistent with Experiments 3.1 – 3.2, similar items were significantly more frequently endorsed than dissimilar items by skilled as well as dyslexic readers (however, this finding was not confirmed in the between-experiment comparison and is further discussed in a later section).

Importantly, unlike Experiments 3.1 – 3.2, low chunk strength items embedding positional and bigram/trigram violations (type I items) were endorsed less frequently than illegal items embedding only bigram/trigram violations (type II items) by both groups of participants. Given that the two sets differed only in terms of their adherence to positional information—there was an equal amount of bigram/trigram violation between type I and type II items—skilled and dyslexic readers' higher ease of rejecting type I items was taken to indicate sensitivity to constraints on where individual letters occurred. This finding is consistent with Wright and

[16] To venture another hypothesis regarding the lack of significant group differences in memorization speed, it is suggested that the string reproduction (i.e., drag and drop) component of the task may have been less of an impediment to dyslexic participants' performance and may have interacted with their ability to memorize the strings. A group by condition ANOVA on participants' mean string reproduction reaction times may have been informative with regards to the claim that impaired typing skill and/or fluent command of the keyboard contributed to the group differences observed in the typing artificial grammar learning variant.

Whittlesea's (1998) proposal that participants' learning is not independent of the training-phase task demands and the processes carried out for them to be met. For example, it has been shown even simple perceptual manipulations of the stimulus presentation conditions (horizontal vs. vertical stimulus arrangement) can affect how—otherwise identical—items are processed and what sources of information are encoded (Wright & Whittlesea, 1998). Our finding of reliable positional sensitivity in the drag and drop, but not in the typing artificial grammar learning variant, corroborates Whittlesea and colleagues' claim that artificial grammar learners are not as passive learners as it was originally suggested (e.g., Reber & Allen, 1978).

Between-experiment comparisons. Further to examining skill learning in Experiments 4.1 – 4.2, a between-experiment analysis (drag and drop variant, typing variant) was performed to quantify performance differences between the two tasks. There were four key findings regarding the *effect of condition* (i.e., our experimental manipulation). First, our new experimental procedure had a significant facilitatory effect on memorization accuracy, such that participants in the drag and drop condition—collapsed across the factor of group—were more accurate in reproducing the strings correctly within a single attempt and took fewer trials to reach criterion. Second, despite the significant group by condition interaction on memorization RTs, no difference was observed in terms of the amount of time participants spent for string memorization in the two variants. That is, the drag and drop condition did not facilitate memorization RT performance. Third, in sharp contrast to the significant facilitatory effect of condition on memorization accuracy, we observed no difference between learners' overall proportion of items endorsed as grammatical ("yes" responses), overall discrimination ability, or their overall response bias in the two variants. The main effect of positional sensitivity was significant across tasks and did not interact with condition. Although this is an unexpected finding (i.e., one would have anticipated an interaction pointing to differences in positional sensitivity between Experiments 3.1 - 3.2 and Experiments 4.1 – 4.2), it is taken to suggest that in a larger sample of participants, sensitivity to positional violations is reliable independent of training phase manipulations. Finally, participants' rate of endorsement of similar versus dissimilar items did not interact with condition. That is, sensitivity to specific similarity was not reliably different between the drag and drop and typing experiment.

Collapsing data across the two variants revealed some further interesting findings regarding the *effect of group* (i.e., skilled vs. dyslexic readers). Overall, dyslexic readers were significantly less accurate in reproducing the strings correctly within a single attempt and needed significantly more trials to reproduce all 96 strings correctly, relative to skilled readers. The effect of group on mean correct memorization RTs, on the other hand, was moderated by a significant interaction with condition, such as that dyslexic readers were significantly slower, relative to controls, in the typing variant (chapter 4; section 4.2.4.2), but not in the drag and drop variant (section 5.2.4.2 in this chapter). Importantly, the aggregated analysis corroborated our previous *d*' analysis by showing that skilled and dyslexic readers' overall level of discrimination ability on the basis of chunk strength was at strikingly similar above chance levels; however, across conditions, dyslexic readers' performance was characterized by a general tendency towards responding "yes", as evident from (a) the omnibus analyses of response bias, (b) dyslexic readers' overall higher rate of endorsements (higher proportion of "yes" responses across variants) and their higher proportion of FAs (incorrect "yes" responses). It is reminded that *d*' scores correct for differences in participants' response bias, therefore, it is still warranted that dyslexic readers performance was similar to that of skilled readers. Nevertheless, the bias *c* analyses results suggest that the same levels of discrimination ability were obtained in different ways by skilled and dyslexic readers. The aggregated analyses also confirmed that there were no statistical group differences with regards to participants' sensitivity to positional information, but a difference emerged in terms of their sensitivity to the specific similarity manipulation. Only skilled readers showed an advantage in discrimination ability for type IV (similar) relative to type III (dissimilar) items. This finding is inconsistent with the lack of group difference between skilled and dyslexic readers' sensitivity to specific similarity in the previous chapter (typing variant; experiments 3.1 – 3.2), as well as in this chapter (drag and drop variant; experiment 4.1 – 4.2) and suggests that lack of statistical power had concealed an impairment in dyslexic individuals' ability to discriminate between similar and dissimilar test phase items. The significant group difference in sensitivity to specific similarity, taken together with the group differences in memorization learning indexes suggest that, in some aspects, dyslexic participants' learning performance may not be spared relative to that of controls. Prior to concluding whether these demonstrations are consistent with the claim of a learning impairment in dyslexia (but certainly not as pervasive as hypothesized; e.g., the group differences in overall *d*' scores were far from significant), it is important to consider whether similar deficits are found when patterned skill learning is assessed in a nonlinguistic drag and drop variant. This was the main purpose of Experiments 4.3 – 4.4.

5.2.5.1 Correlational Analyses

As in our previous study, participants' ability to discriminate between grammar-conforming/nonconforming letter strings at test was unrelated to their literacy skills (reading skill, spelling skill) and literacy-related skills (rapid automatized naming and phoneme awareness skill). Verbal short-term memory was also unrelated to test phase performance, replicating the findings of previous artificial grammar learning studies (Gebauer & Mackintosh, 2007; Robinson, 2005, among others). Once again, test phase performance was not systematically related to either memorization accuracy or memorization speed. Memorization accuracy was closer to ceiling—at least among skilled readers; mean accuracy ≈ .91—and proved to be minimally susceptible to individual variations in skilled/dyslexic readers' performance on the cognitive and literacy measures. However, significant associations were observed between memorization RTs and performance on two of the background measures: First, RTs varied inversely as a function of Symbol Search performance among skilled readers, and a similar trend among dyslexic readers just failed to reach significance, $p = .068$; this finding is consistent with the negative correlation between graphomotor speed of processing skill and mean correct memorization RTs found in experiments 3.1- 3.2.[17] Second, RTs varied inversely as a function of verbal short-term memory skill among dyslexic readers, such that higher Digit Span scores were associated with faster memorization performance. Both of these findings are logical and well-reasoned and suggest that individuals with better memory and faster speed of processing were faster in memorizing the training phase letter strings.

A systematic investigation of the relationship between psychometric intelligence and discrimination ability at test was not an objective in this series of experiments (in the interest of reducing the length of our experimental session, full-scale IQ estimates were not obtained). To

[17]It should be noted that, in our previous study, this relationship was observed for performance on a measure of graphomotor speed of processing skill (Symbol Digit Modalities Test; Smith, 1991), and not for performance on a purer measure of speed of processing skill (WAIS Symbol Search subtest; Wechsler, 1998). Unfortunately, it was not possible to verify whether the association with memorization RTs would also hold for performance on the Symbol Digit Modalities Test, as the latter task was not administered to participants of this study.

the extent that performance on the Vocabulary and Matrices WRIT subtests are reliable estimates of verbal and nonverbal IQ, respectively, we suggest that these two are unrelated. On the other hand, we were interested in examining the association of Symbol Search performance—a measure of speed of processing—with discrimination ability. In our previous artificial grammar learning study, these varied inversely among skilled readers, whereas in Kaufman et al. (2010) implicit sequence learning study with adults, a positive association between learning skill and a latent speed of processing variable was reported. In this study, the relationship was proven unstable. There was no longer a negative correlation between Symbol Search performance and discrimination ability in the skilled reader sample analysis, and a trend for a positive correlation between dyslexic readers' Symbol Search scores and *d*' scores was observed.

To recap, the correlational analyses replicated, by and large, the key findings of experiments 3.1 – 3.2. First, individual differences in *d*' scores were less sensitive to variability among individuals in terms of performance on most cognitive and literacy measures, whereas some correlations were observed between participants' memorization speed and the remaining variables. These were not always identical to those observed in the context of our previous study; at least some of the findings of our correlational analysis (e.g., speed of processing and its relation to implicit learning) needed to be interpreted with caution. Nevertheless, the general picture is consistent with Reber et al.'s (1991) claim that memorization phase performance should show a higher degree of concordance with measures of explicit functioning (e.g., memory, IQ) than well-formedness test phase performance. Second, *d*' scores were unrelated to individual differences in literacy/literacy-related skill. While puzzling and difficult to account for, the lack of significant correlations between implicit learning and literacy performance appears to be a consistent and reproducible finding in this thesis across different tasks and populations varying in age and/or literacy proficiency (typically developing children, skilled and dyslexic adult learners). We return to suggest a possible explanation of this finding in the General Discussion.

5.3 Experiments 4.3 – 4.4

5.3.1 Rationale

Experiments 4.3 – 4.4 were created in direct analogy to Experiments 4.1 – 4.2. The internal structure of the memorization/test phase material and all aspects of the procedure were the same across studies, but stimulus format differed: letter strings were replaced by sequences of nonlinguistic symbols. Our between-group design (skilled readers, dyslexic readers) was maintained and was of paramount importance for establishing the nature of some group differences that were observed in our previous studies. For example, should the observed differences between skilled and dyslexic readers' training phase performance be taken to indicate a causal link between learning skill and literacy impairments? Was dyslexic participants' impaired sensitivity to the similarity between training and test phase letter strings (found in our between-experiment comparison: section 5.2.4.3) caused by the linguistic nature of the stimuli? We expected that if implicit skill learning is unimpaired in dyslexia, and any instances of dyslexic individuals' worse performance were related to the need to process letter stimuli, the deficits documented in the context of our previous studies would not emerge for performance in the shapes task variant. There are some previous studies employing, in similar reasoning, nonlinguistic implicit learning tasks to assess learning in language impaired populations (e.g., dyslexic children and adults: Pavlidou et al., 2010; Pothos & Kirk, 2004; patients with agrammatic aphasia: Christiansen, Kelly, Shillcock, & Greenfield, 2010; children with autism spectrum disorder: Klinger, Klinger, & Pohlig, 2007).

Although the vast of majority of artificial grammar learning research has examined humans' ability to classify letter strings (see however Seger, 1994, for a brief overview of some early studies examining learning using other types of stimuli), there is little doubt that similar learning is possible with various types of stimuli and even in different sense modalities. As reviewed in the Literature review, participants can even transfer their knowledge from one to another type of stimuli (e.g., Altmann et al., 1995). However, we wondered whether learning performance and/or sensitivity to different sources of knowledge would be modulated by differences in the format used to instantiate structure. Pothos, Chater and Ziori's (2006) work pertains to this topic. In a study with adults, they compared their ability to categorize stimuli generated from the same finite-state grammar on the basis of rule knowledge and global associative chunk strength differences under three experimental conditions: One group was exposed standard letter strings, a second group was exposed to embedded geometrical shape sequences (see section 4.1.1 for an example of Pothos' embedded shapes stimuli), and a third group was exposed to "city sequences" (i.e., artificial "airline routes"; e.g., Berlin-Athens-Madrid-London). Pothos et al.'s (2006) study provided little evidence of differences in participants' ability to encode different stimulus aspects between the three conditions, and it was therefore suggested that artificial grammar learning is independent of stimulus appearance. We assessed whether this finding also holds true when verbal stimulus coding is minimized (our shapes did not have commonly accepted names), and whether it also generalized to participants' sensitivity to letter positions and specific similarity.

5.3.2 Method

5.3.2.1 Participants

Typical Readers. Forty skilled readers (14 male; mean age = 20.88 years, range = 18.17 years – 34.58 years) and 31 students diagnosed with dyslexia (12 male; mean age = 21.05 years, range = 18.25 years – 36.25 years) participated in the study. All participants were naïve to the purpose of the study. Skilled readers were recruited through the Bangor University participant panel. Data from 12 of them were not included in the analyses due to one of the following reasons: (a) failure to meet the study language criteria (n = 2), (b) unexpectedly low performance on reading/spelling tasks or verbal/nonverbal IQ tasks (n = 9), (c) failure to complete the memorization phase within 75 minutes (n = 1). Dyslexic readers were recruited through the Miles Dyslexia Center and were compensated monetarily for the participation. Fourteen participants had received a formal diagnosis of dyslexia via the Miles Dyslexia Center's assessment service; the remaining 17 participants were diagnosed by a qualified educational psychologist during secondary school or later. Data from 5 dyslexic participants were excluded due to one of the following reasons: (a) failure to meet the study language criteria (n = 1), (b) known co-occurrence of another developmental or neurological disorder (n = 3), (c) unexpectedly low performance on the Matrices subtest of the WRIT (n = 1).[18] Furthermore, five dyslexic participants in this study failed to complete the memorization phase of the artificial grammar learning task within a reasonable time frame (~75 minutes) and were unable to proceed to the test phase. These participants were not considered in any further analyses. The final sample consisted of 28 skilled readers (10 male; mean age = 20.70 years, range = 18.25 years – 31.25 years) and 21 dyslexic readers (9 male; mean age = 20.54 years, range = 18.42 years – 27.92 years), all of which were monolingual native English speakers and had normal or corrected-to-normal visual acuity.

5.3.2.2 Background Measures

The psychometric tests and literacy-related measures used in experiments 4.1 – 4.2 were also used in experiment 4.3 – 4.4.

5.3.2.3 Material

Stimuli were generated by the same artificial grammar used in Experiments 3.1 – 4.2. Thus, the underlying structure of both training and test phase items was the same as in the previous experiments. Letter strings were converted to sequences of shapes by mapping each of the 20 letters used by Kinder and Lotz (2009) to an abstract easily distinguishable shape developed by Taylor, Plunkett, and Nation (2011). Abstract shapes were used instead of geometrical shapes (e.g., Pothos et al., 2006) to prevent, as much as possible, participants from adopting a verbal encoding strategy. The two sets of shape sequences (counterbalanced across participants) that were used in the memorization phase of experiments 4.3 – 4.4 are shown in Appendix G. The two counterbalanced lists of test phase sequences are shown in Appendix H.

5.3.2.4 Procedure

The artificial grammar learning task procedure was the same as in experiment 4.1 – 4.2, except that participants were presented with shape sequences. Testing conditions were the same as in the previous studies.

[18] Data from one dyslexic participant with a standard score of 83 on WRIT Matrices were retained in order to avoid further loss of statistical power.

5.3.3 Results

5.3.3.1 Background Measures

Skilled and dyslexic readers' performance on the cognitive and literacy measure is summarized in Table 5.8. The two groups did not differ in terms of age, verbal and nonverbal IQ. As in the previous experiments, dyslexic participants performed within the average range of the WRAT Reading and Spelling subtests. Skilled readers' performance was also in the average range; however, independent *t* tests confirmed that the two groups differed reliably in terms of literacy skill. Dyslexic readers obtained significantly lower Digit Span scores than skilled readers; there was no group difference in terms of speed of processing skill (Symbol Search performance). Phonological awareness and rapid automatized naming of digits was significantly impaired in dyslexic relative to skilled readers. The two groups did not differ significantly in terms of RAN objects mean latencies, $p = .066$.

Table 5.8. Mean scores (*SD*s) and Ranges on the Cognitive and Literacy Related Measures as a Function of Group and Results from Independent *t* tests (Experiments 4.3 – 4.4).

Variable	Skilled readers (n = 28)		Dyslexic readers (n = 20)		t	sig.	Cohen's d
	Mean (*SD*s)	Range	Mean (*SD*s)	Range			
Handedness (r; l; a)	25; 3; 0		17; 3; 1				
Age (years)	20.70 (2.58)	18.25 – 31.25	20.54 (2.23)	18.42 – 27.92	0.23	.818	0.07
WRIT Vocabulary[a]	106.68 (9.85)	90.00 – 128.00	102.15 (8.75)	85.00 – 122.00	1.64	.107	0.49
WRIT Matrices[a]	103.75 (10.18)	86.00 – 127.00	103.65 (13.58)	83.00 – 130.00	0.03	.977	0.01
WRAT Reading[a]	104.11 (5.06)	95.00 – 116.00	97.90 (7.17)	86.00 – 114.00	3.52	**	1.00
WRAT Spelling[a]	109.64 (8.87)	95.00 – 129.00	97.65 (7.23)	85.00 – 114.00	4.98	***	1.48
WAIS Digit Span[b]	10.22 (2.87)	6.00 – 17.00	8.35 (1.66)	6.00 – 11.00	2.60[e]	*	0.80
WAIS Symbol Search[b]	13.68 (2.50)	10.00 – 19.00	12.95 (3.28)	7.00 – 19.00	0.87	.387	0.25
RAN digits mean time[c]	14.02 (2.69)	7.80 – 19.95	17.70 (4.43)	11.80 – 27.95	3.31[f]	**	1.00
RAN objects mean time[c]	22.27 (3.13)	17.27 – 31.50	24.91 (5.59)	17.45 – 34.58	1.91[f]	.066	0.58
NWPD accuracy[d]	11.07 (0.91)	8.50 – 12.00	10.33 (1.20)	8.00 – 12.00	2.46	*	0.70
NWPD latencies[c]	38.80 (9.85)	17.52 – 72.77	54.36 (15.67)	34.83 – 94.13	3.92[f]	***	1.19

Note. Dyslexic readers: n =20 due to missing background data from one participant. r = right-handed. l = left-handed. a = ambidextrous. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = Nonword Phoneme Deletion.
[a]Standard Scores. [b]Scaled scores. [c]In seconds. [d]Out of 12. [e]Skilled readers: n = 27 due to missing data from one participant.
[f]Correction for unequal variances applied.
* $p < .05$. ** $p < .01$. *** $p < .001$.

5.3.3.2 Artificial Grammar Learning Task

Memorization phase. In contrast to experiments 4.1 – 4.2, no significant difference emerged between skilled readers' (83%, $SD = 0.11$) and dyslexic readers' (83%, $SD = 0.12$) ability to correctly reproduce the shape sequences within a single attempt, $F(1, 45) < 1$. However, there was a small significant main effect of list, $F(1, 45) = 4.86$, $p = .033$, $\eta^2 = .10$, indicating higher proportion accuracy in list 2 ($M = .87$, $SE = 0.02$) when compared to list 1 ($M = .79$, $SE = 0.02$). List did not interact with group, $F(1, 45) = 1.00$, $p > .05$, $\eta^2 = .02$. The mean number of trials to criterion did not differ reliably for skilled readers ($M = 1.21$, $SE = 0.03$) and dyslexic readers ($M = 1.22$, $SE = 0.03$), $F(1, 45) < 1$, and was marginally affected by list (list 1: $M = 1.26$, $SE = 0.03$; list 2: $M = 1.17$, $SE = 0.03$), $F(1, 45) = 4.07$, $p = .050$, $\eta^2 = .08$. The ANOVA on the mean number of trials to criterion revealed no group by list interaction, $F(1, 45) < 1$. Memorization reaction time data, analyzed by means of an independent *t* test, were not statistically different for skilled readers ($M = 10502.69$, $SE = 613.85$) and dyslexic readers ($M = 10573.41$, $SE = 647.33$), $t(47) = 0.08$, $p > .05$, $d = 0.02$. List did not have a significant effect on mean correct memorization RTs, $F(1, 45) < 1$.

In sum, dyslexic readers' memorization performance matched that of skilled readers when letters strings were replaced by sequences of unpronounceable shapes. The main effect of list on mean proportion accuracy (list 2 > list 1) parallels Experiments' 4.1 – 4.2 finding of longer memorization RTs for list 1 relative to list 2 items, and suggests that the latter set of training items were slightly more difficult to memorize. Given that memorization performance was affected similarly across the two formats, the effect is unlikely to reflect list differences regarding the presence of highly salient letter patterns (e.g., TV). Another aspect of stimulus structure may have contributed to this pattern, although, it is not clear what this may have been (e.g., there were no immediate repetitions of elements within items either in list 1 or in list 2; there were no item differences between lists in terms of a "same first-last letter/shape" criterion, etc.).

Test phase. All test phase responses were associated with RTs above 300 ms. Thus, no data points were excluded from the analyses. Mean endorsement rates for high and low chunk strength sequences are shown separately for skilled and dyslexic readers in the left of panel of

Table 5.9. As in all previous experiments, the mean rate of endorsements was significantly affected by a strong main effect of chunk strength, $F(1, 47) = 122.93$, $p < .001$, $\eta^2 = .72$, due to a higher proportion of "yes" responses for high strength ($M = .68$, $SE = 0.02$) than low strength items ($M = .41$, $SE = 0.02$). There was no group by chunk strength interaction, $F(1, 47) < 1$, or main effect of group, $F(1, 47) = 1.60$, $p > .05$, $\eta^2 = .03$.

d' values and criterion *c* values are shown separately for each group of participants in the right panel of Table 5.9. Dyslexic readers' discrimination ability was actually numerically higher, but not significantly different from that of skilled readers, $t(47) = 0.63$, $p > .05$, $d = 0.18$. One sample *t* tests confirmed that mean *d*' values reliably exceeded chance levels both for skilled readers, $t(27) = 7.45$, $p < .001$, $d = 1.41$, and dyslexic readers, $t(20) = 7.08$, $p < .001$, $d = 1.54$. Skilled and dyslexic readers' mean scores on the measure of response bias, *c,* were in the same direction and did not differ reliably, $t(47) = 1.17$, $p > .05$, $d = 0.34$. However, only skilled readers' bias towards accepting items was reliable, $t(27) = 3.59$, $p = .001$, $d = 0.68$; dyslexic readers: $t(20) = 1.47$, $p > .05$, $d = 0.32$. Note that this is the opposite finding from that obtained in experiments 3.1 – 3.2 (typing variant; only dyslexic readers showed a significant bias towards "yes" responses), and it is also different from the finding of experiments 4.1 – 4.2 (neither group's bias was reliable).

Sensitivity to letter position knowledge. A list by type of violation by group ANOVA showed a main effect of type of violation, $F(1, 45) = 33.60$, $p < .001$, $\eta^2 = .39$, due to a higher proportion of endorsements for type II ($M = .50$, $SE = 0.02$) than type I items ($M = .32$, $SE = 0.02$). The interaction between type of violation and group did not reach significance, $F(1, 45) = 3.30$, $p = .076$, $\eta^2 = .04$. The main effect of group was significant, $F(1, 45) = 4.49$, $p = .040$, $\eta^2 = .07$, and was qualified by an interaction with the factor of list, $F(1, 45) = 4.65$, $p = .036$, $\eta^2 = .07$. In list 2, skilled readers' mean rate of endorsements ($M = .51$, $SE = 0.03$) was significantly higher relative to dyslexic readers' mean rate of endorsements ($M = .39$, $SE = 0.03$), $t(21) = 2.88$, $p = .009$, $d = 1.20$. On the other hand, there was no difference between skilled and dyslexic readers' endorsement rates in list 1, $t(24) = 0.03$, $p > .05$, $d = 0.01$. A significant main effect of list was also observed, $F(1, 45) = 8.09$, $p = .007$, $\eta^2 = .13$ and was qualified by the same interaction, showing that list differences occurred with regards to skilled readers' endorsements, $t(26) = 4.38$, $p < .001$, $d = 1.65$, but not with regards to dyslexic readers' endorsements, $t(19) =$

0.39, $p > .05$, $d = 0.17$. The type of violation by list interaction was not significant, $F(1, 45) < 1$, and neither was the three-way interaction between list, type of violation and group, $F(1, 45) = 3.82$, $p > .05$, $\eta^2 = .04$. On the whole, the aggregated analyses of participants' sensitivity to letter position knowledge confirmed that items that violated both letter positions and bigram/trigram information were endorsed more frequently than items that violated bigrams/trigrams information alone. That is, participants' sensitivity to positional constraints was reliable. There was also evidence of list differences on the main rate of endorsements among skilled readers, as well as evidence of their liberal response bias, however, this only true for performance in list 2.

Sensitivity to specific similarity. A similarity by group ANOVA carried out on the mean rate of endorsement of high chunk strength items revealed a significant main effect of similarity, $F(1, 47) = 11.11$, $p = .002$, $\eta^2 = .19$, with a higher proportion of "yes" responses for sequences similar to those presented during memorization ($M = .72$, $SE = 0.02$) than sequences which differed by 3 shapes or more from specific training sequences ($M = .64$, $SE = 0.02$). There was no effect of group, $F(1, 47) < 1$ or group by similarity interaction, $F(1, 47) < 1$.

Table 5.9. Proportion of Items Endorsed as Grammatical (*SD*s) by Skilled and Dyslexic Readers as a Function of Stimulus Type and the Resulting Signal Detection Theory Measures (Experiments 4.3 – 4.4).

Group	Endorsement rates						SDT measures	
	Low Chunk Strength (FAs)			High Chunk Strength (Hits)				
	Type I	Type II	Mean	Type III	Type IV	Mean	*d*' ± ME	Bias *c* ± ME
Skilled readers	.32 (0.13)	.54 (0.16)	.43 (0.11)	.64 (0.15)	.73 (0.13)	.68 (0.12)	0.70 ± 0.19	-0.17 ± 0.10
Dyslexic readers	.32 (0.17)	.44 (0.15)	.38 (0.11)	.63 (0.19)	.71 (0.15)	.67 (0.14)	0.80 ± 0.23	-0.08 ± 0.12

Note. Skilled adult readers: *n* = 28. Dyslexic readers: *n* = 21. Type I items are low chunk strength items which contain bigram/trigram violations as well as positional violations. Type II items are low chunk strength items which contain bigram/trigram violations only. Type III items are high chunk strength items that differ from any training item by 3 shapes or more, i.e., they are relatively dissimilar to the items used in the memorization phase. Type IV items are high chunk strength items that differ from the closest training item by one shape, i.e., they are similar to a specific training item.
FA= False Alarms. ME = Margin of Error. SDT = Signal Detection Theory.

5.3.3.3 Correlational Analyses

The examination of the distributional properties of performance in each of the cognitive and literacy measures after outlier replacement revealed a small positive skew in skilled readers' Digit Span performance (skew = 1.03, *SE* = 0.45) and dyslexic readers' response latencies in RAN digits (skew = 1.14, *SE* = 0. 51). Correlations involving these variables were carried out following log transformation. Inter correlations among the cognitive and literacy measures are shown in Appendix M (skilled readers) and N (dyslexic readers).

Relationship between *d'* scores and memorization performance. As in the previous experiments, the proportion of stimuli reproduced correctly in a single attempt was extremely highly negatively correlated with the number of trials to criterion in the skilled reader analyses, $r(27) = -.99$, $p < .001$, as well as in the dyslexic reader analyses, $r(19) = -.98$, $p < .001$. Inter correlations between the composite accuracy score and mean correct memorization RTs (Table 5.10) revealed a significant positive relationship between speed and accuracy not only for skilled readers (i.e., as seen in Experiment 4.1), but also for dyslexic readers. Thus, when presented with nonlinguistic shape sequences, both groups of participants traded speed with accuracy, such that slower responses for correct trials were associated with higher memorization accuracy. Memorization performance (accuracy; RTs) was not related to test phase performance in either set of the analyses.

Table 5.10. Correlations Between *d'* scores, the Composite Measure of Memorization Accuracy, and Memorization RTs in Experiments 4.3 – 4.4.

Variable	1	2	3
Skilled readers (*n* = 28)			
1. *d'* scores	—	.19	-.06
2. Memorization ACC		—	**.43***
3. Memorization RTs			—
Dyslexic readers (*n* = 21)			
1. *d'* scores	—	.20	.11
2. Memorization ACC		—	**.51***
3. Memorization RTs			—

Note. ACC = Accuracy
* $p < .05$.

Relationship between *d'* scores and the cognitive/literacy measures. Table 5.11 shows the association of skilled and dyslexic readers' *d'* scores with their cognitive and literacy performance. In sharp contrast to the previous experiments, the analyses revealed a significant positive relationship between skilled readers' performance on the WRAT Reading and Spelling tasks and their well-formedness task performance (scatter plots are shown Figure 5.2). As seen in Appendix M, reading and spelling scores correlated significantly with verbal IQ (i.e., WRIT Vocabulary) performance. When the variance associated with performance on this measure was partialled out, both correlations remained, $r_1 = .39$, $r_2 = .29$, but only the association with reading performance was statistically significant ($p = .042$). *d'* scores were also strongly negatively correlated with mean latencies on the nonword phoneme deletion task in the skilled reader group analyses (the scatter plot is also shown in Figure 5.2).

A similar, but somewhat weaker, trend for a positive relationship between reading and *d'* scores was observed among dyslexic readers, $r(18) = .32$, $p = .173$. Dyslexic readers' spelling performance and nonword phoneme deletion latencies were, on the other hand, unrelated to discrimination ability. Higher nonverbal/verbal IQ, $r_1(18) = .34$, $r_2(18) = .33$, and verbal short-term memory skill, $r_1(18) = .35$ tended to be associated with higher discrimination ability among dyslexic readers; however, none of the correlations reached significance: nonverbal IQ: $p = .143$; verbal IQ: $p = .151$; digit span: $p = .135$.

Table 5.11. Correlations between Discrimination Accuracy, *d*' scores in Experiments 4.3 – 4.4 and Performance on the Cognitive and Literacy Measures (Raw Scores).

Variable	Skilled readers (n = 28)	Dyslexic readers (n = 20)
WRIT Vocabulary	.29	.33
WRIT Matrices	.11	.34
WRAT Reading	**.46***	.32
WRAT Spelling	**.38***	.15
WAIS Digit Span[a]	.15[c]	.35
WAIS Symbol Search	.04	.26
RAN digits mean time	-.30	-.03[c]
RAN objects mean time[b]	.03	.05
NWPD latencies[b]	**-.56****	-.30

Note. Dyslexic readers: n = 20 due to missing background data from one participant. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = NonWord Phoneme Deletion.
[a]skilled readers: n = 27 due to missing data from one participant. [b]In seconds.
[c]Log transformed.
* $p < .05$. ** $p < .01$.

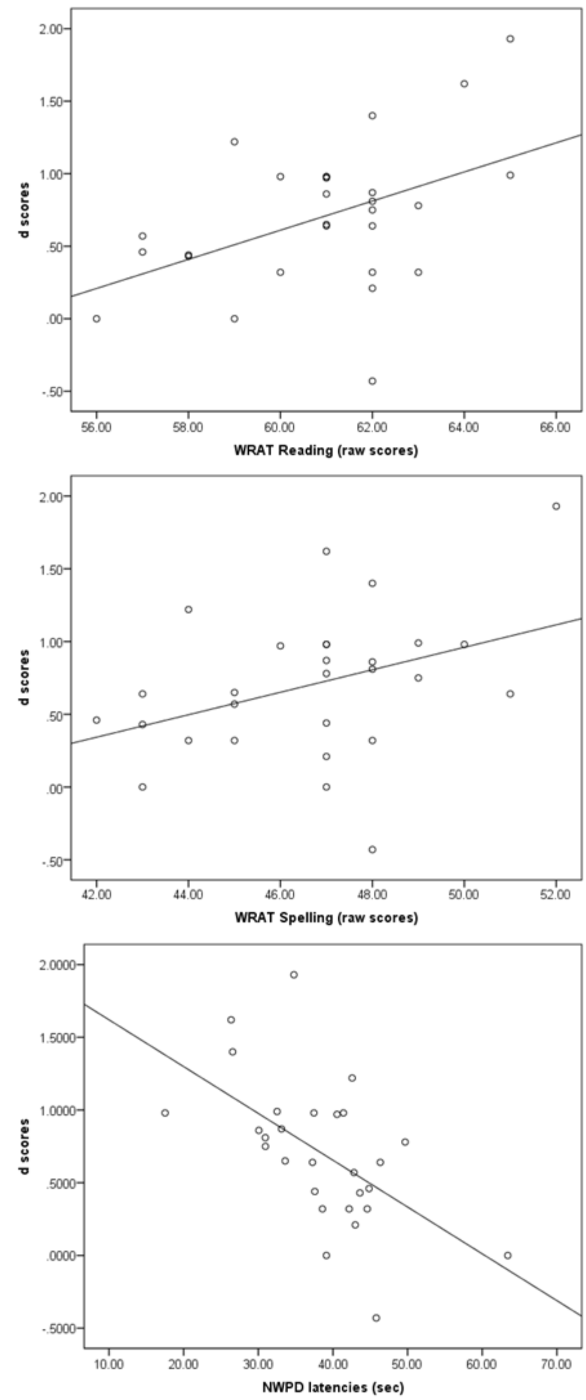


Figure 5.2. Scatter plot showing a significant positive relationship between skilled readers' performance in the drag and drop shape variant (*d*' scores) and performance on the (a) WRAT Reading and (b) WRAT Spelling subtest; (c) a significant negative relationship between skilled readers' performance in the drag and drop letter variant (*d*' scores) and nonword phoneme deletion latencies.

Relationship between memorization performance and the cognitive/literacy measures. The full set of correlations between the composite measure of memorization accuracy and cognitive/literacy performance is shown in Table 5.12. There were no trends or significant correlations, except for a significant positive relationship between dyslexic readers' memorization accuracy and Digit Span performance, and a small trend for positive relationship with the nonverbal IQ estimate, $r(18) = .39$, $p = .088$.

Table 5.12. Correlations between the Composite Measure of Memorization Accuracy in Experiments 4.3 – 4.4 and Performance on the Cognitive and Literacy measures (Raw Scores).

Variable	Skilled readers (n = 28)	Dyslexic readers (n = 20)
WRIT Vocabulary	.23	.08
WRIT Matrices	-.14	.39
WRAT Reading	.15	-.09
WRAT Spelling	.23	.07
WAIS Digit Span[a]	.27[c]	**.49***
WAIS Symbol Search	-.20	.23
RAN digits mean time[b]	-.06	-.15[c]
RAN objects mean time[b]	.07	-.17
NWPD latencies[b]	-.14	-.11

Note. Dyslexic readers: n = 20 due to missing background data from one participant. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = NonWord Phoneme Deletion.
[a]skilled readers: n = 27 due to missing data from one participant. [b]In seconds.
[c]Log transformed.
* $p < .05$

As shown in Table 5.13, the negative relationship found in both previous studies regarding skilled/dyslexic readers' speed of memorizatization and their performance in the Symbol Search task was no longer observed. A significant positive correlation was found between dyslexic readers' RTs and their performance on the WRAT Spelling subtest, as well as the WRIT Vocabulary subtest; the relationship between WRAT Spelling performance and memorization RTs remained when verbal IQ was partialled out, $r(17) = .55$, $p = .014$. In contrast to experiment 4.2, dyslexic readers' Digit Span performance was positively associated with their mean correct memorization RTs, indicating that longer RTs were associated with higher verbal short-term memory skill.

Table 5.13. Correlations between Mean Correct Memorization RTs in Experiments 4.3 – 4.4 and Performance on the Cognitive and Literacy Measures (Raw Scores).

Variable	Skilled readers (n = 28)	Dyslexic readers (n = 20)
WRIT Vocabulary	-.09	**.53***
WRIT Matrices	.02	.15
WRAT Reading	-.21	.20
WRAT Spelling	-.12	**.52***
WAIS Digit Span[a]	-.28[c]	**.53***
WAIS Symbol Search	-.11	-.18
RAN digits mean time[b]	.17	-.34[c]
RAN objects mean time[b]	.07	-.30
NWPD latencies[b]	.15	.08

Note. Dyslexic readers: n = 20 due to missing background data from one participant. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = NonWord Phoneme Deletion.
[a]skilled readers: n = 27 due to missing data from one participant. [b]In seconds.
[b]Log transformed.
* $p < .05$

5.3.3.4 Between-Experiment Comparisons: Experiments 4.1 – 4.4

Experiments 4.3 – 4.4 replicated the methods and procedure of Experiments 4.1 – 4.2, while introducing patterned relationships within a new type of material. Memorization and test phase data from the four experiments reported in this Chapter were subjected to a series of omnibus ANOVAs with the additional factor Stimulus Format (letter strings, shape sequences). Separate one-way ANOVAs with group (skilled readers-letter strings, dyslexic readers-letter strings, skilled readers-shape sequences, dyslexic readers-shape sequences) as a between-subject factor and performance on each of the background measures as a dependent variable (Table 5.14) confirmed that participants across the four experimental conditions were matched in terms of age, performance on the Matrices, Vocabulary and Symbol Search tasks. The main effect of group was significant in all of the remaining analyses. Levene's test for equality of variance was significant in the analyses of response latencies in the RAN Digits task, RAN Objects task and the phoneme deletion task. Subsequent Welch's ANOVAs confirmed that the main effect of group remained significant in each case: RAN digits: $F(3, 43.51) = 8.16$, $p < .001$, RAN objects: $F(3, 42.91) = 6.28$, $p = .001$, nonword phoneme deletion: $F(3, 42.60) = 18.81$, $p < .001$.

Pairwise comparisons showed significant differences in favor of skilled relative to dyslexic readers in the letter variant for performance on all measures (ps < .05). In the shapes variant, skilled readers' performance advantage held true for performance on most measures (ps < .05), but Digit Span[19] ($p = .101$) and RAN objects ($p = .247$). The two groups of skilled readers were also matched on all measures (all ps > .05). There were no statistically significant differences in performance between the two groups of dyslexic participants; however, dyslexic participants in the shapes variant tended to score higher in WRAT Reading than dyslexic participants in the letter variant, $p = .069$. A similar trend for shorter latencies among the former group was observed in the nonword phoneme deletion task, $p = .062$. Thus, it appears that the dyslexic participants in the drag and drop shapes variant were somewhat better compensated than dyslexic participants in the drag and drop letters variant.

[19] Note that the difference between skilled and dyslexic readers' Digit Span performance in the shapes drag and drop variant was significant in the original (i.e., uncorrected) independent t-test analyses reported earlier in this chapter (section 5.3.3.1). The implications of this finding are considered later on in the Discussion.

Table 5.14. Results from one-way ANOVAs Comparing Cognitive and Literacy-Related Performance across the Four Experimental Groups (Experiments 4.1 – 4.4).

Variable	Levene's	*F*	*p*	η^2
Age (years)	.202	0.76	.520	.02
WRIT Vocabulary[a]	.919	1.10	.352	.03
WRIT Matrices[a]	.302	1.99	.121	.06
WRAT Reading[a]	.062	18.82	***	.38
WRAT Spelling[a]	.457	22.33	***	.42
WAIS Digit span[b]	.200	6.56	***	.18
WAIS Symbol Search[b]	.756	2.12	.103	.06
RAN digits mean time[c]	**	10.53	***	.25
RAN objects mean time[c]	***	6.67	***	.18
NWPD time[c]	***	27.98	***	.47

Note. WRIT = Wide Range Intelligence Test. WRAT = Wide Range Achievement Test. WAIS = Wechsler Adult Intelligence Scale. RAN= Rapid Automatized Naming. NWPD = NonWord Phoneme Deletion.
[a]Standard Scores. [b]Scaled scores. [c]In seconds.
** $p < .01$. *** $p < .001$.

Memorization phase. Memorization performance was compared across stimulus format (letter strings, shapes sequences) and group (skilled readers, dyslexic readers) by means of three separate ANOVAs on the proportion of strings reproduced within a single attempt, the number of trials to criterion and mean correct memorization RTs. The analysis of the proportion of strings reproduced correctly within a single attempt revealed a significant effect of stimulus format, $F(1, 90) = 5.86$, $p = .018$, $\eta^2 = .05$, no effect of group, $F(1, 90) = 3.24$, $p > .05$, $\eta^2 = .03$, or group by stimulus format interaction, $F(1, 90) = 1.59$, $p > .05$, $\eta^2 = .02$. However, the main effect of stimulus format was qualified by a significant interaction with list, $F(1, 90) = 7.63$, $p = .007$, $\eta^2 = .07$, such that proportion accuracy was significantly worse in the shapes ($M = .80$, $SE = 0.03$) when compared to the letter variant ($M = .90$, $SE = 0.02$) in list 1, $t(42.25) = 3.30$, $p = .002$, $d = 0.92$, whereas the difference between proportion accuracy in the shapes ($M = .87$, $SE = 0.01$) and letter variant ($M = .87$, $SE = 0.02$) was not different for performance in list 2, $t(46) = 0.05$, $p > .05$, $d = 0.01$. There was no effect of list or list by group interaction, both *F*s < 1, or list by stimulus format by group interaction, $F(1, 90) = 1.36$, $p > .05$, $\eta^2 = .01$. The analyses on the mean number of trials to criterion revealed no effect of stimulus format, $F(1, 90) = 3.40$, $p > .05$, $\eta^2 = .03$, no effect of group, $F(1, 90) = 3.07$, $p > .05$, $\eta^2 = .03$, and no interaction between group and stimulus format, $F(1, 90) = 1.57$, $p > .05$, $\eta^2 = .01$. There was a conceptually similar to the previous analysis interaction between stimulus format and list, $F(1, 90) = 6.62$, $p = .012$, $\eta^2 = .06$, showing that, in list 1, participants took more trials to reach criterion in the shapes variant ($M = 1.26$, $SE = 0.04$) when compared to the letter variant ($M = 1.13$, $SE = 0.03$), $t(43.47) = 2.89$, $p = .006$, $d = 0.81$, whereas the difference between the mean number of trials to criterion in the shapes ($M = 1.17$, $SE = 0.02$) and letter variant ($M = 1.18$, $SE = 0.03$) was not different in list 2, $t(46) = 1.59$, $p > .05$, $d = 0.07$. There was no effect of list, or list by group interaction, both Fs < 1. There was no three-way interaction between list, stimulus format and group, $F(1, 90) = 1.59$, $p > .05$, $\eta^2 = .01$. The effect of stimulus format on mean correct memorization RTs was strong, $F(1, 94) = 71.09$, $p < .001$, $\eta^2 = .43$, and more reliable (that is, there were no interactions with list, and this factor was not included in the final analysis reported here). Participants' mean correct memorization RTs for the shape sequences ($M = 10538.05$, $SE = 387.51$) were almost double when compared to participants' RTs for the letter strings ($M = 5881.06$, $SE = 393.58$). The difference between skilled and dyslexic readers' RTs was not significant, neither was the group by stimulus format interaction, both $Fs < 1$.

Test phase. The omnibus ANOVA revealed a highly significant main effect of chunk strength, $F(1, 94) = 298.59$, $p < .001$, $\eta^2 = .75$ and no effect of stimulus format on mean endorsement rates, $F(1, 94) = 1.07$, $p > .05$, $\eta^2 = .01$, however, the stimulus format by chunk strength interaction reached significance, $F(1, 94) = 4.11$, $p = .046$, $\eta^2 = .01$. Simple effect analyses revealed a lower rate of low chunk strength item endorsements in the letter condition ($M = .35$, $SE = 0.02$) when compared to the shapes condition ($M = .41$, $SE = 0.02$), $t(96) = 2.68$, $p = .009$, $d = 0.54$. This is further broken down in the analyses of low chunk strength item endorsement rates. There was no difference in the rate of high chunk strength item endorsements between the letter and the shapes condition, $t(96) = 0.76$, $p > .05$, $d = 0.15$. The main effect of group was not significant, $F(1, 94) < 1$, and neither were the interaction between group and stimulus format, $F(1, 94) = 1.62$, $p > .05$, $\eta^2 = .75$, group by chunk strength, $F(1, 94) < 1$, or group by chunk strength by condition, $F(1, 94) < 1$.

A two-way group by stimulus format ANOVA on mean *d*' scores showed a marginal effect of stimulus format, due to higher discrimination ability in the letter variant ($M = 0.95$, $SE = 0.08$) when compared to the shapes variant ($M = 0.75$, $SE = 0.07$), $F(1, 94) = .3.70$, $p =$

.058, $\eta^2 = .04$. There was no main effect of group, $F(1, 94) < 1$, or group by condition interaction, $F(1, 94) < 1$. Skilled readers' performance collapsed across experimental conditions ($M = 0.85$, $SE = 0.07$) was significantly above chance, $t(57) = 12.28$, $p < .001$, $d = 1.61$, and so was dyslexic readers' performance ($M = 0.86$, $SE = 0.08$), $t(39) = 10.33$, $p < .001$, $d = 1.63$. Bias c values did not vary as a function of stimulus format, $F(1, 94) < 1$, or group, $F(1, 94) < 1$. Group and stimulus format did not interact, $F(1, 94) = 1.59$, $p > .05$, $\eta^2 = .02$. One sample t tests revealed a significant liberal bias on the part of skilled readers ($M = -0.11$, $SE = 0.03$), $t(57) = 3.58$, $p < .001$, $d = 0.47$, as well as on behalf of dyslexic participants ($M = -0.09$, $SE = 0.04$), $t(39) = 2.24$, $p < .05$, $d = 0.35$.

Sensitivity to letter position knowledge. The main effect of type of violation was highly significant in the pooled dataset, $F(1, 94) = 40.26$, $p < .001$, $\eta^2 = .28$. Furthermore, stimulus format had a significant effect on the mean rate of endorsement of low chunk strength items, $F(1, 94) = 5.61$, $p = .020$, $\eta^2 = .05$, which was qualified by a significant interaction with type of violation, $F(1, 94) = 4.35$, $p = .040$, $\eta^2 = .03$. The interaction indicated that a lower proportion of type II items were endorsed in the letter condition ($M = .39$, $SE = 0.02$), relative to the shapes condition ($M = .50$, $SE = 0.02$), $t(96) = 3.34$, $p = .001$, $d = 0.68$, whereas, the proportion of type I item endorsed were not different between the letter and shapes condition, $t(96) = 0.53$, $p > .05$, $d = 0.11$. The main effect of group was not significant, $F(1, 94) < 1$, and did not interact with stimulus format, $F(1, 94) = 2.62$, $p =.109$, $\eta^2 = .03$, or type of violation, $F(1, 94) = 2.84$, $p > .05$, $\eta^2 = .02$. The three-way interaction between group stimulus format and type of violation was not significant, $F(1, 94) < 1$.

Sensitivity to specific similarity. Mean endorsement rates for high chunk strength items were not significantly affected by stimulus format or group, $Fs(1, 94) < 1$. Stimulus format did not interact with specific similarity or group, $Fs(1, 94) < 1$. The main effect of specific similarity was significant, $F(1, 94) = 20.39$, $p < .001$, $\eta^2 = .17$, due to higher endorsements of type IV items ($M = .73$, $SE = 0.01$) than type III items ($M = .65$, $SE = 0.02$). No other effect or interaction reached significance in the pooled dataset analysis, confirming that participants performed very similarly across groups and conditions.

In sum, the test phase analyses revealed that the chunk strength, letter position, and specific similarity manipulations were effective across task versions. There were a few interactions with stimulus format pointing to differences in low chunk strength item endorsements between task versions, and these are further discussed in the section below.

Importantly, dyslexic participants did not show poorer performance than controls in any of the analyses (no effects of group or interactions involving this factor).

5.3.4 Discussion

Our previous studies revealed that dyslexic readers were equally sensitive, compared to a group of controls, to chunk frequency information and individual letter position constraints, but not to the similarity between test phase and specific training exemplars (between-experiment comparison). Furthermore, in the first study (Experiments 3.1 – 3.2), dyslexic participants tended to need more trials to reach criterion and were significantly slower than skilled readers during string memorization, whereas, in the second study (Experiments 4.1 – 4.2), dyslexic participants were significantly less accurate but not slower in string memorization. It is not clear whether these difficulties should be considered a cause or a consequence of their literacy difficulties. To tease apart between these two explanations, Experiments 4.3 – 4.4 investigated artificial grammar learning of symbol sequences that cannot be easily verbalized. Our results were clear cut and favored the second explanation.

First, in contrast to our previous studies, there were no differences in skilled and dyslexic readers' accuracy or speed of memorization of non letter sequences. There was evidence that participants adopted a more-accurate-but-slower approach to the memorization of the stimuli, despite the fact that they were not explicitly told to emphasize accuracy over speed. Importantly, the presence of an accuracy-speed trade-off was confirmed among literacy unimpaired *as well as* literacy impaired participants. It is therefore clear that all aspects of dyslexic readers' memorization phase performance were spared and very similar to controls' performance. Second, there was no evidence for group differences in test phase performance, except for the presence of a significant bias towards accepting items which was only observed in skilled readers' responses. Note that this is the opposite findings from experiments 3.1 – 3.2, where a similar bias in favor of "yes" responses was observed among dyslexic readers. It is currently not clear why this difference emerged between the two studies. Be that as it may, the crucial point is the demonstrated lack of statistical differences in skilled versus dyslexic readers' discrimination ability (*d*' scores) controlling for the contribution of strategic response biases by one group or another. Over and above the effect of chunk strength, skilled and dyslexic readers' classifications were significantly influenced by the manipulations of positional information and specific similarity. None of these factors interacted with Group, demonstrating that both of them affected similarly skilled and dyslexic readers' endorsements of low and high chunk strength items, respectively. It should be noted, though, that in contrast to these clear results, our memorization and test analyses revealed three significant influences of our manipulations of counterbalancing list: (a) Memorization accuracy for list 1 items was slightly but significantly worse than memorization accuracy for list 2 items, (b) list 1 participants needed significantly more trials to reach criterion than list 2 participants; and (c) list interacted with group in the low chunk strength subset analyses, such that, in list 2, skilled readers endorsed, overall, a significantly higher proportion of items when compared to dyslexic readers (but no difference occurred between skilled and dyslexic readers' overall rate of endorsements in list 1[20]). Even though it is difficult to propose straightforward interpretations of these effects and interaction, we wish to emphasize that there were no interactions between list and (a) chunk strength, (b) type of violation or (c) specific similarity (or any higher-order interactions involving list and any of these factors). That is, our conclusion that skilled and dyslexic readers' endorsements were similarly affected by all variables held for performance in both lists.

Between-experiment comparisons. The lack of significant differences between skilled and dyslexic readers' performance were further confirmed in the combined dataset analyses. The between-experiment analyses revealed no effect of group on any of the three indexes of memorization performance and no interaction between (a) chunk strength, (b) type of violation, or (c) specific similarity and group at test. There was no difference between skilled and dyslexic readers' *d* scores, and in fact, the difference regarding participants' response bias reported in the section above was no longer significant. Interestingly, the lack of group differences was independent of stimulus format. That is, dyslexics performed similarly to controls in the drag and drop shapes *and* letter variant. However, this finding needs to be considered in relation to the results of our previous between-experiment comparison (see section 5.2.4.3 in this Chapter) which yielded a different pattern with regards to memorization accuracy and sensitivity to specific similarity: When we aggregated across experiments 4.1 – 4.2, dyslexic readers' memorization accuracy was significantly worse than skilled readers' memorization accuracy. Similarly, dyslexics' sensitivity to specific similarity was not reliable across the two linguistic variants. It appears such small group differences are to a great extent analysis dependent and should be interpreted with great caution.

[20]It is reminded that a group by list interaction also modulated the effect of group in the context of experiments 3.1 – 3.2. Group differences on high chunk strength items endorsements were shown for list 2 but not list 1 performance. Consistent with the analyses of response bias, it was dyslexic participants who endorsed a higher proportion of high chunk strength items.

We finish by briefly considering the findings regarding the effect of *stimulus format* on memorization performance and participants' sensitivity to the three sources of knowledge manipulated in our experiments. These can be summarized as follows. First, stimulus format had an ambiguous effect on participants' memorization accuracy. That is, a lower proportion of stimuli were reproduced correctly within a single attempt in the shapes relative to the letters variant, but this was only in true for performance in one of the two counterbalanced lists (list 1). Consistently, more trials were needed for the reproduction of shapes sequences in list 1, but not in list 2. Second, presenting participants with shape sequences instead of letter strings had a clear detrimental effect on RT performance that held across both lists: Memorization times nearly doubled in the shapes drag and drop variant relative to the letter drag and drop variant. Third, a trend for weaker discrimination performance in the shapes when compared to the letter variant on the basis of chunk frequency information emerged; however, the effect of stimulus format did not reach statistical significance. The chunk strength by stimulus format interaction did reach significant, and—when broken down further—revealed that a lower proportion of type II low chunk strength items were endorsed in the letter variant relative to shapes variant. Importantly, this interaction did not alter the fact that our experimental manipulation of setting constraints on allowable letter positions was effective across tasks. Consistent with previous studies (e.g., Pothos et al., 2006), it is clear that learning of different types of information is possible independent of stimulus appearance. We further tentatively propose that the complexity of type of material used to instantiate structure may have a small influence on participants' discrimination ability and rate of endorsement of different items. However, differences appear to be subtle and are open to further replication.

5.3.4.1 Correlational Analyses

Consistent with our previous findings, Experiments 4.3 – 4.4 demonstrated that the extent to which participants erred on the first attempt across training strings was almost perfectly predictive of the length of memorization to criterion. These experiments also replicated the lack of a significant relationship between memorization accuracy/memorization speed and test phase discrimination ability among skilled and dyslexic readers. Given the consistent failure to observe a link between *d*' scores and performance in the acquisition phase of the experiment, the possibility that variability in *d*' scores may have been attenuated by the fact that everyone learned the stimuli to 100% success criterion is worthy of consideration. One could speculate that discrimination ability following exposure to a fixed number of trials may have resulted in higher variability at test, allowing more significant patterns of association to be revealed.

The composite measure of memorization accuracy was unrelated to performance in all of the cognitive and literacy measures, except for Digit Span, which was positively linked to dyslexic participants' memorization accuracy. Dyslexic readers' memorization speed was also related to their verbal-short term memory skill, but in contrast to experiment 4.1 – 4.2—where the association between memory and speed was positive—in this study we observed a negative relationship. This shift in the direction of the correlation suggests that in the face of increased stimulus and task complexity, dyslexic participants who exhibited higher verbal short-term memory skill also spent longer memorizing the shape sequences. Another difference between experiments 3.1 – 4.2 and 4.3 – 4.4 concerns the association between speed of memorization and speed of processing skill. These were unrelated in this study, a finding that may be also consistent with the notion that differences in speed of processing skill are associated with differences in memorization speed only when performance is at mastery level (i.e., in the easier letter versions of the task).

Undoubtedly, the most surprising outcome of the correlational analyses was the significant association between literacy skill and artificial grammar learning performance. In contrast to virtually all of our previous experiments, skilled readers' discrimination ability was positively correlated with their performance on the WRAT Reading and Spelling subtests, and correlated negatively with their nonword phoneme latencies. Moreover, the relationship between *d*' scores and reading performance remained largely the same after controlling for verbal IQ, demonstrating convincingly a positive association between learning and reading skill. However, in order to consider these findings in the grand scheme of things, it is reminded that (a) the association of *d*' scores and reading skill did not reach significance among dyslexic readers (and there was no evidence of a relationship between task performance and spelling skill in the dyslexic group analysis), (b) the relationship between skilled readers' *d*' scores and spelling skill was weaker and no longer statistically significant when verbal IQ was partialled out, (c) furthermore, all coefficients were modest—suggesting that only a small amount of variance associated with literacy skill is captured by implicit skill learning measured in terms of an artificial grammar learning performance. There are some open questions regarding these findings. For example, did the relationship between reading and artificial grammar learning performance emerge only among skilled learners due to our relatively restricted dyslexic sample size? Or does this finding truly suggest a weaker—though not fundamentally different—relationship between implicit learning and literacy skill among language impaired individuals? More puzzlingly, why was the relationship between learning performance and literacy skill *only* observed when participants were confronted with nonlinguistic stimuli? As shown in Table 5.15, the discrepancy between the different patterns of associations could not be simply accounted by a failure to find systematic variations in performance on the letter (typing; drag and drop) variant. The range of scores was very similar across variants, although somewhat more restricted among dyslexic relative to skilled learners.

Table 5.15. Minimum and Maximum Values of Discrimination Ability (*d*' scores) Across Artificial Grammar Learning Variants (Experiments 3.1 – 4.4) and Groups of Participants (Skilled Readers, Dyslexic Readers).

Material	Task	Group	
		Skilled readers	Dyslexic readers
Letters	Typing variant	-0.26 – 2.12	0.10 – 1.64
	Drag & Drop variant	-0.21 – 2.20	0.10 – 1.82
Shapes	Drag & Drop variant	-0.43 – 1.93	0.00 – 1.93

Another possible statistical explanation of this discrepancy concerns task variant differences in terms of their internal consistency reliability. Preliminary analyses revealed that Cronbach's alphas for the letter and shape variants (calculated separately for each counterbalanced list condition), were in fact consistently *very* low. These low internal consistencies indicate that the well-formedness task, as a whole, measured a heterogeneous set of abilities (which is probably justified, given that different groups of items were designed to assess participants' sensitivity to different sources of knowledge). Importantly, they may have attenuated the number of correlations between this measure and literacy (among other variables). We conclude by suggesting that alternative measures of implicit learning may be better suited for drawing inferences about the relationship between implicit learning and literacy skill (or even psychometric intelligence).

5.4 General Discussion

A series of experiments compared skilled and dyslexic adults' performance on Kinder and Lotz's (2009) artificial grammar learning task and two of its variants introducing changes in the induction phase task demands and in stimuli's format. First and foremost, it was shown that while a few differences between skilled and dyslexic adults' artificial grammar learning performance arose when stimuli consisted of letter strings, dyslexics' ability to learn structure embedded in nonlinguistic stimuli was spared and at control levels. Therefore, our findings provide evidence against the notion that an implicit learning deficit (affecting, for example, dyslexics' ability to extract orthographic patterns, Sperling et al., 2004) is causally related to reading and spelling disability. It should be noted, though, that the dyslexic sample used in the last study (Experiments 4.3 – 4.4) was possibly, albeit inadvertently, less representative of a general dyslexic population than the samples used in Experiments 3.1 – 3.2 and Experiments 4.1 – 4.2. There are two observations consistent with this suggestion. First, we obtained trends for lower WRAT reading performance and longer nonword phoneme deletion latencies among dyslexic participants in the letter variant compared to the drag and drop shapes variant (section 5.3.3.3). Second, the excessive number of trials to criterion required by five dyslexic participants in the shapes variant made the length of session prohibitive for them, and we did not consider their performance in any of our analyses. Thus, even though we report data from a diagnosed dyslexic population exhibiting significant deficits in literacy (reading, spelling) skill and phonological processing skill (e.g., phonological awareness skill) and our sample is adequately sized (n = 21), one could question whether our findings also hold for the general dyslexic population. Given the challenging nature of our lengthy training phase, this issue would have to be resolved in a future study that used fewer or a fixed numbers of training trials.

Second, taken together, our three studies, provide converging evidence that artificial grammar learners' performance is substantially influenced by their sensitivity to bigrams and trigrams that occur frequently among training stimuli, and, to a smaller extent, whole training stimuli, and positional constraints on the allowable position of stimulus' elements (Kinder, 2000; Kinder & Lotz, 2009). Participants' sensitivity to chunk strength differences is well-established in the artificial grammar learning literature (e.g., Knowlton & Squire, 1994, 1996;

Meulemans & Van der Linden, 1997). However, we wish to highlight again that, sensitivity to this factor was not measured independently of sensitivity to abstract rule knowledge in this study. Items of low associative chunk strength were also grammatical in nature, whereas items of low associative chunk strength were also inconsistent with the underlying grammar. Establishing the extent to which stimuli's rule adherence (or any other factors confounded with rule adherence; see Johnstone & Shanks, 1999) may have influenced grammaticality judgments over and above chunk frequency information was beyond the scope of our study.

We end with a final note on the interpretation of the main effect of specific similarity—shown quite consistently across studies and groups of participants (however, see section 5.2.4.3 for a significant similarity by group interaction). While our demonstration supports Kinder and Lotz's (2009) claim that knowledge of entire training letter strings/shape sequences influences participants' judgments of grammaticality, it is not consistent with previous studies (e.g., Knowlton and Squire's [1994] re-examination of Brooks and Vokey's [1991] findings; see section 1.1.3 of the Literature Review) showing that specific similarity does not influence performance when similar and dissimilar items are matched for associative chunk strength. There are several possible explanations for this discrepancy. First, as Kinder (2000) suggests (see also Vokey & Brooks, 1992), participants' sensitivity to specific similarity may have been a by-product of the lengthy training phase duration (extensive number of repetitions of a small-sized set of training strings) that was adopted in their—as well as in our—experimental investigation. Second, it worth noting that the factor of specific similarity (similar vs. dissimilar) was not orthogonally manipulated with that of associative chunk strength (high vs. low). It is possible that participants may have been less sensitive to this source of knowledge among low chunk strength items. The discrepancy between this and previous findings regarding the effect of specific similarity may have also resulted from Kinder's (2000) criterion of "dissimilarity". That is, while similar items differed from their closest training item by one letter—as in most studies using Brooks and colleagues' operationalization of specific similarity—dissimilar items differed from their closest training item by 3 letters which constitutes a maximal difference relative to previous studies. Finally, closer inspection of the material used by Kinder and Lotz (2009) revealed another dimension which is relevant to participants' sensitivity to the factor of specific similarity. Out of the 12 type IV similar items developed by Kinder (2000), only 2 items (in list 1 as well as in list 2), included a letter change in nonanchor positions (e.g., MP*V*RZQZ differs from its closest training item, MP*D*RZQZ, in terms of one letter located in position 3), whereas the remaining

10 items included a letter change in the first or last position (e.g., *R*TVRTXJ differs from its closest training item, *F*TVRTXJ, in terms of one letter located in position 1). This was not the case in Vokey and Brooks' (1992) study where letter change occurred in initial, middle as well as terminal string positions.

Chapter 6

General Discussion

6.1 Summary of Results

This thesis is concerned with humans' ability to learn statistical information embedded in written language under conditions that involve no instruction to learn. Implicit skill learning, conceptualized as the converse of the ability to learn through a process of deliberate hypothesis testing (e.g., Frensch & Rünger, 2003; Perruchet, 2008; Seger, 1994) has been widely suggested to be important for reading and spelling acquisition (e.g., Gombert, 2003; Pacton et al., 2001; Steffler, 2001). Statistical learning research has also shown that a range of statistics (pair frequencies: e.g., Marcovitch & Lewkowicz, 2009; conditional probability between adjacent and some nonadjacent elements: e.g., Aslin et al. 1998; Newport & Aslin, 2004) can be encoded from as early as infancy and used to perform various language-related tasks, from word boundary identification (e.g., Saffran et al., 1996a, 1996b) to the discovery of rudimentary syntax (e.g., Gómez & Gerken, 1999; Saffran &Wilson, 2003). A statistical learning account of learning to read and spell has been recently proposed (e.g., Pollo et al., 2007), nevertheless, little research has applied the methodological framework of both streams of research (implicit learning, statistical learning) to a systematic investigation of their contribution to spelling acquisition and literacy impairment. The work described in this thesis was aimed to empirically assess whether statistical information governing aspects of correct spelling can be learnt under incidental conditions and whether impaired implicit skill learning/sensitivity to statistical information can account, to some extent, for dyslexic individuals' literacy difficulties. Within such a broad domain, our starting point involved the ability to learn one particular aspect of orthographic knowledge: Graphotactic constraints on the position and context of letter distributions.

Our first series of experiments (Chapter 2 & 3) validated and extended Onishi et al.'s (2002) methods of inducing novel phonotactic learning to the visual domain (the incidental graphotactic learning task). Pilot studies established that novel graphotactic constraints embedded within visually presented syllables can be learnt by adults following brief incidental exposure. This is an important finding corroborating that similar language-wide constraints can be learnt without explicit instruction, as previously hypothesized (e.g., Pacton et al., 2001). The experiments also provided supporting evidence for the widely held notion that statistical learning is domain-general, rather than specific to language (e.g., Kirkham et al., 2002; Saffran et al., 1999). Learning performance was reliable not only with pronounceable CVC strings, but also when similar statistical patterns were embedded within Greek-letter (i.e., nonlinguistic) sequences. In fact, the magnitude of the learning effect was unaffected by stimulus format. Finally, consistent with the findings of previous phonotactic studies (e.g., Warker & Dell, 2006), it was shown that participants' learning was modulated by pattern complexity. Sensitivity to the allowable position of letters was significantly stronger than sensitivity to context-based patterns. However, there were some limitations in our graphotactic learning demonstrations. First, learning did not generalize to unseen items conforming to the graphotactic constraints in the harder-to-learn contextual constraints conditions. Second, participants' speed of responding proved only partially sensitive to our manipulation of legality. In only two out of five experimental conditions, reaction times associated with correct rejections of illegal items were significantly faster than RTs associated with correct hits to pattern-conforming stimuli.

Correcting for some basic methodological issues, two of the pilot study research questions (effect of pattern complexity on learning performance; relationship between learning and literacy skill—discussed at the end of this section) were addressed by means of a similar task among school-aged (7-year-old) children and adults. Further to these more specific questions, the study presented in Chapter 3 was set out to establish whether statistical learning processes underlying graphotactic learning operate from early on in development. This was confirmed in light of evidence that novel graphotactic learning was reliable not only among adults, but also from as early as 7 years of age. The implications of this core finding for "late" models of spelling development (e.g., Frith, 1985) and a consideration of the characteristics (e.g., richness) of the material used to demonstrate learning in our study are discussed in more detail below (sections 6.2 - 6.4). A clear developmental effect was observed for performance in the positional constraints learning condition, such that 7-year-old children performed less well than adults; a similar pattern was observed with regards to performance in the contextual constraints learning condition, although the difference between children's and adults' discrimination ability was not reliable in a secondary set of analyses. By and large, in our study, adults were superior learners than children, a finding that goes against Reber's (1993) claim that implicit learning is age-invariant. This finding also contradicts some previous statistical learning studies showing equivalent levels of performance between different age populations (e.g., Saffran et al., 1997). It is suggested that the extent to which age effects are shown to operate may be dependent on the complexity of the material used to assess learning. Regarding the effect of this factor, it was shown that, as in the pilot studies, both children and adults found positional constraints learning easier than contextual constraints learning. It was originally suggested that this demonstration is consistent with previous developmental studies on the acquisition of language-wide spelling-to-sound conditional patterns (e.g., Treiman & Kessler, 2006). However, following an in-depth consideration of the implications of using a single-letter colour detection task during exposure, it was acknowledged that the effect may have partly resulted from the nature of our cover task: diverting participants' attention to single letters may have worked in favor of learning single-letter constraints, and against the learning of context-based constraints. Finally, while discrimination accuracy was above chance in both age groups, it was only for adult participants that reaction times varied as a function of legality, such that illegal items were rejected faster than legal items were endorsed.

A different population (groups of dyslexic adults—and well-matched controls) was assessed in the context of studies presented in Chapters 4 and 5. In addition, we opted for a significant change in the methods used to assess skill learning by employing a core experimental tool for the study of the implicit learning phenomenon, the artificial grammar learning task. Several questions were addressed in these last two chapters, but in consideration of space (and repetitiveness) we mainly summarize the results of those that are relevant to the dyslexic/control group comparisons. The first hypothesis under scrutiny was whether implicit learning is impaired in dyslexia, as has been previously argued by Nicolson, Fawcett and colleagues (e.g., Nicolson & Fawcett, 1990; Nicolson et al., 2001). We used a non-motor implicit learning task, and going beyond the few previous studies of implicit skill learning in dyslexia (e.g., Pothos & Kirk, 2004; Rüsseler et al., 2006) and the general claims of the automatization deficit theory, we assessed dyslexics' (and controls') sensitivity to *different* sources of information (allowable chunks, positional constrains, specific similarity) on the basis of Kinder and Lotz's (2009) design. It was argued that all three factors manipulated in our studies may map onto real orthographic learning constructs which guide developing children's reading aloud/spelling performance even though they are not explicitly pointed out to them (i.e., through mere but repeated exposure to written language). Therefore, our experimental work holds important implications for our understanding of how impaired implicit learning may be of relevance to dyslexics' literacy difficulties, a point further discussed in section 6.2. From the very beginning of our investigations, we obtained some evidence of spared learning in dyslexia. Dyslexic adults were equally sensitive relative to controls to chunk frequency information, on the basis of which they reliably discriminated between grammatical and ungrammatical test phase items. The effect of positional constraints on participants' rate of endorsement of low chunk strength items was reliable following a modification of the training phase processing demands (that is, in the drag and drop artificial grammar learning variants). However, once a significant contribution of positionally constrained information to learning was obtained, this affected skilled and dyslexic readers' responses similarly. Therefore, there was no evidence of a deficit in dyslexic adults' ability to exploit cues on the allowable position of letters—a type of information that is frequently constrained in written language (Treiman, 1993). Sensitivity to specific similarity on the other hand, mapping indirectly onto the ability to use analogies for reading and spelling words (Goswami, 1988; Ehri, 1997) was not always spared. However, the presence of a selective impairment depended on the type of analysis performed. Nevertheless, equivalent levels of sensitivity to all three types of knowledge, including specific similarity, were unequivocally shown in a nonlinguistic artificial grammar learning variant; importantly, in the latter, there were no group differences either at test or during learning (i.e., during stimulus memorization). Thus, whatever signs of impairment arose among dyslexics in the letter-string version of the learning task, these disappeared when the task comprised nonverbal unpronounceable symbols. However, some caution is drawn to the possibility that dyslexic participants in experiments 4.3 – 4.4 may have been relatively better compensated.

Apart from the specific purposes stated above, another general aim of the thesis was to investigate, in a correlational manner, the relationship between implicit/statistical skill learning and general reading/spelling proficiency. The latter was mainly measured by performance on the WRAT Reading and Spelling subtests (see, however, Chapter 3 for attempts with other literacy measures). Our studies with dyslexic adults necessitated the additional inclusion of oral language measures (e.g., a phoneme awareness task); therefore, we also investigated the association of implicit learning with measures more indirectly related to literacy skill. Finally, motivated by Reber's (1993) claim that individual differences in psychometric intelligence correlate with performance on explicit, but not implicit learning tasks, we also investigated the relationship between IQ (and some of its basic contributors) and artificial grammar learning (Chapter 4 and 5). The findings with regards to the first two questions are very easy to summarize: we obtained no evidence of an association between task performance and reading or spelling skill in either skilled/dyslexic adults or during earlier stages of literacy acquisition regardless of the type of measure used to assess literacy skill. There was one exception. In the shapes variant of the artificial grammar learning task, higher reading skill and faster phoneme deletion latencies were associated with better discrimination ability among skilled readers. We propose an interpretation of these findings and discuss their implications for future research in the next section. Our findings with regard to the last question, however, were relatively inconsistent and perhaps even spurious given the large number of correlations performed (and the resulting increased chance of type I error). Thus, it is safer not to go further than the tentative explanations proposed earlier in the thesis.

6.2 Implications for Literacy Development and Literacy Impairment

As briefly explained in Chapter 3 (see also section 1.4 in the Literature Review), an early body of literature suggested that learning to read and spell involves successful progression through a series of stages or phases, during each one of which, specific milestones are reached (Frith 1985; Gentry, 1982). Two core claims underlie this conceptualization of literacy skill development: First, stages are qualitatively different, that is children "situated" in the so called alphabetic stage approach the reading aloud/spelling task in a different way than children in the so called orthographic stage; and second, progress from one stage/phase to another occurs in a strict sequential manner. There are three such stages involved in the process of becoming a skilled reader and speller in Frith's (1985) terms: (a) a logographic stage, during which children are unaware of the relationship between letters (graphemes) - sounds (phonemes) and vice versa, such as, for example, children's spelling attempts mainly consist of scribbles, drawings, and perhaps reproductions of parts or wholes of very high frequency words in the environment; (b) an alphabetic stage, during which knowledge of letter-sound connections begins to emerge and gradually governs children's reading and spelling, and (c) an orthographic stage, during which more "mature" strategies allow children to begin reading correctly and start producing conventionally accurate spellings. As pointed out by Nunes and Bryant (2009), a comprehensive list of the strategies used by children while in the latter stage is lacking in Frith's (1985) model. However, her framework has been generally taken to imply that it is only in that very *last* orthographic stage that knowledge of orthographic conventions starts to exert its influence on children's spellings (e.g., Pollo et al., 2007).

There is now an abundance of evidence against the view that children's spelling mistakes and experimentally measured judgmental preferences vary in a "phase/stage-appropriate" manner (Cassar & Treiman, 1997; Lehtonen & Bryant, 2005; Pacton et al., 2001; Treiman, 1993). As variously demonstrated by these authors, sensitivity to untaught positional restrictions (e.g., *ck* does not occur in syllable-initial positions in English) and orthographic constraints on allowable doublets (e.g., *hh* never occurs in English) is evident from the very beginnings of literacy development. While, none of the above findings are accountable by Frith's (1985) or Gentry's (1982) account of reading and spelling development, the question of "how" children succeed in incorporating such orthographic knowledge into their own spellings (e.g., Treiman, 1993) has been only hypothesized to date (e.g., Pacton et al., 2001; Pacton, Sobaco, Fayol, & Treiman, 2013; Pollo et al., 2007). The main contribution of the work presented in Chapter 3 is the elucidation of the learning mechanisms that allow children to avoid some orthographically illegal errors in the absence of explicit instruction. Thus, the need to adopt a more flexible view of spelling development is corroborated by our evidence that children are cognitively ready to start exploiting the statistical nature of their orthography. Our evidence also indirectly suggests that other statistically-based sources of knowledge that have similarly hypothesized to emerge "late" may also be learnable early on in development.

Having established that statistical skill learning plays an important role in literacy development, it is interesting to explore whether impaired skill learning/sensitivity to statistical cues may be part of the reason why spelling problems that are resolved relatively easily over the course of typical development persist unexpectedly in dyslexia. On the basis of similar considerations, a general skill learning deficit has been sometimes considered to be a parsimonious account of dyslexic individuals' difficulties in reading and other tasks that become over learnt through extensive repetition (e.g., Nicolson & Fawcett, 1990; Nicolson et al., 2001). Several studies have claimed to show deficits in dyslexic individuals' performance relative to controls in the simplest overlearned tasks (namely, postural stability; e.g., Nicolson & Fawcett, 1990, 1994, 1995) and, more recently, their ability to learn implicitly sequential structure (e.g., Howard et al., 2006; Stoodley et al., 2006; Vicari et al., 2003) and artificial grammars (Pavlidou et al., 2010; Ise et al., 2012). The findings are not consistent, but considering the potential influence of other co morbid disorders and the variety of experimental approaches that have taken to measure implicit learning in dyslexia, this is not a surprising finding. Taken together, the results of Chapter 4 and 5 weaken the claim that deficits in implicit skill learning are a direct cause of literacy impairment. It is therefore suggested that, while statistical learning mechanisms are important for reading and spelling acquisition, they operate similarly in dyslexic and nondyslexic adults.

The final approach adopted in this thesis for establishing links between literacy acquisition and implicit/statistical skill learning was a correlational one. Correlation is, of course, not causation; however, if a link exists between these skills, it is logical to anticipate a positive association showing that individual variation in learning and literacy/literacy-related performance goes hand in hand. Our approach was certainly not unfounded, given, for example, Arciuli and Simpson's (2012) or Howard et al.'s (2006) results discussed in Chapter 2. However, with one exception, the hypothesized positive correlations were never observed in our studies. These null findings, though, are not taken to suggest that implicit/statistical learning is irrelevant to reading or spelling performance. Rather, they may suggest that correlation is not a sensitive way of establishing associations with the type of measures that we chose to employ throughout the thesis. For correlation to occur, statistical properties such high internal consistency reliability (e.g., high Cronbach's alpha), are essential. However, all of the items used in our learning "tests" were selected to provide appropriate balance between different conditions/factors and were not constructed to measure sensitivity to the same type of knowledge. It is hard to envisage a similar level of experimental control if the main criterion for stimulus selection was the maximization of the artificial grammar learning task's internal consistency (as for example, in Gebauer & Macintosh, 2007). As suggested in section 6.5, it may be the case that other measures are less ill-suited for showing correlational links. This aspect of our design may also explain why significant correlations have been sometimes documented in previous serial reaction time task studies such as Howard et al.'s (2006).

6.3 Incidental Graphotactic Learning – Artificial Grammar Learning: Same or Different?

In drawing final conclusions about the type of learning that was shown to operate among literacy unimpaired/impaired adults and typically developing children, it is deemed necessary to take into account the differences, as well as the similarities between the two methodologies employed in this thesis. To begin with, the incidental graphotactic learning and the artificial grammar learning task are procedurally similar in that they both comply with the conceptual design of most implicit/statistical learning tasks, whereby learning is induced and measured via a 2-phase (acquisition, test) procedure. Generally speaking, exposure to pattern-conforming stimuli always occurred incidentally and to the satisfaction of another objective set by the experimenter ("spot the red letter"; or "memorize the letter string/shape sequence"). While learners were blind to the patterned nature of the stimuli, they were informed about the "true" purpose of their exposure prior to test in all of our studies. From that stage on, sensitivity was assessed via participants' explicit judgments on novel stimuli's well-formedness, enquired by means of two similarly worded questions (incidental graphotactic learning task: *does the stimulus go well with the stimuli you saw in the previous phase of the experiment?*; artificial grammar learning task: *does the stimulus conform with the rules used to construct the previous set of stimuli?*). Participants' levels of conscious awareness of what was driving these judgments were not measured in either study, and admittedly, may have varied within, as well as between tasks. However, defining and measuring awareness is a problem that is not straightforwardly solved (Cleeremans et al., 1998; Shanks & St John, 1994) and learning was taken to be implicit to the extent that it occurred under incidental conditions that did not promote explicit "rule" searching strategies (see also, Cleeremans et al., 1998; Saffran et al., 1997).

Beyond these general similarities, it goes without saying that the two tasks used as measures of skill learning differ in a number of ways, among which, the amount of exposure given (number of trials); individual trial duration; the design and complexity of the stimuli used; their induction phase characteristics and demands. The latter appears to be a critical factor over and above stimulus complexity, as suggested by Whittlesea and collaborators' episodic-processing theoretical framework (e.g., Whittlesea & Dorken, 1993), reviewed in Chapter 5. Further to the hypothesized differential influence that our cover (colour detection) task may have had on participants' performance in the positional and contextual constraints learning conditions of the incidental graphotactic learning task (see sections 3.3 and 6.1), we speculate that a similar manipulation within our artificial grammar learning task may have had a more detrimental effect on participants' overall artificial grammar learning performance and/or more specific aspects of informational sensitivity. For example, previous artificial grammar learning studies have shown that under conditions where patterned knowledge is embedded in different stimulus aspects (e.g., letter identity order; letter colour order), only those that have been attended as part of the experimental instructions are reliably learnt (e.g., Eitam et al., 2009). It is therefore, reasonable to assume that being asked to detect single letters may have attenuated artificial grammar learners' sensitivity to small/large chunks (and even constraints on single letters, to the extent that these would not always coincide with the letters selectively attended). Predicting how exposure length may have differentially affected artificial grammar/incidental graphotactic learners is somewhat more unclear. For example, an investigation of the effect of learning exposure in Chapter 2 by comparing performance under a standard-length (i.e., as in the pilot study) and double-length versions of the task was, to our surprise, unsuccessful.

At this point, it is important to point out that task differences are not considered in an attempt to generalize findings from one set of studies to another. There is little doubt, for example, that memorizing 96 seven-letter-long strings would be an impossible task for most 7-year-old children (however, note that child-friendly and even infant-friendly adaptations of the artificial grammar learning procedure have been successfully used to demonstrate learning of some dependencies, such as pairwise element violations; Ise et al., 2012; Gómez & Gerken, 1999). Task similarities and differences are considered, though, for the interpretation of seemingly contradictory findings, such as the influence of positionally constrained information on participants' judgments of legality/grammaticality in Chapters 3 and Chapters 4-5. That is, learning in the positional constraints condition of the incidental graphotactic learning task was very reliable not only among skilled adult learners, but also among 7-year-old typically developing children. Positional sensitivity in Chapter 4 and 5, on the other hand, was reliable only within one of the two artificial grammar learning variants used in the thesis, and unarguably, exerted a smaller influence on participants' judgments of grammaticality compared to chunk frequency information. This apparent inconsistency is easily reconciled by a closer examination of the material used in the two tasks, illustrated by an analysis of the incidental graphotactic learning stimuli properties along the lines of the metrics used to quantify stimuli differences in the artificial grammar learning task. Legal unseen stimuli (e.g., *det*) in the positional constraints incidental learning condition always conformed to the zero-order constraints that the letters *d* and *t* can appear in position 1 and 3, respectively. In addition, the stimuli consisted solely of allowable and non-novel bigrams (*de*, *et*): their global/anchor associative chunk strength was 1.33, and their global/anchor chunk novelty was zero; finally, all of them were highly similar (i.e., differed by a single letter) to 4 specific training items (*den*, *des*, *dot*, *met*, *fet*). On the other hand, illegal unseen items (e.g., *tod*) in the same condition violated both positional constraints (i.e., letters *t* and *d* cannot appear in position 3 and 1, respectively), and also consisted solely of illegal and novel chunks (*to*, *od)*. Their mean global/anchor associative chunk strength was zero, their global/anchor chunk novelty was 1.33. Finally, they were all dissimilar (i.e., differed by 3 letters) from *each* of the training items. The important message is that more than one of these systematically confounded factors may have contributed to participants' higher rate of endorsements for legal items in Chapter 3. That is, sensitivity to positionally constrained information was not assessed over and above participants' sensitivity to bigram/trigram information or overall stimulus similarity, as in experiments 3.1 – 4.2. Adding to the fact that positional constraints were always embedded in anchor positions in experiments 2.1 – 2.8 (but not in experiments 3.1 - 4.2), it may be contended that positional sensitivity was much more salient in the former series of experiments, hence exerted a stronger influence than in the artificial grammar learning task.

6.4 Limitations

Specific limitations of the studies reported in this thesis were addressed in each of the previous chapters. In this section, we discuss a more general methodological concern to be considered in the interpretation of our findings: the lack of control group data to serve as the basis of chance performance (instead of assuming that chance accuracy always corresponds to 50%, that is, *d*' = 0). The need for control groups has been noted early on in the literature (e.g., Redington & Chater, 1996) and is further highlighted in more recent paper by Reber and Perruchet (2003; however, see Dienes & Altmann, 2003 for a different view as to whether controls group should not be "untrained", i.e., proceed immediate to the test phase of the experiment). As Redington and Chater (1996) point out, grammatical and ungrammatical stimuli may differ in a number of unpredictable ways, which can potentially lead to an inherent bias towards endorsement. Thus, simply comparing against chance does not rule out that sensitivity might have been driven, to some unknown extent, by task-irrelevant stimuli characteristics. This argument is taken further by Reber and Perruchet's (2003) demonstration that stimuli similarities/differences in terms of *nonspecific* factors are usually outside the control of the experimenter, whose primary concern is the manipulation of stimuli differences/similarities in terms of *specifi*c variables (e.g., grammaticality; chunk strength, similarity). While these authors provide a comprehensive list of potentially confounding factors (e.g., same first/last letter within a string identity, repetition of consecutive letters, etc), the list is by no means exhaustive. It is therefore suggested, that the work reported in this thesis should be complemented by control group conditions.

We conclude with another note of caution on the type of materials used in this and in previous lab-based demonstration of implicit/statistical learning. As discussed in chapter 1, the lack of ecological validity has been a longstanding issue in statistical learning research and has important implications for the generalizability of lab-based findings to real language learning (Pacton et al., 2001; Pelucchi et al., 2009). This concern is further highlighted by Taylor and Houghton's (2005) demonstration that, artificial phonotactic learning induced under strict laboratory settings fades much quicker than similarly induced "learning" (i.e., reinforcement) of real phonotactic constraints. Along the lines of these concerns, it is acknowledged that the grammar, as much as the stimuli used in our artificial grammar learning experiments are indeed, quite artificial. To name only a few of their differences from the words embedding patterns in real language, none of the stimuli are fully pronounceable or carry meaning. However, this is not necessarily only a drawback. The level of manipulation/control achieved with this type of material is certainly very appealing.

The problem of ecological validity was, to some extent, taken into account in our initial experimental investigations. We used pronounceable graphotactically constrained syllables, instead of using visuo-spatial arrays (e.g., Fiser & Aslin, 2001); sequences of shapes (Fiser & Aslin, 2002; Kirkham et al., 2002); or "alien" figures (Arciuli & Simpson, 2011, 2012). However, opting for a frequent-syllable-structure (Consonant-Vowel-Consonant) and pronounceable stimuli was of cost to another aspect of ecological validity, namely complexity. Participants were presented with identical-length strings, embedding deterministic (i.e., all-or-nothing) constraints, and in some cases, more than one sources of consistent information (as explained in section 5.3). All of these parameters point to the fact that our task may have been an underestimation of the extent of variation and complexity found in natural orthographies. For example, contextual constraints regarding permissible orders are usually probabilistic rather than deterministic, at least in orthographies as inconsistent as English (note though, that this is not the case in more consistent orthographies such as Turkish, or Finnish). Furthermore, we targeted learning of a single type of statistical information, whereas natural languages are replete with statistical cues assisting learners in different linguistic-related tasks. Therefore, even though Chapter 3 demonstrated convincingly that incidental learning occurs remarkably quickly in 7-year old children and adults, it is very likely that the challenge faced by statistical learners in the laboratory is not truly comparable to the challenge of real statistical learning.

6.5 Future Research

There are several possibilities to extend the research presented in this thesis. Future studies into the relationship between literacy skill and implicit learning could, for example, benefit from a different and/or a more indirect approach to measuring learning than the well-formedness task. A prime example of such indirect measures of learning, whereby participants need never to be informed about the rule-governed nature of the stimuli, is the RT-based score used within Nissen and Bullemer's (1987) serial reaction time task. There are more options, though, than simply replicating the finding that RT-based serial learning sometimes correlates with individual differences in literacy (e.g., in Howard et al., 2006, but not in Waber et al., 2003). Another alternative is, for example, Karpicke and Pisoni's (2004) immediate memory span task that was briefly presented in Chapter 4 (section 4.2.4.3). It is suggested, that such indirect measures of learning could be a promising way forward in establishing whether implicit learning is in general invariant of literacy or general intellectual skill, as suggested by Reber (1993), or rather, whether grammaticality judgments are insensitive to such individual variability.

Many interesting questions and variations of the incidental graphotactic learning task await investigation. Future research into this area of experimentation could benefit from a replication of our main effect of pattern complexity under conditions where attention does not bias participants towards attending a single letter. For example, an equally simple but more neutral task relative to the one used in our experiments could involve responding to all three letters turning red. It would be also interesting to directly compare learning differences as a function of the extent of overlap between the size of the letter unit targeted in the exposure phase and the size of the unit needed for stimulus discrimination. In addition, as addressed in Chapter 3, learners' ability to discriminate between graphotactically legal and illegal test phase items may have been additionally guided by the extraction of novel phonotactic constraints (e.g., the sound /d/ is only legal as an onset), which systematically covaried with the novel graphotactic constraints within our material (i.e., the letter d is only legal as an onset). Assessing learning of similar patterned relationships embedded within sequences of nonliteral symbols (such as the Greek letters used in the pilot study reported in Chapter 2) is certainly a legitimate option. However, it is conceivable that a nonlinguistic variant may have required longer training with 7-year-old children. Using some homophonic letters (i.e., letters representing the same phoneme; e.g., c and k map the phoneme /k/) would not only allow us to tease apart purely graphotactic from joint grapho- and phonotactic learning, but would also comply with the aim of using stimuli from a true orthography. A 2 by 2 design pitting stimuli's legality (legal, illegal) against their compliance with graphotactic versus grapho-phonotactic constraints would therefore provide a direct comparison between learners' ability to discriminate stimuli on the basis of these two sources of statistical information.

There are also possibilities for future research with regard to the artificial grammar learning task employed in this thesis, many of which were mentioned in the discussion of the findings of Chapter 4 and 5. Our finding of spared learning in the nonlinguistic variant has crucial implications for theories viewing dyslexia as a developmental disorder that involves impaired automatization and clearly merit further research. For example, the role of the memorization phase duration and its contribution to the strikingly similar levels of above chance discrimination ability demonstrated by our samples of skilled and dyslexic readers remains somewhat unexplored. All three artificial grammar learning experiments reported in this thesis employed an extensive number of memorization trials, which was in fact further prolonged, as most participants failed to reproduce all of the items within their first attempt. One could perhaps argue that, by training participants to a 100% success criterion, we effectively enabled dyslexic participants to compensate for any difficulties in pattern induction skill, and therefore indirectly ensured that skilled and dyslexic participants would be similar in discrimination ability at test. It would be interesting to compare skilled versus dyslexic readers' test phase performance following a similarly large but *fixed* number of memorization trials (e.g., 8 repetitions/trial regardless of whether the stimulus was correctly reproduced or not) or a smaller number of trials to criterion (e.g., at least 4 repetitions/trial).

Last but not least, an interesting question to be addressed by future research is whether statistical information governing other aspects of literacy can be efficiently extracted by dyslexic individuals under implicit conditions. Learning orthography to phonology associations is widely accepted as a critical step of reading acquisition (e.g., Frith, 1985), whereas failure to do has been so has been proposed as a "missing link" in explaining the relationship between dyslexic readers' phonological impairments and reading aloud impairments (e.g., Aravena, Snellings, Tijms, & van der Molen, 2013; Froyen, Willems, & Blomert, 2011; Hatcher et al., 1994). Recent evidence by Taylor et al. (2011) shows that, under laboratory conditions, the ability to map artificial characters onto phonemes may occur among adults via mere (implicit) exposure to monosyllabic "words" spelled in an artificial orthography. Artificial orthography paradigms such as the one used by Taylor and colleagues could be used to extend our work by comparing skilled and dyslexic readers' ability to learning statistical patterns governing phoneme-to-grapheme associations rather than orthographic patterns in the visual modality.

6.6 Conclusions and Final Remarks

The experimental work reported in this thesis furthers the literature by showing that statistical learning processes that have been previously implicated in phonotactic, prosodic and syntax acquisition (to name a few domains of linguistic performance) come similarly into play in the orthographic knowledge acquisition process. Our findings do not deny the importance of explicit instruction. Instead, they are taken to show that, while explicitly taught orthographic conventions (e.g., i before e except after c) may aid children in learn to spell, as pattern complexity increases, there are other, and perhaps more efficient learning mechanisms operating that also account for children's ability to read and spell by convention. This body of work also constitutes one of the first well-controlled studies of implicit learning among dyslexic adults. There is more to learn about the effectiveness of implicit/statistical learning processes underlying orthographic knowledge acquisition and sensitivity to other statistical cues governing correct reading and spelling by studying learning in dyslexic children.

References

Alegria, J., & Mousty, P. (1996). The development of spelling procedures in French-speaking, normal and reading-disabled children: effects of frequency and lexicality. *Journal of Experimental Child Psychology, 63*, 312-338.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 899-912.

Aravena, S., Snellings, P., Tijms, J., & van der Molen, M. W. (2013). A lab-controlled simulation of a letter–speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology, 115*, 691-707.

Arciuli, J. & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science, 14*, 464-473.

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

Arciuli, J., & von Koss Torkildsen, J. (2012). Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. *Frontiers in Psychology*, 3, 1-9.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*, 170-176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology, 46*, 81-86.

Beaman, C. P., & Röer, J. P. (2009). Learning and failing to learn within immediate memory. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bell, R., Röer, J. P., & Buchner, A. (2013). Irrelevant speech disrupts item-context binding. *Experimental Psychology, 60*, 376-384.

Bennett, I. J., Romano, J. C., Howard Jr, J. H., & Howard, D. V. (2008). Two forms of implicit learning in young adults with dyslexia. *Annals of the New York Academy of Sciences, 1145*, 184-198.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*, 858-888.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., et al. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49*, 422-432.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness *Nature, 271*, 746-747.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature, 301*, 419-421.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 170-211). Hillsdale, NJ: Erlbaum.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120*, 316-323.

Caravolas, M. (2004). Spelling development in alphabetic writing systems: A cross-linguistic perspective. *European Psychologist, 9*, 3-14.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751-774.

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology, 92*, 307-321.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litaviský, M., Onochie-Quintanilla, E., … Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23*, 678-686.

Caravolas, M. & Mikulajová, M. (2008). Effects of sound letter consistency, letter-form complexity, and frequency in learning canonical and contextually conditioned letter spellings in Slovak. *Phonetica Pragensia, XI*, 1-10.

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology*, *92*, 107-139.

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89*, 631-644.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69-77.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A Vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 821-828.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2011). Representations for phonotactic learning in infancy. *Language Learning and Development, 7*, 287-308.

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition, 116*, 382-393.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*, 406-416.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NJ: Erlbaum.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356-371.

Conway, C. M., & Christiansen, M. H. (2002). Sequential learning by touch, vision, and audition. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, N.J.; Lawrence Erlbaum.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1119-1130.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience, 9*, 522-533.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Deacon, S. H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology, 49*, 118-124.

Deacon, S. H., LeBlanc, D., & Sabourin, C. (2011). When cues collide: children's sensitivity to letter- and meaning-patterns in spelling words in English. *Journal of Child Language, 38*, 809-827.

Dienes, Z., & Altmann, G. T. (2003). Measuring learning using an untrained control group: Comments on R. Reber and P. Perruchet. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56(A)*, 117-123.

Dienes, Z., Broadbent, D., Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 875-887.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355-1367.

Ehri, L. C. (1976). Do words really interfere in naming pictures? *Child Development, 47*, 502-505.

Ehri, L. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 323-358). Newark, DE: International Reading Association.

Ehri, L. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*, 116-125.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice* (pp. 237-269). Mahwah, NJ: Erlbaum.

Eitam, B., Schul, Y., & Hassin, R. (2009). Goal relevance and artificial grammar learning. *Quarterly Journal of Experimental Psychology, 62*, 228-238.

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 235-250.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13*, 348-353.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321-335.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin and Review, 4*, 460-493.

Fawcett, A. J., & Nicolson, R. I. (1992). Automatisation deficits in balance for dyslexic children. *Perceptual and Motor Skills, 75*, 507-529.

Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Portsmouth, NH: Heinemann.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458-467.

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society, 15*, 501-508.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*, 3-17.

Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PLoS ONE, 8*, e52500.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science, 12*, 13-18.

Froyen, D., Willems, G., Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter-speech sound processing. *Developmental Science, 14*, 635-648.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception and Psychophysics, 54*, 287-295.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart, & J. Marshall (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301-330). London: Erlbaum.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Froyen, D., Willems, G., & Blomert, L. (2010). Evidence for a specific cross-modal association deficit in dyslexia: An electrophysiological study of letter-speech sound processing. *Developmental Science, 14*, 635-648.

Gabay, V., Schiff, R., & Vakil, E. (2012). Dissociation between online and offline learning in developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology, 34*, 1-10.

Gaillard, V., Destrebecqz, A., Michiels, S., & Cleeremans, A. (2009). Effects of age and practice in sequence learning: A graded account of ageing, learning, and control. *European Journal of Cognitive Psychology, 21*, 255-282.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 34-54.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review, 16*, 486-490.

Gentry, J. R. (1982). An analysis of developmental spelling in GNYS AT WRK. *The Reading Teacher, 36*, 192-200.

Glutting, J. J., Adams, W., & Sheslow, D. (2000). *Wide Range Intelligence Test Manual*. Wilmington, DE: Wide Range.

Gombert, J. E. (2003). Implicit and explicit learning to read: implication as for subtypes of dyslexia. *Current Psychology Letters, 10*.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gómez , R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. J*ournal of Experimental Psychology: Learning, Memory and Cognition, 20*, 396-410.

Goswami, U. (1988). Orthographic analogies and reading development. *The Quarterly Journal of Experimental Psychology, 40A*, 239-268.

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences, 7*, 534-540.

Grigorenko, E. L. (2001). Developmental dyslexia: An update on genes, brains, and environments. *Journal of Child Psychology and Psychiatry, 42*, 91-125.

Habib, M. (2000). The neurological basis of developmental dyslexia. An overview and working hypothesis. *Brain, 123*, 2373-2399.

Hasher, L., & Zacks, R., (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist, 39*, 1372-1388.

Hatcher, P., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41-57.

Hayes, N., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition, 28*, 249-276.

Hayes, H., Kessler, B., & Treiman, R. (2005). English spelling: Making sense of a seemingly chaotic writing system. *Perspectives: The International Dyslexia Association, Summer*, 8-10.

Hayes, H., Treiman, R., & Kessler, B. (2006). Children use vowels to help them spell consonants. *Journal of Experimental Child Psychology, 94*, 27-42.

Hazeltine, E., Grafton, S. T., & Ivry, R. B. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding a PET study. *Brain, 120*, 123-140.

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20*, 115-162.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*, 341-348.

Howard Jr., J. H., Howard, D. V., Japikse, K., Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia, 44*, 1131-1144.

Hsu, H. J., & Bishop, D. V. M. (2010). Grammatical difficulties in children with specific language impairment. Is learning deficient? *Human Development, 53*, 264-277.

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language, learning and cognition*. Oxford, England: Wiley-Blackwell.

International Phonetic Association. (1996). *Reproduction of the International Phonetic Alphabet*. Retrieved from http://www.langsci.ucl.ac.uk/ipa/

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press.

Irannejad, S., & Savage, R. (2012). Is a cerebellar deficit the underlying cause of reading disabilities? *Annals of Dyslexia, 62*, 22-52.

Ise, E., Arnoldi, C. J., Bartling, J., & Schulte-Körne, G. (2012). Implicit learning in children with spelling disability: Evidence from artificial grammar learning. *Journal of Neural Transmission, 119*, 999-1010.

Jaffré, J. P., & Fayol, M. (1997). *Orthographe: Des systèmes aux usages*. Paris: Flammarion.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46*, 723-750.

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

Johnson, E. K., & Jusczyk, P. W. (2001).Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548-567.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339-345.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 524–531.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*, 61-112.

Jonaitis, E. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33*, 951-968.

Judge, J., Caravolas, M., & Knox, P. C. (2006). Smooth pursuit eye movements and phonological processing in adults with dyslexia. *Cognitive Neuropsychology, 23*, 1174-1189.

Judge, J., Caravolas, M., & Knox, P. C. (2007). Visual attention in adults with developmental dyslexia: Evidence from manual reaction time and saccade latency. *Cognitive Neuropsychology, 24*, 260-278.

Jusczyk, P. W., Aslin, R. N., (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420.

Kaplan, B. J., Dewey, D., Crawford, S. G., & Wilson, B. N. (2001). The term comorbidity may be of questionable value in reference to developmental disorders: Data and theory. *Journal of Learning Disabilities, 34*, 555-565.

Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition, 32*, 956-964.

Karpicke, J. D., & Roediger, H. L. III. (2007). Expanding retrieval practice promotes short term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704-719.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 45-66). Amsterdam: Elsevier.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37*, 374-382.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*, 321-340.

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

Kemp, N., & Bryant, P. (2003). Do beez buzz? Rule-based and frequency-based knowledge in learning to spell plural-s. *Child Development, 74*, 63-74.

Kessler, B. (2009). Statistical learning of conditional orthographic correspondences. *Writing Systems Research, 1*, 19-34.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 44*, 592-617.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171-182.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it implicit and long-term? *Neuroscience Letters, 461*, 145-149.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models. *Psychological Research, 63*, 95-105.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Memory and Cognition, 28*, 1321-1332.

Kinder, A, & Lotz, A. (2009). Connectionist models of artificial grammar learning: what type of knowledge is acquired? *Psychological Research, 73*, 659-673.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, 35-42.

Klinger, L. G., Klinger, M. R., & Pohlig, R. (2007). Implicit learning impairments in autism spectrum disorders: Implications for treatment. In J. M. Pérez, P. M. González, M. L.

Comí, & C. Nieto (Eds.), *New developments in autism: The future is today* (pp. 76–103). London: Jessica Kingsley.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 79-91.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 169-181.

Landerl, K., & Moll, K. (2010). Comorbidity of specific learning disorders: Prevalence and familial transmission. *Journal of Child Psychology and Psychiatry, 51*, 287-294.

Lehtonen, A., & Bryant, P. (2005). Doublet challenge: form comes before function in children's understanding of their orthography. *Developmental Science, 8*, 211-217.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45*, 764-781.

Lewicki, P., & Hill, T. (1989). On the status of nonconscious processes in human cognition: Comment of Reber. *Journal of Experimental Psychology: General, 118*, 239-241.

Lorch, R. R, & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 149-157.

Lotz, A., Kinder, A., Lachnit, H. (2009). Multiple regression analyses in artificial grammar learning: The importance of control groups. *The Quarterly Journal of Experimental Psychology, 62*, 576-584.

Lovegrove, W. J., Bowling, A., Badcock, B., Blackwood, M. (1980). Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science, 210*, 439-440.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*, 185-199.

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford: Oxford University Press.

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science, 12*, 1020-1025.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science, 283*, 77-80.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1083-1100.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

Mattys, S., & Jusczyk, P. (2001). Do infants segment words or continuous recurring patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27*, 644-655.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar based classification in artificial grammar learning. *Memory & Cognition, 13*, 469-475.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 239-245.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffini, M.., & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. *Neuropsychologia, 48*, 863-872.

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

Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., et al. (2008). Structural correlates of implicit learning deficits in subjects with developmental dyslexia. *Annals of the New York Academy of Sciences, 1145*, 212-221.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1007-1028.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176-185.

Mintz, T. H., Newport, E. L. & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393-424.

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

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*, 138-153.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665-681.

Nation, K. (1997). Children's sensitivity to rime unit frequency when spelling words and nonwords. *Reading and Writing: An Interdisciplinary Journal, 9*, 321-338.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Nicolson R. I., & Fawcett, A. J. (1990). Automaticity: a new framework for dyslexia research? *Cognition, 35*, 159-182.

Nicolson, R. I., & Fawcett, A. J. (1994). Comparison of deficits in cognitive and motor skills among children with dyslexia. *Annals of Dyslexia, 44*, 147-164.

Nicolson, R.I., & Fawcett, A.J. (1995). Dyslexia is more than a phonological disability. *Dyslexia, 1*, 37-45.

Nicolson, R. I., Fawcett, A. J., Berry, E. L., Jenkins, I. H., Dean, P., & Brooks, D. J. (1999). Association of abnormal cerebellar activation with motor learning difficulties in dyslexic adults. *Lancet, 353*, 1662-1667.

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences, 24*, 508-511.

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

Nunes, T., & Bryant, P. (2009). *Children's reading and spelling: Beyond the first steps*. Oxford: Wiley-Blackwell.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13-23.

Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from English and French. *Cognitive Development, 23*, 339-359.

Pacton, S., Fayol, M., & Perruchet, P. (2002). The acquisition of untaught orthographic regularities in French. In L. Verhoeven, C. Erlbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 121-137). Dordrecht: Kluwer.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development, 76*, 324-339.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 80-96.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130*, 401-426.

Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings? *Frontiers in Psychology, 4*, 1-10.

Pavlidou, E. V., Kelly, L. M, & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia, 16*, 143-161.

Pavlidou, E. V., Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of Dyslexia, 59*, 55-77.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*, 674-685.

Pennington, B. F., Van Orden, G., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development, 61*, 1753-1778.

Pennington, B. F., Willcutt, E., & Rhee, S. H. (2005). Analyzing comorbidity. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 33, pp. 263-304). Oxford: Elsevier.

Perruchet, P. (1994). Defining the units of a synthetic language: Comment on Vokey and Brooks (1992). J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 223-228.

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (Vol. 2: Cognitive psychology of memory, pp. 597-621). Oxford: Elsevier.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General, 119*, 264-275.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233-238.

Pollo, T. C., Treiman, R., & Kessler, B. (2007). Three perspectives on spelling development. In E. J. Grigorenko & A. Naples (Eds.), *Single word reading: Cognitive, Behavioral, and Biological Perspectives*. Mahwah, NJ: Erlbaum.

Pothos, E. M., & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 847-862.

Pothos, E. M., Chater, N., & Ziori, E. (2006). Does stimulus appearance affect learning? *American Journal of Psychology, 119*, 277-301.

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

Rack, J. E., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29-53.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology, 13*, 212-218.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126*, 841-865.

Rae, C., Lee, M. A., Dixon, R. M., Blamire, A. M., Thompson, C. H., Styles, P., Talcott, J., Richardson, A. J., & Stein, J. F. (1998). Metabolic abnormalities in developmental dyslexia detected by 1H magnetic resonance spectroscopy. *Lancet, 351*, 1849-1852.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics, 27*, 537-544.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Reber, A. S. (1976). Implicit learing of synthetic languages: The role of the instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 88-94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition, 6*, 189-221.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human, Learning and Memory, 6*, 492-502.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888–896.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56(A)*, 97-115.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A re-evaluation. *Journal of Experimental Psychology: General, 125*, 123-137.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 425-469.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235-268.

Rochelle, K. S. H., & Talcott, J. B. (2006). Impaired balance in developmental dyslexia? A meta-analysis of the contending evidence. *Journal of Child Psychology and Psychiatry, 47*, 1159-1166.

Romberg, A. R. & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 906-914.

Roodenrys, S., & Dunn, N. (2008). Unimpaired implicit learning in children with developmental dyslexia. *Dyslexia, 14*, 1-15.

Rose, J. (2009). Identifying and teaching children and young people with dyslexia and literacy difficulties. DCSF- 00659-2009. London: DCSF Publications. Retrieved from http://publications.dcsf.gov.uk/eOrderingDownload/00659-2009DOM-EN.pdf

Rüsseler, J., Gerth, I., & Münte, T. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology, 28*, 808-827.

Saffran, J. R. (2003). Statistical language learnings: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101-195.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy, 4*, 273-284.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh, PA: Psychology Software Tools.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367-447.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 50*, 216-252.

Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K., et al. (1999). Persistence of dyslexia: The Connecticut longitudinal study at adolescence. *Pediatrics, 104*, 1351-1359.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*, 163-163.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science, 298*, 553-554.

Seidenberg, M., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*, 434-435.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior, 8*, 248-251.

Smith, A. (1991) *Symbol Digit Modalities Test*. Los Angeles: Western Psychological Services.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Application to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Snowling, M. (2000). *Dyslexia* (2nd ed.). Oxford: Blackwell.

Snowling, M., Hulme, C., & Nation, K. (1997). A connectionist perspective on the development of reading skills in children. *Trends in Cognitive Sciences, 1*, 88-91.

Sperling, A. J., Lu, Z. L., & Manis, F. R. (2004). Slower implicit categorical learning in adult poor readers. *Annals of Dyslexia, 54*, 281-303.

Stanback, M. L. (1992). Syllable and rhyme patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42*, 196-221.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and garden-variety poor reader: The phonological-core variable-model. *Journal of Disabilities, 21*, 590-612.

Steffler, D. J. (2001). Implicit cognition and spelling development. *Developmental Review, 21*, 168-204.

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences, 20*, 147-152.

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.

Stoodley, C. J., Harrison, E. P. D., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia, 44*, 795-798.

Stoodley, C. J., Ray, N. J., Jack, A., & Stein, J. F. (2008). Implicit learning in control, dyslexic, and garden-variety poor readers. *Annals of the New York Academy of Sciences, 1145*, 173-183.

Tagarelli, K., Borges Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2061–2066). Austin, TX: Cognitive Science Society.

Tallal, P. (1980). Auditory temporal perception, phonics and reading disabilities in children. *Brain and Language, 9*, 182-198.

Tanaka, D., Kiyokawa, S., Yamada, A., Dienes, Z., Shigemasu, K. (2008). Role of selective attention in artificial grammar learning. *Psychonomic Bulletin & Review, 15*, 1154-1159.

Taylor, C. F., & Houghton, G. (2005). Learning artificial phonotactic constraints: Time course, durability, and relationship to natural constraints. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 31*, 1398-1416.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 60-76.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706-716.

Tillmann, B., & McAdams, S. (2004) Implicit Learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1131-1142.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency (TOWRE)*. Austin, TX: Pro-Ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in children with phonological processing difficulties: group and individual responses to instruction. *Journal of Educational Psychology, 81*, 579-593.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.

Treiman, R. (2004). Phonology and spelling. In P. Bryant & T. Nunes (Eds.), *Handbook of children's literacy* (pp. 31-42). Dordrecht, the Netherlands: Kluwer.

Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders, 20*, 1-18.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98*, 642-652.

Treiman, R., & Kessler, B. (2013). Learning to use an alphabetic writing system. *Language Learning and Development, 9*, 317-330.

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47*, 448-468.

Treiman, R., Kessler, B., & Bourassa, D. (2001). Children's own names influence their spelling. *Applied Psycholinguistics, 22*, 555-570.

Tremblay, S., Macken, W. J., & Jones, D. M. (2000). Elimination of the word length effect by irrelevant sound revisited. *Memory & Cognition, 28*, 841-846.

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

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934-1945.

Valdois, S., Bosse, M.-L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*, 339-363.

Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing: An Interdisciplinary Journal, 16*, 541-572.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery, and Psychiatry, 76*, 1392.

Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia, 41*, 108-114.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 328-344.

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. Jo*urnal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1504-1510.

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., et al. (2003). Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits? *Journal of Experimental Child Psychology, 84*, 338-354.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 387-398.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1289-1295.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale - Revised: Manual*. New York: Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children – Third Edition UK*. London: Psychological Corporation.

Wechsler, D. (1998). *Wechsler Adult Intelligence Scale – Third Edition*. London: The Psychological Corporation.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122*, 227-248.

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 181-200.

Wilkinson, G. S., & Robertson, G. J. (2006). *The Wide Range Achievement Test–4*. Lutz, FL: Psychological Assessment Resources.

Wimmer, H., Mayringer, H., & Landerl, K. (1998). Poor reading: A deficit in skill automatization or a phonological deficit? *Scientific Studies of Reading, 2*, 321-340.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-revised (WJ-r)*. Allen, TX: DLM Teaching Resources.

Wright, R. L., & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: Active adaptation and selective processing in acquisition and application. *Memory & Cognition, 26*, 402-420.

Appendix A: Stimuli used in the Latin-letter positional constraints variants of the incidental graphotactic learning task (Chapters 2 and 3).

List 1	List 2	List 3	List 4
2-vowel version			
dot	det	tod	ted
den	don	tom	tem
dop	dep	tel	tol
des	dos	tef	tof
met	mot	ned	nod
mon	men	nem	nom
mep	mop	nol	nel
mos	mes	nof	nef
lot	let	pod	ped
len	lon	pom	pem
lop	lep	pel	pol
les	los	pef	pof
fet	fot	sed	sod
fon	fen	sem	som
fep	fop	sol	sel
fos	fes	sof	sef
3-vowel version[a]			
dit	din	tid	tim
dip	dis	tif	til
min	mit	nim	nid
mis	mip	nil	nif
lit	lin	pid	pim
lip	lis	pif	pil
fin	fit	sim	sid
fis	fip	sil	sif

Note. [a]*n* = 24 items/list. Stimuli in the 3-vowel version consisted of the16 strings/list used in the 2-vowel version and an additional 8 C_iiC_j strings/list.
In one of the counterbalanced list conditions, den was presented during exposure (and served, in addition, as a legal seen test item in the study presented in Chapter 2), don as a legal unseen test item and tom as an illegal test item.

Appendix B: Stimuli used in the Latin-letter contextual constraints variants of the incidental graphotactic learning task (Chapter 2 and 3).

List 1	List 2	List 3	List 4
dot	don	det	den
dop	dos	dep	des
mon	mot	men	met
mos	mop	mes	mep
lot	lon	len	let
lop	los	les	lep
fon	fot	fet	fen
fos	fop	fep	fes
tem	ted	tod	tom
tef	tel	tol	tof
ned	nem	nom	nod
nel	nef	nof	nol
pem	ped	pom	pod
pef	pel	pof	pol
sed	sem	sod	som
sel	sef	sol	sof

Note. In one of the counterbalanced list conditions, dop was presented during exposure (and served, in addition, as a legal seen test item in the study presented in Chapter 2), dos as a legal unseen test item and dep as an illegal test item.

Appendix C: Stimuli used in the Greek-letter variants of the incidental graphotactic learning task (Chapter 2).

Table C1. Stimuli used for positional constraints learning

List 1	List 2	List 3
θαξ	θηξ	ξαθ
θηφ	θαφ	ξαδ
θαγ	θηγ	ξητ
θηλ	θαλ	ξηπ
δηξ	δαξ	φηθ
δαφ	δηφ	φηδ
δηγ	δαγ	φατ
δαλ	δηλ	φαπ
ταξ	τηξ	γαθ
τηφ	ταφ	γαδ
ταγ	τηγ	γητ
τηλ	ταλ	γηπ
πηξ	παξ	ληθ
παφ	πηφ	ληδ
πηγ	παγ	λατ
παλ	πηλ	λαπ

Note. In one of the counterbalanced list conditions, θηφ was presented as an exposure and legal seen test item, θαφ as a legal unseen test item and ξαδ as an illegal test item.

Table C2. Stimuli used for contextual constraints learning

List 1	List 2	List 3
θαξ	θαφ	θηξ
θαγ	θαλ	θηγ
δαφ	δαξ	δηφ
δαλ	δαγ	δηλ
ταξ	ταφ	τηφ
ταγ	ταλ	τηλ
παφ	παξ	πηξ
παλ	παγ	πηγ
ξηδ	ξηθ	ξαθ
ξηπ	ξητ	ξατ
φηθ	φηδ	φαδ
φητ	φηπ	φαπ
γηδ	γηθ	γαδ
γηπ	γητ	γαπ
ληθ	ληδ	λαθ
λητ	ληπ	λατ

Note. In one of the counterbalanced list conditions, θαγ was presented as an exposure and legal seen test item, θαλ as a legal unseen test item and θηγ as an illegal test item.

Appendix D: Stimuli used in the exception word reading task.

Low frequency (friends<enemies)	Low frequency (friends>enemies)	High frequency (friends<enemies)	High frequency (friends>enemies)
blown	brook	bear	both
breast	bull	blood	bread
bush	calf	break	breath[a]
cough	dread	broad	child
crow	fold	foot	chose
deaf	ghost	give	come
dough	grind	gone	does
hearth	halt	gross	friend
mould	hood	lose	front
pear	hose	move	heard
pint	mall	push	kind
plaid	mild	put	none
rouse	poll	rough	please
sew	prey	says	post
shove	soup	show	pull
soot	stalk	son	roll
sown	stealth	though	stood
steak	tease	were	tall
ton	tread	whom	took
wool	yearn	youth	touch

[a]Item excluded from the analyses

Appendix E. Stimuli used in the rapid automatized naming digits task (block 1).

7	**2**	**9**	**3**	**6**	**9**	**2**	**6**
9	**3**	**7**	**6**	**3**	**7**	**3**	**2**
7	**9**	**2**	**6**	**2**	**3**	**6**	**7**
9	**6**	**3**	**7**	**9**	**2**	**3**	**9**
7	**2**	**6**	**2**	**7**	**3**	**9**	**6**

Appendix F. Stimuli used in the rapid automatized naming objects task (block 1).

Appendix G: Stimuli used in the memorization phase of the artificial grammar learning tasks.

Table G1. List 1 Items

FSGFBNB	
FTGMBQV	
FTVRCNV	
FXDRZQB	
JPDRCXJ	
JPKMBWJ	
JSGMTXH	
JXDHCWH	
MPDFBQL	
MPVRTXL	
RTGHCXZ	
RTKMZQZ	

Table G2. List 2 Items

FSGMBWH	
FTKMBQL	
FTVRTXJ	
FXDHCXH	
JPKMZQV	
JPVRCXZ	
JSGFBQB	
JXDRCNB	
MPDFBNV	
MPDRZQZ	
RTGHCWJ	
RTGMTXL	

Appendix H: Stimuli used in the test phase of the artificial grammar learning tasks.

Table H1. List 1 Items

FXDHFML

FXLKCXJ

FSXLBWJ

FTVRTKX

RTVNLXJ

MBXMZQV

JKFMZQB

JSGFGDZ

MPVLPQZ

BQDRZQV

HLKMTXL

RTGMTVS

FXDHZNL

FXVFCXJ

FSKRBWJ

FTVRTWZ

RTVFZXJ 

MTDMZQV 

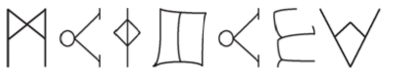
JTDMZQB 

JSGFZNZ 

MPVFCQZ 

JTDRZQV 

MXKMTXL 

RTGMTQJ 

FXDHCXL 

FXDHCXJ 

FSGMBWJ 

FTVRTXH 

RTVRTXJ 

MPKMZQV 

JPKMZQB 

JSGFBQZ 

MPVRZQZ 

MPDRZQV 

RTKMTXL 

RTGMTXJ 

FXDRZQL 

FXDRZQV 

JSGMTXJ 

FTVRCNB 

RTKMZQB 

MPVRTXH 

JPKMBWH 

JSGMTXZ 

MPKMBWJ 

MPDRTXL 

RTVRCNV 

RTGMZQZ 

Table H2. List 2 Items

FXMSZQL	
FXDRXCV	
JSPGTXJ	
FRFRCNB	
RTKZJQB	
MPVRTDM	
LBKMBWH	
JSGSQXZ	
DCKMBWJ	
MPDRTFX	
RWQMZQZ	
RTVRFKV	
FXKFZQL	
FXDRBXV	
JSKHTXJ	
FPGRCNB	
RTKFCQB	

MPVRTQH 

MXKMBWH 

JSGRBXZ 

MXKMBWJ 

MPDRTWL 

RSVMZQZ 

RTVRBXV 

FXDRZQL 

FXDRZQV 

JSGMTXJ 

FTVRCNB 

RTKMZQB 

MPVRTXH 

JPKMBWH 

JSGMTXZ 

MPKMBWJ 

MPDRTXL 

RTVRCNV 

RTGMZQZ	
FXDHCXL	
FXDHCXJ	
FSGMBWJ	
FTVRTXH	
RTVRTXJ	
MPKMZQV	
JPKMZQB	
JSGFBQZ	
MPVRZQZ	
MPDRZQV	
RTKMTXL	
RTGMTXJ	

Appendix I. Inter correlations between skilled readers' performance on the cognitive and literacy measures (Chapter 4).

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. WRIT Full-scale IQ[a]	—	.58**	.40	.69**	.77**	.50*	.58**	.31	.21	.26	.32	.05	-.21	-.13	.36
2. WRIT Verbal IQ[a]		—	.67**	.88**	-.07	-.18	.12	.20	.45*	-.07	.12	.14	-.05	-.04	.13
3. WRIT Verbal Analogies[b]			—	.58**	-.04	-.25	.24	.21	.34	.04	-.11	.16	.17	.02	.12
4. WRIT Vocabulary[b]				—	.16	-.02	.27	.29	.38	-.11	.23	.14	-.13	.01	.13
5. WRIT Nonverbal IQ[a]					—	.75**	.63**	.22	-.09	.36	.30	-.06	-.23	-.13	.34
6. WRIT Matrices[b]						—	-.04	-.01	-.11	.31	.38	.08	-.22	-.03	.22
7. WRIT Diamonds[b]							—	.35	.03	.19	.02	-.17	-.09	-.18	.30
8. WRAT Reading[b]								—	.64**	-.14	.26	.22	-.32	-.18	-.18
9. WRAT Spelling[b]									—	-.31	.25	.45*	-.13	-.18	-.30
10. WAIS Digit Span[b]										—	-.11	.13	.11	-.21	.47*
11. WAIS Symbol Search[b]											—	.22	-.51*	-.64**	-.35
12. SDMT[b]												—	-.14	-.40	-.24
13. RAN digits mean time[c]													—	.55**	.25
14. RAN objects mean time[c]														—	.20
15. NWPD latencies[d]															—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; SDMT = Symbol Digit Modalities Test; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Standard Scores. [b]Raw scores. [c]In seconds. [d]Log transformed.
* $p < .05$. ** $p < .01$.

Appendix J. Inter correlations between dyslexic readers' performance on the cognitive and literacy measures (Chapter 4).

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. WRIT Full-scale IQ[a]	—	.79**	.75**	.52**	.81**	.75**	.46*	.21	-.16	.42*	.32	.23	-.20	-.32	-.27
2. WRIT Verbal IQ[a]		—	.85**	.79**	.29	.46*	.85**	.38	.21	.43*	.07	-.09	-.06	-.26	-.20
3. WRIT Verbal Analogies[b]			—	.52**	.38	.36	.21	.31	-.01	.34	.04	-.14	-.04	-.20	-.20
4. WRIT Vocabulary[b]				—	.05	.30	-.20	.51**	.42*	.29	.05	-.03	.01	-.20	-.18
5. WRIT Nonverbal IQ[a]					—	.75**	.74**	-.05	-.46*	.27	.42*	.44*	-.27	-.26	-.22
6. WRIT Matrices[b]						—	.11	.01	-.21	.00	.24	.23	-.06	-.27	-.13
7. WRIT Diamonds[b]							—	-.08	-.47*	.42*	.38	.41*	-.35	-.12	-.19
8. WRAT Reading[b]								—	.56**	.24	-.11	-.26	.16	.10	-.50*
9. WRAT Spelling[b]									—	.01	-.09	-.20	.06	.08	-.28
10. WAIS Digit span[b]										—	.31	.11	-.32	-.26	-.27
11. WAIS Symbol Search[b]											—	.75**	-.68**	-.71**	-.19
12. SDMT[c]												—	-.71**	-.60**	-.19
13. RAN digits mean time[d]													—	.86**	.24
14. RAN objects mean time[d]														—	.12
15. NWPD latencies[e]															—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; SDMT = Symbol Digit Modalities Test; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Standard Scores. [b]Raw scores. [c]Square root transformed. [d]Log transformed. [e]In seconds.
* $p < .05$. ** $p < .01$.

Appendix K. Inter correlations between skilled readers' performance on the cognitive and literacy measures (Chapter 5: experiments 4.1 – 4.2).

Variable	1	2	3	4	5	6	7	8	9
1. WRIT Matrices[a]	—	.51**	.35	.26	-.03	.14	-.18	.11	-.02
2. WRIT Vocabulary[a]		—	.45*	.14	.16	.28	-.17	-.07	-.14
3. WRAT Reading[a]			—	.71**	.40*	.12	-.35	-.17	-.22
4. WRAT Spelling[a]				—	.29	.13	-.36	-.26	-.44*
5. WAIS Digit Span[a]					—	.15	-.21	.07	-.32
6. WAIS Symbol Search[a]						—	- .35	-.49**	-.24
7. RAN digits mean time[b]							—	.44*	.28
8. RAN objects mean time[b]								—	.33
9. NWPD latencies[c]									—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Raw scores. [b]In seconds. [c]Log transformed
* $p < .05$. ** $p < .01$.

Appendix L. Inter correlations between dyslexic readers' performance on the cognitive and literacy measures (Chapter 5: experiments 4.1 – 4.2).

Variable	1	2	3	4	5	6	7	8	9
1. WRIT Matrices[a]	—	.33	-.19	-.12	-.03	.41	-.06	-.02	.43
2. WRIT Vocabulary[a]		—	.14	.07	.20	.36	.07	.01	-.06
3. WRAT Reading[a]			—	.82**	.27	.35	-.13	.04	-.59**
4. WRAT Spelling[a]				—	.11	.33	-.16	.07	-.75**
5. WAIS Digit Span[b]					—	.27	-.16	-.02	-.10
6. WAIS Symbol Search[a]						—	-.58**	-.51*	-.09
7. RAN digits mean time[c]							—	.82**	.13
8. RAN objects mean time[c]								—	.02
9. NWPD latencies[c]									—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Raw scores. [b]Log transformed. [c]In seconds
* $p < .05$. ** $p < .01$.

Appendix M. Inter correlations between skilled readers' performance on the cognitive and literacy measures (Chapter 5: experiments 4.3 – 4.4).

Variable	1	2	3	4	5	6	7	8	9
1. WRIT Matrices[a]	—	.30	.16	.30	.33	.36	-.20	-.11	-.18
2. WRIT Vocabulary[a]		—	.39*	.48**	.08	-.16	.15	.10	.02
3. WRAT Reading[a]			—	.58**	.21	.09	-.36	-.05	-.46*
4. WRAT Spelling[a]				—	.26	-.18	-.08	.25	-.35
5. WAIS Digit Span[b]					—	.18	-.48*	.10	-.31
6. WAIS Symbol Search[a]						—	-.37	-.39*	-.38*
7. RAN digits mean time[c]							—	.33	.35
8. RAN objects mean time[c]								—	-.12
9. NWPD latencies[c]									—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Raw scores. [b]Log transformed. [c]In seconds.
* $p < .05$. ** $p < .01$.

Appendix N. Inter correlations between dyslexic readers' performance on the cognitive and literacy measures (Chapter 5: experiments 4.3 – 4.4).

Variable	1	2	3	4	5	6	7	8	9
1. WRIT Matrices[a]	—	.23	.18	-.33	-.09	.15	-.01	.25	-.22
2. WRIT Vocabulary[a]		—	.47*	.10	.24	-.03	-.26	-.05	.02
3. WRAT Reading[a]			—	.31	-.13	-.24	.28	.41	-.09
4. WRAT Spelling[a]				—	.43	-.14	-.16	-.36	-.25
5. WAIS Digit Span[a]					—	.34	-.49*	-.53*	-.22
6. WAIS Symbol Search[a]						—	-.58**	-.57**	-.58**
7. RAN digits mean time[b]							—	.80**	.36
8. RAN objects mean time[c]								—	.32
9. NWPD latencies[c]									—

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = Nonword Phoneme Deletion.
[a]Raw scores. [b]Log transformed. [c]In seconds.
* $p < .05$. ** $p < .01$.