Procedural and Declarative Memory and Language Ability in Children

Gillian West

University College London

A thesis submitted for the

degree of Doctor of Philosophy (PhD)

November 2017

Declaration

I, Gillian West, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Impaired procedural learning has been suggested as a possible cause of developmental language disorder and dyslexia (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005). However, studies investigating this hypothesis have so far delivered inconsistent results. These studies typically use extreme group designs, frequently with small sample sizes and measures of procedural learning with unreported reliability.

This thesis first used a meta-analysis to examine the existing evidence for a procedural deficit in language disorders. The experimental studies then took a different approach to previous studies, using a concurrent correlational design to test large samples of children unselected for ability on a wide range of implicit (serial reaction time, Hebb serial learning, contextual cueing and probabilistic category learning) and declarative learning tasks and literacy, language and arithmetic attainment measures. The reliability of the tasks was also carefully assessed. A final study explored the hypothesis from an extreme group design perspective, comparing a typically developing sample with a group of dyslexic children matched for reading ability. None of the studies found evidence of a relationship between procedural learning and language-related abilities. By contrast, a relationship between verbal declarative learning and attainment was found replicating earlier studies. Crucially, the first large-scale study showed that procedural learning tasks of a similar length to those typically used in earlier studies had unacceptably low reliability and correlated poorly with each other and with attainment. The second large-scale study, used extended procedural learning tasks that had proved reliable in adults, but found similar low levels of reliability in children. Additionally, the level of attention children paid during these extended tasks accounted entirely for the relationship between procedural learning and attainment.

The results in this thesis highlight the importance of establishing task reliability, as well as considering the potential effects of individual differences in basic cognitive processes such as attention in all investigations of procedural learning.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Charles Hulme for such tremendous guidance and support throughout the last three years, for shaping this thesis, alongside my development as a researcher, and for making my time at UCL such an enriching experience. Thank you also to my secondary supervisor Professor David Shanks for his support and insightful comments on papers and on the draft of this thesis. I would also like to acknowledge my gratitude to the ESRC for the funding I received that enabled me to undertake this PhD.

I want to thank several people who have been integral to my success during my time at UCL. Thank you to Miguel Vadillo for his programming skill and support on the contextual cueing and weather prediction tasks. His enthusiasm for cognitive psychology and the investigation of procedural learning has been contagious. Thank you to Mike Coleman for the expert programming of the serial reaction time task that forms the backbone of this thesis. I have also been lucky to be part of a very special and supportive community of PhD students, postdocs and lecturers at UCL and I thank them all, but I am particularly grateful to Francina Clayton who was unfailingly helpful and full of good advice. My gratitude also goes to the head teachers and teachers who welcomed me into their schools and the many wonderful, interested and interesting children who took part in the research.

Last, but by no means least, I would like to thank my lovely family for giving me the courage to start this PhD and see it through to the end. Your confidence in me has made all the difference. I thank my parents and brother for their unwavering and unquestioning love and support, always. I thank my husband, James, for his love and patience and for putting up with so many lost evenings and weekends. He has been my partner in this adventure, as in all things. I thank my children, Alice and Casper, for surviving three years of benign neglect with such good humour. You are still and always the best things I have ever done. I will pay more attention to you now, I promise.

For my family

Table of Contents

11

List of Tables

List of Figures

Chapter 1   Developmental disorders of language

Developmental disorders of language involve problems in acquiring spoken and written language. The focus in this thesis will be on two of these disorders: developmental language disorder (DLD) and developmental dyslexia (DD).

1.1   Developmental language disorder

Developmental language disorder is a disorder affecting the development of oral language, in spite of normal non-linguistic development (Bishop, 2006). Until recently the disorder was termed specific language impairment (SLI). The word "specific" in the title was justified on the basis of the co-occurrence of language impairment alongside unimpaired non-verbal intelligence (Mareschal, Butterworth, & Tolmie, 2013) or where language impairments were disproportionately greater than any impairments in non-linguistic domains (Webster & Shevell, 2004). However, the requirement for a definition based on a discrepancy diagnosis has now been dropped and the term developmental language disorder recommended in its place for all cases where language disorder presents without a known biomedical aetiology (Bishop, Snowling, Thompson, & Greenhalgh, 2016b). Diagnosis of developmental language disorder no longer requires non-verbal ability to outstrip verbal ability, not least because this excluded many children with other co-occuring difficulties from diagnosis and, therefore, from much–needed clinical support. The term developmental language disorder will be used throughout this thesis, even where cited authors have referred to SLI.

Developmental language disorder is estimated to affect around 3% - 7% of the population (Norbury et al., 2016; Tomblin et al., 1997). Language development in children with developmental language disorder is delayed, with children using and combining words later and plateauing earlier than typically developing children (Leonard, 2014). Children with developmental language disorder do not follow a typical profile of language development. They have particular problems with morphology and syntax (Tager-Flusberg & Cooper, 1999). In particular, they show impairments in learning language structure, for example, in marking tenses and with

grammatical inflections. They also have difficulty with phonology, finding it more difficult than typically developing children to both process and produce word sounds correctly (Bishop, 1994; Webster & Shevell, 2004). They also have impaired verbal working and short-term memory (Archibald & Gathercole, 2006; Weismer, Evans, & Hesketh, 1999), as evidenced by poorer non-word repetition and word recall performance.

Developmental language disorder can persist into adolescence (Clark et al., 2007; Conti-Ramsden & Durkin, 2007; Stothard, Snowling, Bishop, Chipchase, & Kaplan., 1998) and even adulthood (Clegg, Hollis, Mawhood, & Rutter, 2005; Snowling, Bishop, & Stothard, 2000; Whitehouse, Line, Watt, & Bishop, 2009), with ongoing difficulties with grammar, phonological awareness, speech production and verbal short-term memory, which frequently extend into problems with literacy too (Young et al., 2002). The outlook for children classed as having persistent developmental language disorder is poor. In many children, however, developmental language disorder can appear short-lived. In a longitudinal study of 87 children diagnosed with language difficulties at the age of four, Stothard et al. (1998) reported that 44% of the children who had received a diagnosis of developmental language disorder no longer qualified for it 18 months later. It should be noted that only 11% of children with language difficulties alongside low non-verbal IQ had shown similar levels of improvement. A different picture emerged in adolescence, however. Although spoken language improved, at age 15 these same children showed verbal short-term memory and phonological impairments and many were experiencing academic difficulties, leading to the conclusion that "resolving" language difficulties in developmental language disorder may be an illusion for many (Duinmeijer, 2013; Snowling, Adams, Bishop, & Stothard, 2001).

1.2   Developmental dyslexia

Developmental dyslexia (DD) is a disorder characterised by impaired printed word recognition and spelling in spite of normal IQ and educational opportunities (Bishop & Snowling, 2004). It is associated primarily with difficulties in phonological processing, most commonly with deficits in phonological awareness and decoding,

22

impairing the ability to access and manipulate the sounds of speech (Bradley & Bryant, 1983; Snowling, 2013; Vellutino, Fletcher, Snowling, & Scanlon, 2004). This core deficit is frequently compounded by others, such as slow lexical retrieval and poor verbal short-term memory and these factors collectively affect the development of reading fluency and spelling (Ramus, 2004). Dyslexia is characterized by a difficulty in acquiring letter-sound knowledge, and learning to read words accurately and fluently. Difficulty with spelling are typically more severe and more persistent than difficulties with word reading. (Treiman, 1985). Developmental dyslexia is estimated to affect between 3% to 7% of the population, but estimates vary depending on the stringency of the diagnostic criteria (Barbiero et al., 2012; Snowling, 2013). Proportionately more boys than girls receive a diagnosis of dyslexia - from 1.5:1 to 3.1:1, according to Rutter et al. (2004). However, as many as 6 boys are referred for clinical diagnosis for every one girl, possibly as a result of increased rates of comorbidity with other disorders, such as attention deficit hyperactivity disorder (ADHD), in boys (Willcutt & Pennington, 2000).

The traditional IQ discrepancy definition of dyslexia is based on the idea that low IQ sets a cap on achievement as a result of general learning difficulties, while the deficits seen in dyslexia are caused by something else. For example, genes are thought to contribute more to high than low IQ dyslexia (Wadsworth, Olson, & DeFries, 2010). The utility of an IQ discrepancy definition is that by excluding those whose language and / or literacy impairments are due to more general learning difficulties, it is easier to isolate and investigate the cognitive deficits specifically implicated in the disorders. However, just as with developmental language disorder, discrepancy definitions have proved highly controversial, particularly in the light of evidence that individual differences in response to remedial instruction are not related to IQ (Jiménez, Rodríguez, & Ramírez 2009; Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009).

1.3   Developmental language disorder & dyslexia are distinct disorders

Research to date suggests that developmental language disorder and dyslexia are distinct developmental disorders, but with considerable overlap and frequent comorbidity (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005;

Krishnan, Watkins, & Bishop, 2016). Both reading and writing are scaffolded on oral language (Fletcher, 2009) and early problems with oral language place children at greater risk of problems with literacy later on.

Many older children receiving a diagnosis of reading disability qualify for a diagnosis of developmental language disorder (McArthur, Hogben, Edwards, Heath, & Mengler, 2000), just as many children receiving a diagnosis of developmental language disorder go on to experience difficulties with reading. For example, Ramus, Marshall, Rosen, & van der Lely (2013) found circa 50% of their sample of language-disordered children qualified for a diagnosis of both developmental language disorder and dyslexia. Similarly, Snowling et al. (2000) found 43% of children with earlier developmental language disorder had a reading disability aged fifteen. Snowling, Duff, Nash, and Hulme (2016) classed children diagnosed with language impairment at age three into two groups dependent on their subsequent developmental trajectory. Children with language impairment that had resolved by the start of literacy instruction had similar literacy outcomes to typically developing children. Children with persisting language impairment were likely to experience reading difficulties. A third group of children was also identified, with late-emerging language impairment. These children also had poor literacy outcomes and were likely to have a family history of dyslexia. The study concluded that children who have language impairments at school entry are likely to go on to have reading difficulties.

Although the above may suggest that developmental language disorder and dyslexia be conceived of as different points on the same developmental continuum, Bishop and Snowling (2004) highlight clear differences between the disorders. Primarily, children with developmental language disorder have non-phonological difficulties with semantics and syntax alongside their phonological difficulties, while the impairments of children with dyslexia are often confined to problems with phonological processing. In addition Ramus et al. (2013) suggested that a distinction be made between phonological representations, which are more predominantly impaired in developmental language disorder and phonological skills, such as verbal short-term memory or retrieval skills, which are more impaired in dyslexia.

24

1.4 Both disorders are dimensional

The impairments seen in both developmental language disorder and dyslexia are dimensional in nature (Fletcher, 2009; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). People with the disorders represent the tail end of a normal distribution in spoken language or reading ability, rather than a distinct category. As such, research into developmental language disorder and dyslexia and research on individual differences in language and reading development can be seen as two sides of the same coin.

1.5 Heterogeneity

The profiles of impairment in both disorders show substantial heterogeneity (developmental language disorder: Webster & Shevell, 2004; dyslexia: Petersen & Pennington, 2015). This may reflect variations in the distribution of subtle and multi-factorial cortical abnormalities that underlie the disorders (Ramus, 2014). To complicate matters phenotypic heterogeneity at a single time point is not a simple reflection of biological heterogeneity within the disorders. For example, one study classified over 2000 children with developmental language disorder into five groups based on language disability profiles (Conti-Ramsden & Botting, 1999), but a year later only 55% still fitted the same profiles. In dyslexia, attempts have been made to classify dyslexics into sub-groups based on reading profiles (Castles & Coltheart, 1993), but this has been criticized as a poor description of the dyslexic population as a whole (Griffiths & Snowling, 2002). Even those with similar profiles of reading impairment do not exhibit a homogenous profile of cognitive deficits (Zoubrinetsky, Bielle, & Valdois, 2014).

1.6 Comorbidity with other developmental disorders

Developmental language disorder and dyslexia often co-occur with other developmental disorders including ADHD (McGrath et al., 2011; Pennington et al., 2006) and dyspraxia (Rochelle & Talcott, 2006). A study by Dyck, Piek, and Patrick (2011) administered a battery of language and other cognitive tests to a mixed group of typically developing children and those with developmental coordination disorder,

language disorder, autism, ADHD and mental retardation, subjecting test scores to latent class analysis. Only two classes were identified from test scores: typically developing and disordered, with 50% of children with developmental coordination disorder and 20% of language-disordered children classed with the typically developing group. The authors concluded that boundaries between disorders and between disorder and normality are far from clear-cut. These findings echo those of Kaplan, Dewey, Crawford, and Wilson (2001) who found that 50% of a large sample of children referred for learning or attention disorders qualified for two or more diagnoses, with children that qualified for a diagnosis of ADHD having an 80.4% probability of meeting the criteria for at least one other developmental disorder. Grouping children into fine-grained, distinct categories of developmental disorder depends as much on the person making the diagnosis as on any impairment profile (Bishop, 2013). The same may be said of researchers grouping participants into experimental groups. Comorbidity complicates research into the underlying cause of developmental language disorder and dyslexia because measured deficits may not be reflective of a specific disorder.

How do we explain the comorbidity between different developmental disorders and the heterogeneity of symptoms seen in a given disorder? One suggestion is that having one condition could lead to an increased risk for others. For example, a child with developmental language disorder may go on to develop dyslexia, because development of efficient reading skill is hampered by deficits in oral language (Fletcher, 2009). Another possibility is that the different conditions may share a common cause (more of this later), with particular presentations dependent on differences in the extent and location of atypical brain development and subsequent connectivity (Powell & Bishop, 1992).

1.7   Genetic and environmental factors

Genetic risk factors play a role in the etiology of both dyslexia and developmental language disorder. There is a significantly higher rate of developmental language disorder in children with a first degree relative with language impairment compared to those with no family history of impairment (Stromswold, 1998). Twin studies have

26

also shown that genetic risk factors are important in developmental language disorder (Bishop 2002), with higher concordance rates in monozygotic twins (72%) than dizygotic twins (49%). However, the heterogeneity of symptoms and comorbidity issues makes identifying specific genetic risk factors for any disorder difficult.

Genetic risk factors appear to be important in the etiology of dyslexia. Twin studies provide evidence that phonological impairments in reading are particularly heritable (Castles, Datta, Gayan, & Olson, 1999; Olson et al., 2013; Hensler, Schatschneider, Taylor, & Wagner, 2010; Logan et al., 2013). The heritability of dyslexia has been estimated to be around 0.50 (Plomin et al., 2013), which is in line with other developmental cognitive disorders. A recent review by Peterson and Pennington (2015) summarised huge advances in the understanding of the genetic factors related to dyslexia at a molecular level. Linkage studies have identified risk loci and candidate genes for the disorder, several of which (DYX1C1, DCDC2, KIAA0319, and ROB01) affect neuronal migration during early brain development. KIAA0319 has been shown to lead to atypical auditory processing in rats (Szalkowski et al., 2013) and these findings are in line with biological evidence of the existence of ectopias in superficial cortical layers in the brains of dyslexics.

However, it appears that any developmental cognitive disorder will be influenced by multiple genes, working together and in synergy with multiple environmental risk factors. In this way gene environment interactions increase or decrease risk for developmental language disorder and dyslexia. For example, in dyslexia these may range from the genetic influence on parental reading skill and the number of books in the home and / or parental interest in fostering good reading habits to the amount of time children at genetic risk of dyslexia may spend with books. Dyslexic children read less than typically developing children and their lack of reading experience may have an ever-increasing negative influence on reading fluency and vocabulary development (Torgesen, 2005). In a similar vein, dyslexia has been found to be more heritable in families with higher levels of parental education (Friend, DeFries, & Olson, 2008). This is consistent with the idea that in supportive environments genetic influences will be the main reason for reading difficulties, while in families with lower parental

education environmental factors may play a greater role. The fact that relatives of dyslexics may show deficits on phonological processing tasks, yet possess reading ability within the normal range reinforces the complex multi-factorial etiology of the disorder (Petersen & Pennington, 2015).

Perhaps unsurprisingly, both developmental language disorder and dyslexia show a relationship with socioeconomic status (SES), with delayed language acquisition and poorer word reading, as well as a flatter trajectory of improvement, associated with lower SES (Hecht, Burgess, Torgesen, Wagner, & Rashotte, 2000; Kelly, 2010). However, Peterson and Pennington (2015) note that systematic reviews suggest that approximately 90% of the variation in reading outcome is not related to SES. The remaining 10% variation can be whittled down still further, as around half of this is thought to be mediated by genetic factors that themselves set a limit on social mobility within society (Petrill, Deater-Deckard, Schatschneider, & Davis, 2005).

1.8   The biological basis of developmental language disorder and dyslexia

Brain-imaging studies fail to show any evidence of obvious atypical structure in children or adults with language disorders (Bishop, 2013). However, even though gross brain structure appears to be the same in brains of individuals with language disorders and normal brains, research does point to the existence of subtle differences in brain structure and function. For example, magnetic resonance imaging has found numerous, small differences in grey matter volume between developmental language disorder and control brains, such as subtle reductions in grey matter in perysylvian structures (Jernigan, Hesselink, Sowell, & Tallal, 1991). Badcock, Bishop, Hardiman, Barry, and Watkins (2012) found reduced grey matter volume in the caudate nucleus and superior temporal cortex bilaterally and increased volume in the left inferior frontal cortex, with corresponding reduced activity in the latter two areas as well as in the right putamen in the brains of those with developmental language disorder compared to controls.

A study investigating atypical cerebral lateralization in young adults using functional transcranial Doppler ultrasonography (Whitehouse & Bishop, 2008) found

greater right-sided lateralization and bilaterality in participants with developmental language disorder, compared to normal adults and those with remediated childhood developmental language disorder. Functional MRI studies have also shown greater bilaterality in children with developmental language disorder than controls during language tasks (Bernal & Altman, 2003). However, while atypical cerebral dominance may be "a biological marker" of persisting developmental language disorder, it is as likely to be a consequence of the impaired learning of language, as any heritable cause of the disorder itself (Bishop, 2013).

Investigations into atypical brain development in developmental dyslexia are not clear-cut either. Pernet, Andersson, Paulesu, and Demonet (2009) have shown similar overall brain volume in dyslexics and controls, but subtle differences in grey matter distribution and lateralization in several areas, most strikingly, in the cerebellum. Imaging studies have also shown abnormal activation of left hemisphere language networks in the disorder (Demonet, Taylor, & Chaix, 2004; Richlan, Kronbichler, & Wimmer, 2009) and many studies have reported subtle differences in the left peri-sylvian cortex, the thalamus, corpus callosum, and cerebellum, suggestive of abnormality during brain maturation (Ramus, 2004; Habib, 2000). Ectopias in the surface layers of language-related cortex in the brains of dyslexics, suggestive of abnormal neuronal migration during early brain development, may explain some of the differences in connectivity (Kaufman & Galaburda, 1989; Galaburda, LoTurco, Ramus, Fitch, & Rosen, 2006). This atypical neuronal migration is thought to be of genetic origin and results in abnormal neuronal connectivity between cortical and thalamic areas in the dyslexic brain.

Differences in brain symmetry between dyslexic and typically developing brains have also been documented. The planum temporale, an important structure within Wernicke's area, associated with both auditory and phonological processing (Blau et al., 2010; Dehaene et al., 2010), is larger on the left than on the right side in typically developing brains. Indeed, it is one of the most lateralized structures found in the normal brain (Geschwind & Levitsky, 1968). However, it has been found to be the same size on both hemispheres in dylsexics (Cohen, Campbell, & Yaghmai, 1989;

Bloom, Garcia-Barrera, Miller, Miller, & Hynd, 2013). This difference in lateralization between normal and dyslexic brains suggests a functional link with the cognitive impairments found in the disorder. However, asymmetry of the planum temporale has not been found to be significantly associated with language laterality in normal adults (Eckert, Leonard, Possing, & Binder, 2006) and hemispheric symmetry in dyslexics has not been replicated in all studies (Leonard & Eckert, 2008). Indeed, adult dyslexic brains in some studies have been found to have increased, not decreased, planum temporale asymmetry (Leonard et al., 1993; 2001).

More recently, focus on the perisylvian language-related areas of the brain has shown promise; under-activation, reduced grey matter and atypical white matter microstructure has been found in the left temporoparietal and left inferior frontal gyrus (Broca's area) of dyslexics, with the reduced grey matter shown to pre-date reading instruction (Raschle, Chang, & Gaab, 2011; Richlan et al., 2009; Rimrodt, Peterson, Denckla, Kaufmann, & Cutting, 2010). Both of these areas are critically involved in phonological processing. Additionally, reduced grey matter in the left occipitotemporal area involved in whole word recognition, has also been related to dyslexia.

Ultimately, however, findings of structural or functional brain differences in either developmental language disorder or dyslexia must still be accepted with some caution. Individual differences in neural structure, combined with frequent small sample sizes and methodological differences between studies, as well as intrinsic difficulties in establishing cause and effect, make it difficult to draw any definitive conclusions.

1.9   Causal explanations of developmental language disorder

A number of causal explanations for developmental language disorder have been put forward. A description of the most influential accounts follows.

1.9.1   Language specific accounts of developmental language disorder

Several language-specific accounts of the root causes of developmental language disorder exist. These theories hinge on the assumption that language processing forms a distinct module in the brain and are rooted in the morpho-syntactic weaknesses that are a core symptom of the disorder. Gopnik and Crago's (1991) study of a family with autosomal language impairment led to their suggestion that the language difficulties they exhibited, such as the over-regularization of irregular verbs, were the result of impaired acquisition of the rules of grammar at an abstract morpho-syntactic level. This was attributed to a single faulty gene: the FOXP2 gene. However, subsequent research has shown that the family's impairments include a complex mix of intellectual and articulatory deficits in addition to linguistic ones, refuting the notion of a single gene for grammar. This research has failed to find evidence of any clear association between the family's impairments, genes and language and most people with developmental language disorder have normal FOXP2 genes (Bishop, 2006).

An alternative language-specific view concentrates on the difficulty children with developmental language disorder have with tense-marking. There is a stage in normal language development when young children do not yet apply the correct marking to verbs, preferring to use the simpler infinitive instead. The extended optional infinitive theory (Rice & Wexler, 1996) suggests that children with developmental language disorder never progress past this stage or are, at least, much delayed. It is certainly true that children with developmental language disorder do not use agreement or mark tenses properly. For example, Rice et al. (1995) demonstrated that children with developmental language disorder performed even more poorly than younger language-ability matched children on a past tense generation task and far below the level of their age-matched typically developing peers.

However, while language-specific theories are excellent at a descriptive level, they fail to explain why language develops in this atypical fashion in developmental language disorder, how the range of typical symptoms might relate to each other, comorbidity with other developmental disorders, or the level of heterogeneity in the disorder. In addition, the inconsistent use of the rules of grammar in developmental

language disorder is a stumbling block for any language specific theory, even if inconsistency between correct and incorrect application of grammar rules is ascribed to the learning of words and phrases by rote, rather than as lexical items. Attempts to explain developmental language disorder (or indeed both developmental language disorder and dyslexia together) at a more basic domain-general cognitive level appear to be more promising in this regard.

1.9.2 Cognitive level explanations of developmental language disorder

A prominent cognitive explanation for developmental language disorder relates to impairments in processing speed. This explanation also extends to dyslexia (see section 1.10.3.1). Processing speed theories of developmental language disorder range from the suggestion that slower information processing across both linguistic and non-linguistic domains is behind the language deficits in the disorder (Kail & Salthouse, 1994; Miller, Kail, Leonard, & Tomblin, 2001) to a tighter focus on a specific deficit in the auditory processing of rapidly changing sounds (Poldrack et al., 2001; Tallal & Piercy, 1973; Tallal, Sainburg, & Jernigan, 1991) that leads to difficulties with speech perception.

Impaired performance on a range of auditory processing tasks, such as auditory repetition, backwards recognition masking and frequency discrimination are characteristic of the disorder. For example, children with developmental language disorder have difficulty in discriminating between morphologically similar syllables, characterized by extremely short formant transitions, but perform as well as controls when formant transitions are artificially increased (Tallal & Piercy, 1975). This speed of processing deficit has also been found in visual and sensory motor tasks (Farmer & Klein, 1995). However, no relationship between processing speed and severity of the disorder has been found (Lahey, Edwards, & Munson, 2001), which would be expected if they were causally related.

Impaired phonological working memory has also been put forward as a cognitive-level explanation for developmental language disorder, supported by findings of poorer performance on tests of non-word repetition in those with the disorder

(Dollaghan & Campbell, 1998). Non-word repetition is a good predictor of language impairments in older children with "resolved" developmental language disorder (Bishop, North, & Donlan, 1996) and has been documented as a better predictor of performance than auditory repetition across a wide range of attainment tests from receptive grammar to word finding (Bishop, Carlyon, Deeks, & Bishop, 1999). However, these studies used concurrent designs, so they are not able to attribute a causal direction to their findings. Longitudinal studies have not found any evidence that non-word repetition is a predictor of language difficulties. For example, no evidence was found of any influence of non-word repetition on vocabulary growth in children followed from the age of 4 to 7 (Melby-Lervag et al., 2012). Indeed, the opposite pattern has been suggested (Bowey, 2001; Metsala, 1999), with non-word repetition ability considered as a possible consequence of vocabulary improvements.

Bishop (2006) suggests that searching for a single causative factor for developmental language disorder may be misguided. She cautions that developmental language disorder is better thought of as a disorder of multiple underlying deficits. This is not least the case because language acquisition appears to be a remarkably robust process and may well proceed unimpaired in the face of single deficits and only manifest when several deficits combine to block the developmental routes to normal language acquisition.

1.10 Causal explanations of dyslexia

Just as with developmental language disorder, there are a number of causal explanations for dyslexia.

1.10.1 A core phonological deficit

The most prominent theory of the cause of developmental dyslexia is that impaired processing of the sounds of language leads to both difficulties with oral language skills and later to problems with reading and writing (Bradley & Bryant, 1983; Ramus, 2004; Snowling, 1981, 2001; Vellutino et al., 2004).

Learning to read proceeds in stages from early visually driven associations between printed letters and word pronunciations to later more sophisticated use of phonological information to generate the letter-sound associations that drive efficient word recognition (Hulme & Snowling, 2009). It has been suggested that the process of learning to read consists of the creation of mappings between printed words, phonology and meanings, with both phonological and semantic skills interacting during the process (The triangle model of reading: Seidenberg & McClelland, 1989). It is the phonological pathways in this model that are impaired in dyslexia (Plaut, McClelland, Seidenberg, & Patterson, 1996), as a result of weakness in the phonological system. Certainly, the ability to attend to and manipulate the sounds of speech is a necessary prerequisite for the automization of the letter to sound correspondences that subserve accurate and fluent recognition of printed words. Children who go on to receive a diagnosis of dyslexia have been shown to have deficits on phonological tasks testing grapheme-phoneme knowledge prior to any reading instruction (Snowling, Gallagher, & Frith, 2003). However, not all children with pre-school phonological deficits go on to be dyslexic, which suggests that a range of cognitive risk and protective factors may be at work.

Successful results from randomized controlled trials offering remedial programmes that emphasize phonological training to children with dyslexia highlight the central role phonological awareness and processing plays in the disorder. Early, intensive and one-to-one or small group instruction in phoneme awareness, letter knowledge and the linkages between these two systems appears to be the best way to improve reading skill in children falling behind their peers (Hatcher, Hulme, & Ellis, 1994; Hatcher, Hulme, & Snowling, 2004; Hatcher et al., 2006; Bowyer-Crane et al., 2008). Such randomized controlled trials have shown promising results in closing the gap between those with reading disorders and typically developing children and evidence from imaging studies appears to support the efficacy of remedial intervention, e.g., normalized activity in the left hemisphere language networks of the brains of dyslexic children post intervention (Gabrieli, 2009). However, there is much that is still not known about how best to treat dyslexia. There are individual differences in how well children respond to intervention, as well as the length of time for which

they are able to maintain these gains (Petersen & Pennington, 2015). Those who do not respond tend to be at the more severe end of the impairment spectrum, with co-occurring issues with poor attention (Snowling & Hulme, 2011).

1.10.2 Phonological short-term memory

Impaired phonological working memory has also been mooted as a separate explanation for dyslexia over and above the acknowledged phonological deficit (Adams & Gathercole, 2000; Gathercole & Baddeley, 1990). Children with dyslexia are less good at non-word repetition than age-matched controls, in the same way children with developmental language disorder are (Gathercole & Baddeley, 1990). They also display poorer verbal short-term memory, as measured by word span and verbal immediate serial recall tests, as well as verbal but not non-verbal paired-associate learning tasks (Hulme & Snowling, 2009). Additionally, verbal short-term memory has been shown to predict reading ability in pre-schoolers (Melby-Lervag, Lyster, & Hulme, 2012).

However, the extent to which non-word repetition and other verbal working memory and short-term memory tasks are an index of phonological working or short-term memory, rather than simply a measure of phonological processing difficulties themselves is unclear (Snowling, Chiat, & Hulme, 1991; Snowling, 2006; Gathercole et al., 2005). Indeed, the range of deficits in dyslexia, from poor phonological short-term memory, poor vocabulary, broader deficits in oral language skill and grapho-motor processing speed, as well as difficulty in rapid automized naming tasks are all indicative of a phonological processing deficit (Petersen & Pennington, 2015).

1.10.3 Other causal explanations of dyslexia

There are a number of other causal explanations for dyslexia that, in the main, either seek to augment the phonological deficit explanation or provide an explanation at a more basic cognitive or neural level.

1.10.3.1 Processing Speed in Dyslexia

Slower domain-general information processing has been suggested as a cause of dyslexia (Catts, Gillispie, Leonard, Kail, & Miller, 2002), as well as developmental language disorder. For example, dyslexic adults have shown speed of processing deficits on a range of auditory and visual tasks (Breznitz & Meyler, 2003), but the deficit is most pronounced for measures relating to phoneme awareness. However, the existence of a generalized processing speed impairment in dyslexia, separable from an impairment in phonological processing has been disputed (Mody, Studdert-Kennedy, & Brady, 1997), as has the existence of a specifically auditory processing impairment. For example, although dyslexic children performed worse than typically developing ones in discriminating between CVC syllables, stretching the synthesized stimuli to increase their duration did not improve dyslexic performance (McAnally & Stein, 1997).

The inconsistent results relating to processing speed (for both dyslexia and developmental language disorder) may be due to the variable age of children at testing, the heterogeneity of the disorders, as well as difficulty in sustaining attention across tasks (Hulme & Snowling, 2009). It has been suggested that processing impairments may be a contributory rather than the core risk factor in both disorders (Bishop, Carlyon, & Deeks, 1999) and while processing speed serves as a good explanation of general learning difficulties, it is less good at explaining the language-specific impairments in either dyslexia or in developmental language disorder (Hulme & Snowling, 2009).

1.10.3.2 Magnocellular deficit hypothesis of dyslexia

Another explanation for dyslexia attributes it to abnormalities in the magnocellular part of the visual system, which responds to changes in stimulation caused by movement in the visual field (Skottun, 2000; Stein, Talcott, & Walsh, 2000). Deficits in this system in dyslexics lead to poor suppression of visual sensitivity during the saccades that are an integral part of the reading process (Lovegrove, 1991). While the magnocellular deficit hypothesis is no longer thought of as the single factor underlying dyslexia, it is still considered as a possible contributory cause. A recent study showed

that motion detection is impaired in dyslexic children and that pre-reading visual motion perception is a predictor of later reading skill, independently of phonological skill (Gori, Seitz, Ronconi, Franceschini, & Facoetti, 2016).

However, the symptoms attributed to the magnocellular deficit (problems with saccades in dyslexics, differences in timing, duration and steadiness of fixations during reading) has also been related to another general causal account of dyslexia and language disorder, the cerebellar deficit hypothesis (Stoodley & Stein, 2013).

1.10.3.3 Cerebellar dysfunction

Cerebellar dysfunction has been implicated in dyslexia with depressed cerebellar activation during phonological and reading tasks seen in adult dyslexics compared to controls (Brunswick, McCrory, Price, Frith, & Frith, 1999), as well as broader cerebellar activation in dyslexia that is more normally seen in younger readers (Marien et al., 2014). Similarly, dyslexic children and adults may perform more poorly on motor-related, eye-movement control and postural stability tasks that depend on cerebellar function than controls, in addition to the impaired implicit motor learning noted by some, but not all, researchers on the serial reaction time tasks. Nicolson and Fawcett (1999) cite cerebellar dysfunction as the root cause of dyslexia as a result of the disrupted automization of learned skills that leads to impaired phonological awareness. An alternative explanation for the diffuse cerebellar activation seen in dyslexia has been put forward by Baillieux et al. (2009), who suggest the dysfunction reflects an impairment in the processing and transfer of information within the cerebellar cortex.

However, the relationship between cerebellar deficits and dyslexia is not clear-cut. For example, although cerebellar damage has been linked to acquired reading difficulties, not all patients with cerebellar damage have reading impairments (Stoodley & Stein, 2013). Similarly, not all studies assessing language-disordered groups on tasks indexing cerebellar function have found a link between them (Irannejad & Savage, 2012). Finally, while smaller grey matter volume in the right cerebellar lobule IV has been found to be a reliable biomarker for dyslexia (Pernet,

Andersson, Paulesu, & Demonet, 2009), it is not found in pre-readers at risk for the disorder, so could just as well be a result of reading difficulties rather than the cause (Bishop, 2002).

In a recent review Stoodley and Stein (2012) concluded that cerebellar dysfunction is unlikely to be the primary cause of dyslexia, but is more likely to be an outcome of a more fundamental and general abnormality in the dyslexic brain. This theory, therefore, arguably serves as a precursor to the procedural deficit hypothesis (Nicolson & Fawcett, 2007; 2011; Ulman, 2004; Ullman & Pierpont, 2005) that will be considered in detail in Chapter 3.

In summary, both developmental language disorder and dyslexia are relatively common developmental disorders of language learning. Dyslexia is most commonly linked to deficits in phonological processing, while the impairments seen in developmental language disorder are more wide-ranging and involve difficulties with semantics and syntax, as well as with phonological processing. Although they are distinct disorders, their dimensional nature, heterogeneity and frequent comorbidity with each other and with other developmental disorders renders diagnosis problematic, as well as making it difficult to classify language-disordered participants accurately into groups for research purposes. It also complicates any search for causal explanations, either those specific to a particular disorder or those that attempt to explain the pattern of impairments seen more generally across both disorders. A number of causal explanations have been put forward, ranging from language specific accounts of developmental language disorder to domain-general cognitive level explanations, such as processing speed. While these and other accounts are good at explaining aspects of the disorders, none of them adequately explain the full, complex and multi-factorial nature of the disorders. We will meet a causal explanation that attempts to do this in Chapter 3.

Chapter 2   Implicit and Explicit memory

Before going on to describe the procedural deficit hypothesis in detail it is necessary to outline some key theories and research relating to the organization of human memory.

2.1   Multiple Memory Systems

Long-term memory as multiple parallel systems in the brain is an idea that has been explored by scientists for well over a hundred years. William James (1890) deliberated over the different processes underlying memory and habit over 120 years ago. Among many others exploring the organization of memory in the ensuing years, Ryle (1949) eloquently summed up the distinction between two kinds of memory as the difference between "knowing that" and "knowing how". The advent of computers, led to further exploration of the idea of parallel and separate memory systems, which first introduced the terms procedural and declarative knowledge to differentiate between the results of these different processes in the brain (Winograd, 1975). Cohen and Squire (1980) distinguished between declarative and procedural memory, while Graf and Schacter (1985) wrote about the distinction between what they termed explicit and implicit memory.

2.1.1   Multiple Memory Systems Taxonomy

With evidence of dissociations between different kinds of memory, increasingly researchers felt the need to develop more nuanced taxonomies, in order to accommodate their experimental data (e.g., Tulving, 1985; Tulving & Schacter, 1990). Accounts were put forward that partitioned memory into declarative and non-declarative systems, but within non-declarative memory sat several additional memory systems (e.g., Shacter & Tulving, 1994).

These classifications were developed further to fit the data coming in from biological studies of the brain. See Figure 2.1 for the taxonomy put forward by Squire (1987; 2004; Squire & Dede, 2015), which is representative of the multiple memory systems view. It allocates declarative learning for facts and events to the medial

temporal lobe and diencephalon, which includes the hippocampus. Under the umbrella of non-declarative learning, it aligns skills and habit learning with the basal ganglia (the striatum in particular); priming and perceptual learning is associated with the neocortex; classical conditioning with the amygdala (for emotional responses) and the cerebellum (for skeletal responses); and non-associative learning with the reflex pathways. Subsequent developments have added associations for procedural learning with the motor cortex and cerebellum (e.g., Bartsch & Butler, 2013).

It should be noted that this taxonomy relates specifically to long-term memory. The other distinction of the multiple systems view distinguishes between short-term or working memory and long-term memory (see Chapter 3 for further detail on this distinction and how it relates to the debate surrounding the procedural deficit hypothesis).
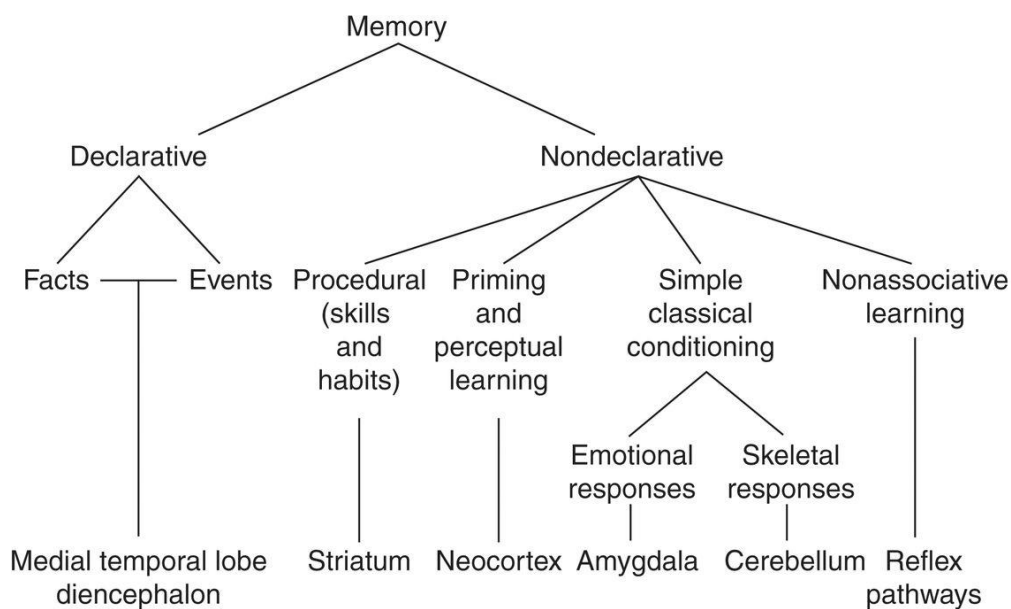
Figure 2.1 The hypothesized taxonomy of long-term memory systems and associated brain structures. Taken from Squire and Dede (2015).

2.1.2　Declarative memory and learning

Declarative memory is what is typically referred to as memory in everyday language. It refers to the ability to consciously recollect facts and events. Declarative memory uses deliberate strategies and is representational, allowing material to be compared and contrasted. This is what enables memories to be encoded in terms of their relationships with other items and events (Squire, 2004).

Declarative memory can be divided into semantic memory for facts about the world, concepts, and meanings and episodic memory for events, which gives us the ability to re-experience an event in the context in which it originally occurred (Tulving, 1985). Episodic memory is autobiographical, involving conscious recollection of past events. Semantic memory is knowledge of the world that we accrue without recollection of the context surrounding its learning and, as such, it is for the most part dissociated from episodic memory (Tulving, 2002). Crucially, information encoded into declarative memory can be expressed with language (Squire, 1987).

2.1.2.1　The neural substrates of declarative learning

The medial temporal lobe network underpins the declarative memory system. The hippocampus is at the centre of this system, receiving projections from most neocortical association regions either directly (linking to the CA1, CA2, CA3 areas of the medial temporal lobe and subiculum) or indirectly via the parahippocampal, perirhinal and entorhinal cortices (Bartsch & Butler, 2013). Such high levels of connectivity provide it with access to an enormous wealth of information. The hippocampus enables swift memory formation as a result of a particular form of long-term potentiation that depends on N-methyl D-aspartate (NMDA) receptors (Martin, Grimwood, & Morris, 2000) to ensure rapid synaptic plasticity. This superior connectivity and plasticity results in memory formation in terms of associative representations, such that presentation of any part of the representation leads to retrieval of the whole (Treves & Rolls, 1994). This includes the rapid formation and retrieval of temporally sequential representations (Eichenbaum, 2004). The hippocampus is also connected to subcortical circuitry, in particular the thalamus and

this connection is particularly important for episodic memory and to facilitate interaction with other memory systems (Bartsch & Butler, 2013).

2.1.3   Non-declarative memory and learning

A definition of non-declarative learning is not straightforward, as different avenues of research define and use terminology in different ways. Firstly, the multiple memory systems account refers to the procedural memory system, which is the focus of this thesis, as only one type of implicit memory system. It is the system which regulates the acquisition, consolidation and automization of both motor and cognitive skills and habits (Squire, 2004) and that is required for performance of skilled motor actions, such as bike-riding and the perceptual-cognitive skills that make the fluent use of language possible (Ullman & Pierpont, 2005). Priming (as well as conditioning and non-associative implicit memory) is considered as a separate implicit memory system according to the multiple memory systems view. However, it is generally agreed that the terms procedural learning and implicit learning are largely synonymous (Shanks, 2005; Berry & Dienes, 1993) or at least overlapping (Seger, 1994). A task is learned implicitly if procedural knowledge develops without, or at least before, any declarative knowledge and it is the procedural system which is necessary in order to perform implicit tasks (Berry & Dienes, 1991). In what follows implicit learning and procedural learning will be used interchangeably, as will explicit and declarative learning.

Secondly, a distinction can be made between research into implicit memory and into implicit learning that echoes the distinction between the procedural and priming pathways of the multiple memory systems taxonomy. Berry and Dienes (1991) refer to how little cross-referencing there is between these two research traditions, leading to a frequent supposition that they refer to very different things. Implicit memory has traditionally been investigated using priming tasks, such as word stem completion, while implicit learning research uses paradigms that will be the focus of this thesis, such as the serial reaction time and Hebb serial order learning tasks, artificial grammar learning, probabilistic categorization and contextual cueing (see Chapter 3 for a detailed explanation of these paradigms). The two research traditions are separated by

their experimental approach. However it can be argued that both fields are investigating similar distinctions and that the same cognitive processes underlie performance on the paradigms in both traditions (Reber, 2008).

Frensch & Runger (2003) refer to there being at least a dozen different definitions of implicit learning, most of which benefit from being given in relation to explicit learning. At the most surface level the distinction is made between implicit and explicit learning as learning without or with awareness respectively. Reber, Walkenfield and Hernstadt (1991) also focus on a dissociation from awareness as being the crucial factor in distinguishing between the two, such that implicit knowledge is acquired without awareness of both the learning process and the information learned. However, defining implicit learning only in terms of what it lacks does not give us a full understanding of how it differs from explicit learning (Reber, 2013). Other characteristic features of implicit learning distinguish it from explicit learning. The knowledge acquired in implicit learning is difficult to access; frequently combines with a subjective sense of intuition; is associated with incidental learning conditions; is robust to decay and interference; is rigid and is subject to considerable specificity of transfer, so it can typically only be applied within the specific circumstances in which it was learned (Berry & Dienes, 1993; Reber, 1993). In the case of procedural implicit learning it is also slow to develop, as it gradually extracts the common elements from strings of separate events (Reber, 2013). All these features are seen as in opposition to the characteristics of explicit learning, which uses deliberate strategies; is accessible to consciousness; is flexible and, once learned, can be applied in various ways; and can be expressed on demand.

2.1.3.1 The neural substrates of implicit learning

The procedural memory system is made up of a network of several interconnected brain structures – the cortico-striatal-pallidal-thalamo-cortical circuitry system (Seger & Miller, 2010; Squire, 2004). The basal ganglia are arguably the hub of this network and are a distributed set of sub-cortical structures that include the globus pallidus and the striatum (itself divided into the caudate nucleus and putamen), as well as the more distant, but connected subthalamic nucleus and sustantia nigra. Within the system the

caudate and putamen are the main input nuclei from the rest of the cortex (Grahn et al., 2009). The system is split into dorsal and ventral streams. The former is connected primarily via the caudate (Grahn et al., 2009) and is particularly implicated in learning and memory (Packard & Knowlton, 2002). The latter loop, connected to sensory and motor areas via the posterior putamen, appears to be more specialized for motor learning (Grahn et al., 2009). Neurobehavioural experiments in animals and humans have shown this area to mediate the learning of incrementally acquired stimulus-response associations, probabilistic rule learning (Knowlton, Mangels, & Squire, 1996; Packard, Hirsch, & White, 1989) and sequence learning (Doyon et al., 1997) and working memory (Wise, Murray, & Gerfen, 1996) among other processes.

The cortico-striatal circuitry of the procedural memory system has been divided into four striatal loops with different cognitive specifications during learning, associated with different cortical connections (Seger, 2006). The executive loop links the anterior caudate and the prefrontal cortex; the visual loop links the posterior caudate and visual cortex; the motor loop links the putamen and motor cortex; and the motivational loop links the ventral striatum and ventromedial frontal cortex. These loops may be differentially involved depending on the nature of learning required. Seger (2006) suggested that the role of the striatum is to react to the learning context to modulate subsequent cortical processing and in support of this, striatal activation during learning has been demonstrated prior to cortical activity. However, alternative interaction processes have also been put forward (Packard & Knowlton, 2002).

There are other important structures involved in the procedural learning system including the frontal cortex, the pre-motor cortex (which includes the supplementary motor cortex) and Broca's area. In addition, research has also pointed to the involvement of the cerebellum in both motor and non-motor implicit learning. Sequence processing is at the heart of cerebellar function (Leggio et al., 2008) and it is suggested that the cerebellum is involved in the prediction of sequences based on the comparison of incoming sequences of stimuli across multiple cognitive domains. As such, it must work in tandem with working memory in order to maintain this information for comparison.

To recap, the multiple memory systems view divides memory into separate systems. The conscious, declarative memory system is associated with the medio-temporal lobe, incuding the hippocampus, and is responsible for the encoding and storage of semantic facts and memory for events. The procedural memory system operates without consciousness and is involved in the learning of motor and cognitive skills and habits. The neural substrate of procedural memory is the cortico-striatal system, which includes the basal ganglia and cerebellum.

2.2   Evidence for multiple memory systems

Much of the evidence for separate declarative and procedural memory systems comes from dissociations in learning found in patients following brain damage. In one of the earliest mentions of a dissociation between memory processes, Korsakoff (1889) wrote about an amnesiac patient who identified the purpose of an electric shock appliance that had been used on him previously, in spite of having no conscious recollection of earlier shock treatment. Experimental exploration of dissociations between implicit and explicit memory began in earnest with Milner's (1968) discovery that a severely amnesiac patient, HM, was able to "learn", improving in skill at mirror drawing over time, while simultaneously being unable to explicitly recall any of the training sessions. The earlier bi-lateral removal of HM's hippocampi had resulted in his complete loss of ability to form any new declarative memories, yet his ability to learn procedurally was still, at least partially, intact. HM and other amnesiacs with medial temporal lobe damage (e.g., Zola-Morgan, Squire, & Amaral, 1986) also suffered from retrograde amnesia that extended back several years, but displayed ongoing intact short-term memory. This supported the theory that the hippocampus is responsible for the formation of new long-term memories, a process that happens over time, but gradually becomes less and less involved with retrieval of consolidated declarative memories (Squire, Cohen, & Nadel, 1984), although some recent research points to a role for the hippocampus in retrieval too (Rekkas & Constable, 2005; Wincour & Moscovitch, 2011).

Dissociations between memory systems in global amnesia have also been shown using many other tasks such as visual motor tracking using the pursuit rotor task, where

participants try to follow a small disc on a rotating turntable (Cermak, Lewis, Butters, & Goodglass, 1973) and the reading of mirror writing. For example, amnesiacs perform as accurately as controls on the pursuit rotor task, in the face of poorer verbal long-term memory (Brooks & Baddeley, 1976) and improve their mirror reading at the same rate as controls, despite being less able to remember the words they read (Cohen & Squire, 1980). Over the last 35 years, investigations of visuo-motor implicit sequence learning using many variations of the serial reaction time task have more often than not shown that amnesiacs can learn about the sequential structure in the task without intention and that they very often do so without any explicit knowledge of the learning they have accomplished in order to perform the task. For example, patients with Korsakoff's syndrome (which involves bilateral hippocampal damage) performed as well on the serial reaction time task as a normal control group (Nissen & Bullemer, 1987), with equivalent levels of consolidation on a repeat of the task one week later (Nissen, Willingham, & Hartman, 1989). Hebb sequence learning is also intact in amnesic patients with hippocampal lesions across both visuospatial and verbal modalities (Gagnon, Foster, Turcotte, & Jongenelis 2004). Amnesic patients also achieve above chance performance on tasks assessing their implicit learning of the grammatical rules required to perform artificial grammar tasks (Knowlton, Ramus, & Squire, 1992), but are less able than controls to recognize exemplars from the learning phase of the experiment afterwards.

Evidence from amnesic patients on two other implicit learning paradigms is less clear-cut. Normal learning rates have been found in amnesiacs on the weather prediction task, a measure of probabilistic category learning (Knowlton, Squire, & Gluck, 1994; Knowlton, Mangels et al., 1996). However, closer inspection has shown intact learning early in training, but impaired performance later in the task. Knowlton et al. (1994) postulated that implicit associative learning supported amnesics' early performance, but that later in the task hippocampus-dependent declarative processes may have aided the performance of the control group. Impaired contextual cueing, a task that assesses the implicit learning processes involved in the development of visual search efficiency, has also been found in some amnesic patients (Chun & Phelps, 1999; Manns & Squire, 2001). Although this can be seen as indicative of impairments in

perceptual priming, it has been suggested that performance in the task also benefits from explicit memory processes that involve the medial temporal lobe (Westerberg, Miller, Reber, Cohen, & Paller, 2011).

Studies of priming effects have demonstrated the functional, as well as stochastic, independence of implicit memory. Amnesic patients have shown impaired free recall and recognition for previously presented list of words, while at the same time demonstrating unimpaired retention when prompted using implicit word stem or fragment completion primes (Warrington & Weisenkrantz, 1968; 1970; 1974). Furthermore, the strategy patients were asked to use in order to complete the tasks led to differences in performance (Graf, Squire, & Mandler, 1984). When amnesiacs were asked to use word stem prompts to aid recall, they were impaired compared to controls, but they performed at an equivalent level when asked to write down the first word that came to mind upon hearing the word stem prompts.

However, inferences arising from demonstrations of the stochastic independence of implicit and explicit memory have been subject to a variety of criticisms. One criticism hinges on the fact that some studies demonstrating stochastic independence of implicit and explicit memory are demonstrably underpowered (Poldrack, 1996). Sample sizes in these studies are too small and the number of test items too few to find statistical dependence of implicit and explicit measures, which in turn undermines any resulting conclusions about their statistical independence. Additionally, the changes in performance that relate to memory are small compared to baseline performance on such tasks (Ostergaard, 1992) and this means that statistical dependence in the portion of the task that relates to memory can be masked by the statistical independence on the greater portion of performance that is unrelated to memory. In this way such tests may be biased towards independence (Poldrack, 1996).

2.2.1 Evidence from Alzheimer's disease and schizophrenia

Intact implicit learning has been found across a wide range of other neuropsychological disorders acknowledged to affect explicit learning. Patients with Alzheimer's disease display gains in mirror drawing similar to those made by

amnesiacs, without the recollection of the activity to go with it (Gabrieli, Corkin, Mickel, & Growdon, 1993). Intact implicit learning on serial reaction time tasks as well as contextual cueing effects have also been found in schizophrenia (Dominey & Georgieff, 1997; Lamy, Goshen-Kosover, Aviani, Harari, & Levkovitz, 2008).

2.2.2 Evidence from diseases of the basal ganglia

Evidence for a dissociation between declarative and procedural memory systems is also found in patients with disorders of the basal ganglia. For example, impaired procedural visuo-motor learning using the serial reaction time task has been found in patients with Parkinson's and Huntington's disease, both diseases affecting the striatum. Huntington's patients reliably show impaired learning on the serial reaction time task (Knopfman & Nissen, 1991). Impaired implicit learning has also been found in Parkinson's disease, with two meta-analyses of serial reaction time task performance in Parkinson's patients finding a moderate overall effect size reflecting impaired performance, albeit with significant heterogeneity between studies (Siegert, Taylor, Weatherall, & Abernathy, 2006; Clark, Lum, & Ullman, 2014). However, not all behavioural investigations of implicit sequence learning in Parkinson's patients are consistent with the above and it is suggested that disease severity may be one moderating factor on performance (Grahn et al., 2009). Corkin (2013) reported experiments that demonstrated intact learning on the serial reaction time task in Parkinson's patients, including consolidation of such learning, although they did show less improvement than controls on HM's mirror tracing task. There is a difference in how the striatum is affected in Parkinson and Huntington disease. Parkinson disease initially affects the putamen as a result of decreased dopamine transfer from the substantia nigra (Dauer & Przedborski, 2003), while Huntington disease initially causes neuronal death in the caudate nucleus (Lawrence et al., 1998). Corkin interpreted her findings as evidence for a specific role for the caudate nucleus in implicit sequence learning.

Both Parkinson and Huntington disease patients have shown impairments on other implicit learning paradigms that involve less of a motor learning component, such as artificial grammar learning (Smith, Siegert, McDowell, & Abernathy, 2001) and the

weather prediction task (Knowlton, Mangels & Squire, 1996; Knowlton, Squire et al., 1996; Poldrack et al., 2001).

2.2.3   Demonstrations in normal participants

There is evidence of memory system independence in normal participants too. For example, Willingham et al. (1989) showed that implicit and explicit learning of a sequence on a serial reaction time task developed independently of one another. Similarly, in several fragment completion tasks priming effects were uncorrelated with recognition performance (Hayman & Tulving, 1989a), with priming effects enduring for months, while recognition memory degraded substantially over the same time period (Tulving et al., 1982). Neuro-imaging also points to differential recruitment of brain areas associated with declarative or procedural learning dependent on the mode of learning used to perform artificial grammar tasks (Yang & Li, 2012) categorization tasks (Reber, Gitelman, Parrish, & Mesulam, 2003) and the serial reaction time task (Destrebecqz et al., 2005).

However, not all research is consistent with these findings, as medial temporal lobe activation has been found on both implicit and explicit versions of a serial reaction time task (Schendan, Searl, Melrose, & Stern, 2003). Additionally, demonstrations of the functional independence of implicit and explicit learning on the serial reaction time task in normal adults have been criticized. For example, development of implicit learning of a sequence prior to explicit learning in serial reaction time tasks is claimed as evidence to support the multiple memory systems view. Perruchet and Amorim (1992) found a difference in RTs between performance on repeated and random transitions in normal adults within only 60 trials, which at first glance is consistent with this interpretation, yet they were able to demonstrate that this corresponded with explicit awareness of the sequence within the same time frame. The portions of the sequence that led to faster RTs in the implicit learning measure were the same chunks that were available to declarative memory in a post-task free generation test. Similarly, Perruchet, Gallego, and Savy (1990) provided an alternative explanation for the faster RTs to transitions that conformed to a complex second order conditional sequence on a serial reaction time task (Lewicki, Hill, & Bizot, 1988), since participants were

shown to have explicit knowledge that the target in this task moved through all possible locations before returning to previous ones. This knowledge was sufficient to explain the improvement in RTs that had previously been ascribed to implicit learning of the complex sequence.

2.2.4   Age and IQ independence

Support for a multiple systems view comes also from a developmental perspective. It is suggested that declarative learning develops with age, while intact implicit learning has been demonstrated from early infancy and appears to be age-independent (Kirkham, Slemmer, & Johnson, 2002; Moscovitch, 1985; Nelson, 1995; Ofen et al., 2007). Intact implicit learning on the task has been shown in children, with no age-related differences in performance between adults and children. (Meulemans, Van Der Linden, & Perruchet, 1998; Thomas & Nelson, 2001). For example, Meulemans et al. (op cit) demonstrated equivalent levels and patterns of implicit learning on a serial reaction time task in six year olds and adults. Intact Hebb sequence learning has also been found in typically developing children (Mosse and Jarrold, 2008). Evidence is less clear-cut in research using non-sequential implicit learning tasks though. Contextual cueing effects have been found in typically developing children as young as five years old (Dixon, Zelazo, & De Rosa, 2010; Merrill, Conners, Roskos, Klinger, & Klinger 2013). However, the degree to which contextual cueing is present in childhood is disputed, with ten year olds (Couperus, Hunt, Nelson, & Thomas, 2011) and six to thirteen year olds (Vaidya, Huger, Howard, & Howard, 2007) not exhibiting the cueing effects shown in adults.

Such implicit learning is thought to play a large part in early development (Karmiloff-Smith (1992). Even one year-old children have been shown to listen preferentially to legal generalizations from an artificial grammar (Gomez & Gerken, 1998), suggesting not just the existence of intact learning at an early age, but also the likelihood that these processes are involved in language acquisition. Indeed, Bulf, Johnson, and Valenza (2011) have even demonstrated evidence of visual statistical learning for sequences of black and white shapes in babies as young as one to three days old, using a habituation procedure.

50

Reber (1993; 2013) suggests an evolutionary justification for this apparent lack of a developmental trajectory for implicit learning: that it is supported by primitive learning mechanisms that mature early to aid survival. This is consistent with the finding of implicit learning in other species, from contextual cueing in baboons (Goujon & Fagot, 2013) and pigeons (Wasserman et al, 2014) to rule-abstraction in rats (Murphy, Mondaragon, & Murphy, 2008) and statistical learning of some types of non-adjacent regularities in primates (Newport, Hauser, Spaepen, & Aslin, 2004). Cotton-top tamarins have even been shown to use statistical information to segment artificial words from a speech stream in a similar way to human babies (Hauser, Newport, & Aslin, 2001; for a review, see Conway & Christiansen, 2001). Results are similar at the other end of the age scale too, with the same level of implicit learning in undergraduates and over-65's, combined with poorer declarative learning in the older group (Howard & Howard, 1992).

In addition, it is suggested that implicit memory is insensitive to IQ (Reber et al., 1991). Procedural learning on the serial reaction time task has been shown to be independent of measures of IQ, but correlated to declarative learning performance (Feldman, Kerr, & Streissguth, 1995). Similarly, contextual cueing effects have been found in participants with learning difficulties (Merrill, Conners, Yang, & Weathington, 2014). While results in this area are once again inconsistent, this could be ascribed to contamination by explicit knowledge, which is both age- and IQ-dependent (Vintner & Perruchet, 2000; Shanks & John, 1994).

2.3   Interaction and Competition

Although the multiple systems view of memory sees declarative and procedural memory as separate systems, it is thought that there is some interaction between them (Brown & Robertson, 2007). Research has shown that consolidation of procedural learning can be blocked by subsequent declarative learning and vice versa. This suggests a dynamic relationship between the two systems. Procedural dorsal stream basal ganglia circuits and declarative medial temporal lobe mechanisms may compete with each other during learning (Poldrack & Packard 2003), such that once automaticity is acquired conscious knowledge can interfere with performance. For

51

example, activation of the caudate nucleus within the basal ganglia and concurrent deactivation of the medial temporal lobe has been shown in participants undertaking the probabilistic categorization "Weather Prediction" task, such that activity in the two brain regions is negatively correlated (Poldrack, Prabakharan, Seger, & Gabrieli, 1999).

In summary, there is much research to suggest a dissociation between declarative and implicit memory processes, along with evidence from lesion studies and brain imaging that suggests these processes are served by different systems in the brain: declarative learning, by the cortico-hippocampal system and procedural learning by the corticostriatal system. However, some of this research has been subject to criticism on methodological grounds. Although evidence for separable memory systems in the brain is strong, there is not universal agreement that the multiple systems view is correct.

2.4   Other accounts of the organization of memory

Alternative views of the organization of memory in the brain exist. At the extreme, the very existence of implicit memory is disputed, with implicit learning performance on tasks accounted for by fragmentary knowledge of specific events (Perruchet & Pacteau, 1990). Connectionist networks have been able to simulate many of the findings in the implicit learning literature using network models that assume no more than stimulus representations and the associations between them, which would make implicit learning the incremental, distributed change in an associative pattern that is sensitive to the underlying statistical structure of the task (Cleeremans & McClelland, 1991; Stark & McClelland, 2000; Perruchet & Vinter, 1998).

For example, Cleeremans and McClelland (1991) showed how an apparently simple connectionist information processing model was able to account for the procedural learning of sequential patterns found in several noisy probabilistically structured artificial grammar tasks. Their recurrent network simply allowed for back-propagation, such that processing at time $t-1$ was able to influence processing at time t. This enabled all the forward-going connections in the model to be modified by back

propagation, even though the model's processing resources were extremely limited. Once it incorporated some simple factors related to short-term priming effects that were evident from analysis of their experimental work, it yielded a good fit to the data. On a theoretical level the model shared many of the characteristics of implicit learning: the representations it learned associatively could only be expressed through performance, yet learning resulted in complex and highly structured knowledge of the optimal conditional probabilities of the task.

An alternative connectionist explanation for the learning evident in implicit tasks uses a connectionist model (PARSER: Perruchet & Vinter, 1998) that initially forms chunks randomly as a natural consequence of attentional processing. These are then strengthened or weakened dependent on subsequent input. The components of repeated chunks are associated to form new representational units and the process repeats, while those that are not repeated rapidly decay. The model uses forward interference to make chunk strengths sensitive to the transitional probabilities of the implicit learning tasks or language input, but once again does so with minimal memory demands and no requirement for separate memory systems. The authors suggest these same processes are at work during word segmentation in infants.

In support of these models Berry, Shanks, Speekenbrink, & Henson (2012) point out that the existence of just one memory system can be seen as a more parsimonious explanation of learning and memory and that connectionist models such as those above should be preferred if for that reason alone.

Others also dispute the existence of a separate system dedicated to implicit learning. Marsolek and Bowers (2003) claim instead that implicit learning is the by-product of the activity of brain systems that engage in perceptual pattern recognition, conceptual processing and motor behavior. In other words it is simply the physical structure of the brain that enables it to retain traces of stimulation. This does not require an implicit learning system per se. Marsolek and Bowers (2003) specifically focus on implicit memory and priming, but similar views are also seen in the implicit learning literature. Reber (2013) suggests that implicit learning should be viewed as a highly

adaptive, emergent property of the plasticity of the information processing circuits of the brain. It is, therefore, involved to an extent in every form of adaptive behavior change and is what enables people to extract the statistical relationships in the environment to their advantage in as efficient a manner as possible. He labels this "pervasive plasticity" (p2028). Where these plastic changes occur outside the medial temporal lobe brain area allied to explicit memory and learning, the result is simply dissociated from awareness. Reber's view provides a shared universal mechanism for all implicit learning, seeing implicit learning as a unitary phenomenon, but at the same time denies there is such thing as an implicit learning system in the brain that exists as a separate entity from a declarative learning system.

Other single system views see the difference between implicit and explicit learning as a difference in processing, rather than systems. Processing views in the 1980s and 1990s provided an alternative to the multiple memory systems account, ascribing the apparent differences between implicit and explicit memory to the level of processing applied to the learning episode. Mandler's activation view (1980) held that implicit and explicit "memory" represented the difference between automatic processing and integration of existing knowledge and effortful processing and elaboration that integrated new knowledge, all within the same memory system. However, this view did not convincingly explain why implicit learning is so robust over time. Roediger's transfer-appropriate processing account (Roediger & McDermott, 1993) claimed that the apparent dissociations between implicit and explicit memory are the result of different processing requirements of the information learned. Implicit memory uses bottom up, data-driven processing, while explicit memory uses conceptually driven, top-down processing. Jacoby's (1991) process dissociation framework similarly viewed explicit memory to be the result of effortful, controlled and deliberate processing and compared this to the effortless, automatic and involuntary processing that results in unconscious implicit memory. Experimental evidence of an apparent dissociation between implicit (perceptual learning) and explicit (recognition) learning dependent on the level of processing required supports this view (Jacoby & Dallas, 1981). Participants demonstrated better recognition of previously presented words when they had been asked to think about them in terms of meanings (elaborative

54

processing), than when they had simply been asked to count letters (superficial processing). However, when words were primed by flashing them briefly on a screen the level of processing had no effect on performance. However, coming seemingly full circle, this dissociation can also be explained by the multiple memory systems view (Tulving & Schacter, 1990), as different memory systems are likely to be associated with different processes (Berry & Dienes, 1991).

Arguments against processing views cite the considerable body of biological evidence mentioned above that would seem to support the existence of multiple memory systems. However, a final neural-based processing view worthy of mention does support the existence of separate memory systems, but not based on the criterion of consciousness (Henke, 2010). This view is consistent with the neurophysiological evidence of brain-based dissociations. It proposes that the medial temporal lobe is involved in the formation of both conscious and unconscious memory, but what distinguishes memory involving the medial temporal lobe is the flexible, associative nature of the encoding it provides. Such memory can be reactivated in many different ways because it is compositional. The account implicates the CA3 area of the medial temporal lobe system in particular as the source of the associative binding of memories that enables storage both as many individual elements and as their associations separately. This is contrasted with the rigid formations of unitized procedural memory that occurs without recruitment of this part of the medial temporal lobe. Formation of this type of procedural memory happens in the basal ganglia, cerebellum, parahippocampal gyrus and neocortex and is characterized by slow encoding over multiple trials. The result of this rigidity is that this type of memory is modality dependent, requiring the same cues that were present during learning for retrieval. Priming is similarly stored as a single inflexible unit with the same restrictions on retrieval, but is rapidly encoded as a single item. It is these differences in how memories are processed, stored and retrieved that combine to give the impression of memory that exists above and below the criterion of consciousness.

The above processing views are compatible with findings that the hippocampus is also involved in the statistical learning of sequential and probabilistic regularities

(Shapiro, Gregory, Landau, McCloskey, & Turk-Brown, 2014; Shapiro, Turk-Browne, Bovinick, & Norman, 2016). For example, an fMRI study of serial reaction time learning under both implicit and explicit conditions (Schendan, et al., 2003) did show activation of the striatum during implicit learning, but also activation of the medial temporal lobe under both conditions, perhaps pointing to the recruitment of this brain area for the learning of higher order associations, regardless of conscious knowledge of the sequence. Additionally, implicit visual associative learning has also been found to recruit the medial temporal lobe (Degonda et al., 2005). While medial temporal lobe involvement during implicit learning is difficult to reconcile with multiple memory systems accounts, Henke (2010) suggests that it may be recruited in conditions that benefit from flexibility to increase the efficiency of learning, such as conditions that require both contextual and temporal bindings.

In summary, the multiple systems view of memory divides long-term memory into two separate systems of conscious declarative memory and non-declarative implicit memory. Non-declarative memory operates without consciousness. Procedural memory for motor and cognitive skills and habits is one of the components of this non-declarative memory system. Evidence to support the multiple systems view consists mainly of dissociations between implicit and explicit declarative learning in patients with brain damage: either intact procedural learning and impaired declarative learning in patients with hippocampal damage or the opposite pattern in patients with diseases of the basal ganglia. Dissociations are also demonstrated in the absence of brain damage too, with implicit learning showing evidence of both age and IQ invariance, which is not seen in declarative learning. Although separate, these systems are thought to interact competitively with one another, in order to support optimal learning.

However, although evidence for separable memory systems is strong, the multiple systems view is not universally supported. Alternative unitary accounts of memory exist that explain the apparent difference between the two types of learning in different ways: implicit learning may be the inevitable outcome of neuronal plasticity; or the result of limited processing resources like those required by a connectionist model; or it may reflect a difference in the type processing used in its formation, e.g., bottom-

up, rather than top-down processing or the creation of rigid, rather than flexible, memory representations.

With an understanding of the main positions in the ongoing debate surrounding the organization of memory in the brain, we now return to the exploration of the underlying causes of developmental disorders of language.

Chapter 3   The Procedural Deficit Hypothesis

3.1   Summary of the hypothesis

This chapter focuses on a recent cognitive-level, memory-based explanation for both developmental language disorder and dyslexia, which relates the impairments found in these disorders to an underlying deficit in the procedural memory system (Nicolson & Fawcett, 2007; 2011; Nicolson, Fawcett, & Dean, 2001; Ullman & Pierpont, 2005). Such a deficit may represent the cause of the grammatical and phonological processing impairments seen in developmental language disorder (Ullman & Pierpont, 2005) and the phonological processing impairments in dyslexia (Nicolson & Fawcett, 2007).

The procedural deficit hypothesis is rooted in the multiple memory systems model considered in Chapter 2. It takes a dual route view of language processing as its starting point (Pinker, 1994, Squire, 2004). The procedural memory system is required for the learning of motor and cognitive skills, which include the perceptual-cognitive skills that make the fluent use of language possible (Ullman & Pierpont, 2005). Specifically, it is responsible for the learning of context-dependent sequential or probabilistically structured information. In language, the procedural system subserves the "mental Grammar", which is concerned with the rule-based procedures that govern the regularities of language (Ullman, 2004), combining temporally and hierarchically sequential or probabilistically structured information into complex representations (Christiansen & Chater, 2015) and is involved in the learning, storage and retrieval of the statistically regular, rule-based features of grammar and phonology. Procedural learning happens slowly and below the level of consciousness, as has been outlined in Chapter 2, but repetition over time leads to rapid, automatic processing and would account for the seemingly effortless retrieval of rule-based, grammar-related and phonological information required for fluency in language and reading. The hypothesis states that it is sequence-based implicit learning in particular that is implicated in language disorder (Nicolson & Fawcett, 2010).

By contrast (as outlined in Chapter 2), the declarative memory system is involved in the acquisition, storage and use of facts and events (Squire et al., 2004). In language, declarative learning occurs via associative binding of phonological or orthographical representations and meanings (Ullman, 2004). The procedural deficit hypothesis proposes that the relatively unimpaired lexical knowledge seen in these disorders suggests that declarative memory systems remain intact. Indeed, Ullman and Pierpont (2005) propose that over time the declarative memory system may compensate for procedural weakness (for a review, see Ullman & Pullman, 2015), since in normal learning both systems are brought to bear in order to learn (Poldrack & Packard, 2003), and work in competition in order to ensure learning is optimized (Foerde, Knowlton, & Poldrack (2007).

3.2 The neural basis for the procedural deficit hypothesis

There are a number of brain regions allocated to the procedural memory system that have been implicated in both language processing and language disorder.

3.2.1 The basal ganglia

There is evidence that the basal ganglia are involved in language processing, at phonological, morphological and syntactic levels, also in lexical selection and retrieval, as well as in higher-order language processing (Ullman, 2004). Evidence for basal ganglia involvement in language impairment comes from different avenues of research. For example, Parkinson's disease patients display more difficulty in marking regular versus irregular past tenses (Ullman & Pierpont, 2005). This is consistent with the claim that implicit memory is involved in the former, while declarative memory and the mental lexicon subserves the latter. However, it is also found in brain-imaging research showing differential activation of dyslexic participants and controls during implicit learning (Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006), as well as structural differences in the brain areas recruited to perform the implicit learning tasks (Menghini et al., 2008). Reduced overall volume of the basal ganglia is also found in developmental language disorder populations, as well as reduced grey matter mainly within the thalamus (Jernigan et al., 1991). Additionally, research into brain

functionality within the KE family, who display a range of symptoms associated with developmental language disorder, has shown both over-activation of the caudate nucleus (Vargha-Khadem et al., 1998), alongside under-activation of the putamen (Liégeois et al., 2002).

3.2.2 Broca's area

The basal ganglia project into the posterior Broca's area within the frontal cortex. Broca's area is also thought to be involved in the learning and manipulation of sequential information, which is a fundamental requirement of mastering language (Conway & Christiansen, 2001; Dominey, Hoen, Blanc, & Lelekov-Boissard, 2003) and maintaining verbal information, such as phonological sequences, in working memory (Smith & Jonides, 1999). It has been hypothesized that these two areas play somewhat different roles within the umbrella of procedural learning. While the basal ganglia play an important part in acquiring procedural and grammatical knowledge, Broca's area is involved in the use of such knowledge (Ullman, 2006). Although long-suspected, a recent imaging study has found evidence of connections between the anterior putamen (part of the dorsal striatum within the basal ganglia), the thalamus and both the anterior portion of Broca's area, which is involved in semantic processing and the posterior portion which is involved in phonology and syntax (Ford et al., 2013). Once again, the KE family display atypical structure and function of this area (for a review of studies of the KE family, see Ullman & Pierpont, 2005). Atypical connectivity between Broca's area and other brain areas has been shown in dyslexia. Brain-imaging in normal participants during phonological processing tasks showed activation of Broca's area alongside the temporo parietal cortex and insula, while in dyslexics, these tasks showed activation of Broca's area alone (Paulesu et al., 1996). Disconnection between Broca's area and the angular gyrus in dyslexia has also been mooted (Hampson et al., 2006), as well as greater symmetricality of both these areas in dyslexia (Habib, 2000).

3.2.3   The cerebellum

The role of the cerebellum in motor control is well-known (for a review, see Manto et al., 2012), but sequence-based cerebellar mechanisms are also crucial for the broad range of cognitive processes that subserve language (Marien et al., 2014), from processing the timing-dependent phonetic features required for speech perception (Keele & Ivry, 1990), to involvement in the articulatory processes needed for speech production, as well as for verbal working memory (as referred to in Chapter 1 as support for a cerebellar deficit as a cause of language disorder). It is involved in the processing of both receptive and expressive grammar (Justus, 2006), via the retrieval of implicitly learnt grammar rules (Ullman, 2001). It coordinates the many processes required for efficient word identification in reading from phonologic assembly to lexical-semantic processing (Vlachos, Papathanasiou, & Andreou, 2007), as well as being involved in the planning as well as the execution of writing.

Dysfunction of the cerebellum has also been suggested as a cause of both dyslexia and developmental language disorder (Nicolson, Fawcett, & Dean, 2001; Stoodley & Stein, 2013). The cerebellum is implicated in a number of processes that are involved in language (Marien, Engelborghs, Fabbro, & De Deyn, 2001) and especially in reading, from the direction of attention to error detection. At a physiological level, right-sided asymmetry of the cerebellum has been shown in typical readers, along with decreased right side volume in dyslexics (Leonard, et al., 2001) and overall decreased cerebellar volume related to oral and written language comprehension skills. Additionally, lesion studies have linked phonological processing deficits on a visual rhyme judgement task, as well as more variable performance on a non-word repetition task indexing phonological working memory, to the cerebellum (Ben-Yehudah & Fiez, 2008).

In summary, the procedural deficit hypothesis claims that the underlying cause of the deficits seen in developmental language disorder and dyslexia is an impairment of the procedural memory system, while declarative memory remains intact. At a neural level the hypothesis is persuasive, since brain structures allocated to the procedural memory system are involved in both language and literacy and atypical structure and

activation of these brain regions have been found in both disorders. As we shall now see, the procedural deficit hypothesis has also generated a substantial body of behavioural research using a variety of different implicit learning tasks. However, results of these behavioural investigations to date have been inconsistent.

3.3   Measures of implicit learning

Research has used a variety of tasks ranging from artificial grammar learning (Reber, 1967) to mirror-drawing (Vicari et al., 2005) with both child and adult samples. The wide range of implicit learning tasks used, as well as variations in the age, diagnosis and classification of participants make generalizing about the nature of procedural learning and its relationship with language and language disorder difficult.

The following figure (Figure 3.1) from Krishnan, Watkins, and Bishop (2017) illustrates the general pattern of findings for implicit learning paradigms used to investigate the procedural deficit hypothesis, with impaired learning found in disordered groups in the tasks used to test sequential procedural learning and probabilistic category learning. However, these results are far from consistent or conclusive. We shall now consider each of the main behavioural paradigms in more detail.
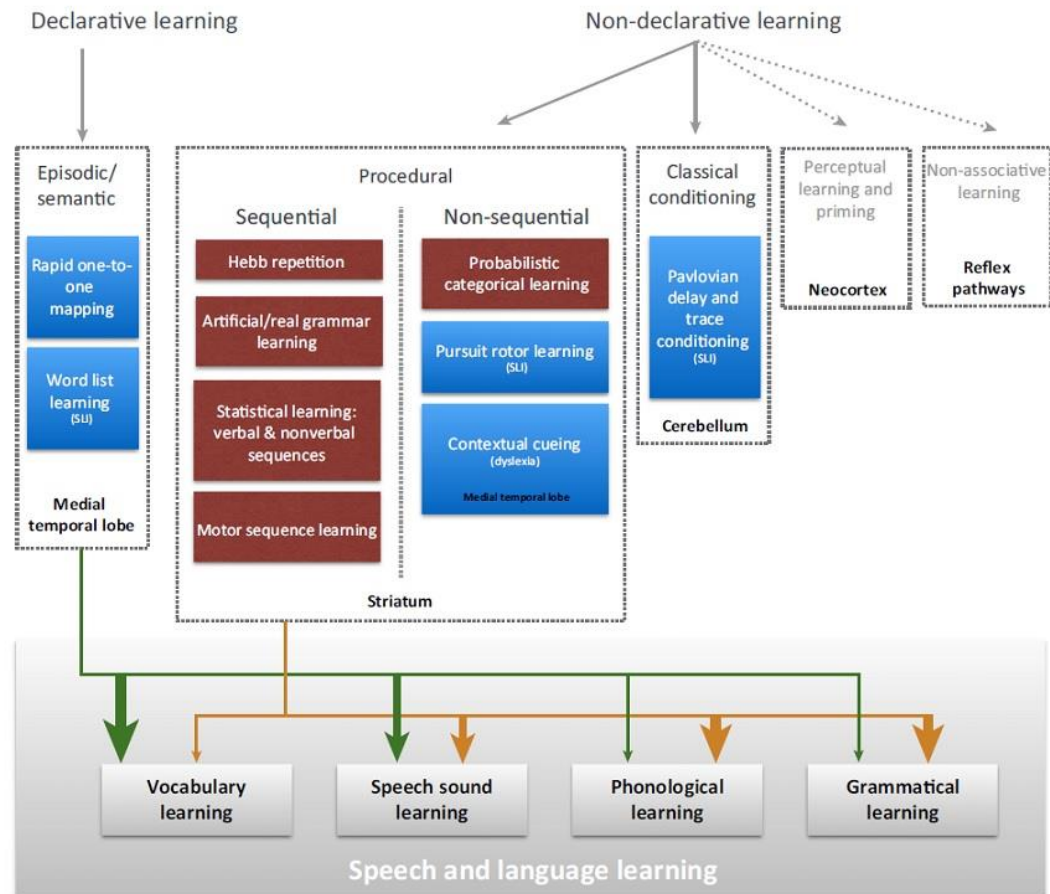
Figure 3.1 The contribution of learning and memory systems to language learning difficulties. Reprinted from Krishnan, Watkins, and Bishop (2017). The boxes below each memory system show the tasks that have most frequently been used to assess learning in children with developmental language disorder and dyslexia. The bottom panel highlights the hypothesized contribution of each learning system to language learning.

3.3.1   The serial reaction time task (SRT)

Support for the procedural deficit view comes, in the main, from impaired performance of language-disordered participants on implicit serial learning tasks, such as the serial reaction time task (Nissen & Bullemer, 1987). In the serial reaction time task participants respond to a stimulus appearing in 1 of 4 locations as fast as possible. Faster response times to trials that follow a covert sequence compared to random trials are taken as evidence of implicit learning (Seger, 1994). Successful implicit learning performance on the serial reaction time task has been shown to recruit the prefrontal

cortex, striatum and cerebellum, all structures implicated in the implicit learning circuitry of the brain (Pascual-Leone, 1993; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Laforce & Doyon, 2001).

The original deterministically structured serial reaction time task has been criticized for not fully dissociating implicit and explicit learning (Shanks & Johnstone, 1998; Shanks & St John, 1994). Shanks, Green & Kolodny (1994) demonstrated that unaware subjects performed above chance in post-task "generate" tests that tested awareness of the sequence by asking participants to predict future locations. Therefore, they could not be said to be properly unaware. More complex, probabilistically structured (Schvaneveldt & Gomez, 1998) or alternating versions (Howard & Howard, 1997) that render the sequence indiscernible have since been developed to minimize the risk of explicit learning. The proportion of unaware participants on these tasks is far higher, yet they still show significant implicit learning. For example, participants on a complex probabilistic serial reaction time task (Cleeremans & McClelland, 1991) spanning 60,000 trials over 20 sessions were shown to develop progressive sensitivity to the sequential dependencies within the task up to and including three elements of context, but developed limited reportable knowledge of the sequences. Although they were significantly better able to predict strings that followed the sequential structure of the task than those that did not during a post-task generation test, this effect was small.

Most research using the serial reaction time task to investigate the procedural deficit hypothesis uses extreme group designs of language-disordered and control participants. In these studies, language-disordered children have been shown to perform poorly both on deterministic serial reaction time tasks (developmental language disorder: Conti-Ramsden, Ullman, & Lum, 2015; Gabriel, Maillart, Stefaniak, Lejeune, Demottes, & Meulemans, 2013; Hsu & Bishop, 2014; Lum, Conti-Ramsden, Page, & Ullman, 2012; Lum, Gelgic, & Conti-Ramsden, 2010; Lukacs & Kemeny, 2014; Sengottuvel & Rao, 2013; 2014; Sengottuvel, Rao, & Bishop, 2016; Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011; dyslexia: Stoodley, Ray, Jack, & Stein, 2008; Vicari et al., 2005; Vicari, Marotta, Menghini, Molinari, & Petrosini,

64

2003) and on more complex alternating task versions (Desmottes, Meulemans, & Maillart, 2016a; 2016b; Hedenius, 2013; Howard, Howard, Japikse, & Eden, 2006). However, other papers do not support these results with null findings on deterministic tasks in children with developmental language disorder (Gabriel, Maillart, Guillaume, Stefaniak, & Meulemans, 2011; Gabriel, Meulemans, Parisse, & Maillart, 2015; Gabriel, Stefaniak, Maillart, Schmitz, & Meulemans, 2012; Lum & Bleses, 2012; Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014); and in children with dyslexia (Menghini et al., 2010; Vakil, Lowe, & Goldfus, 2015; Yang & Hong-Yan, 2011).

In adults with dyslexia the story is similarly inconsistent, with null findings on several deterministic tasks (Rüsseler, Gerth, & Münte, 2006; Kelly, Griffiths, & Frith, 2002; Laasonen et al., 2014), as well as a recent study using an alternating version of the serial reaction time task (Henderson & Warmington, 2017). However, several studies do find poorer implicit learning in dyslexic adults on deterministic tasks (Gabay et al., 2012a; Menghini et al., 2006; 2008; Stoodley, Harrison, & Stein, 2006) and on complex alternating versions (Howard, Howard, Japikse, & Eden, 2006). Only two studies examine procedural learning deficits in adults with developmental language disorder and both report null results (Lee & Tomblin, 2015; Lee, Mueller, & Tomblin, 2016).

Few correlational studies examining serial reaction time learning and language ability exist. Those that do are in children and all report a predominantly null result for the relationship, either between implicit learning and knowledge of past tense morphology or with receptive vocabulary on either the BPVS-II or the PPVT (Lum & Kidd, 2012; Kidd & Kirjavainen, 2011), as well as with reading ability (Waber et al., 2003). The lack of significant findings in large-scale correlational studies raises a red-flag about the legitimacy of significant findings in the extreme group studies, which will be explored further in Chapter 4.

3.3.1.1 Noise within the data

There are a number of possible explanations for the inconsistent results in the literature. Mayor-Dubois et al. (2014) found implicit learning impaired in children with developmental language disorder who were comorbid for developmental coordination disorder, but not for children with developmental language disorder alone. This suggests that impairments on the task may not be related to language difficulties, but to motor deficits instead. In line with this suggestion, Gabriel et al. (2012) reported that children with developmental language disorder made more learning-related errors than typically developing children when using a keyboard, but not a touch-screen, suggesting that implicit learning measures may be affected by task-specific response mechanisms.

3.3.1.2 Sequence complexity

It is also possible that differences between groups may be related to sequence complexity. Learning of first order conditional (FOC) sequences can be based on knowledge of a single preceding location, while second order conditional (SOC) sequences are more complex, with learning requiring knowledge of the two preceding sequence locations. Robertson (2007) suggested that any differences in implicit learning would relate purely to the greater computational complexity of SOC sequences, not the type of learning they engender. However, it is possible that FOC sequences are easier to learn explicitly (Curran, 1997) and that SOC sequences, therefore, may provide a purer measure of implicit learning.

In support of this, Kelly, Jahanshahi, & Dirnberger (2004) found Parkinson disease patients predominantly impaired for SOC not FOC sequence learning, concluding that FOC sequence learning is aided by declarative memory mechanisms. At a first glance the use of tasks with SOC sequences supports the procedural deficit hypothesis, with significant group differences found in dyslexic adults (Howard et al., 2006) and in children with developmental language disorder (Gabriel et al., 2013). However, studies that analyse FOC and SOC learning side by side are once again inconsistent. While Du & Kelly (2013) also found dyslexic adults impaired for higher order, not lower order, transitions within the same task, Clark & Lum (2017) found the opposite,

with children with developmental language disorder impaired on the FOC version only. Finally, Deroost et al. (2010) found dyslexic children unimpaired on both FOC and SOC versions of the same deterministic serial reaction time task.

3.3.1.3   Consolidation

The extent to which consolidation of implicit learning may explain the difference between language-disordered and control groups has also been examined. Two studies have suggested that the implicit learning impairment in developmental language disorder is confined to consolidation rather than acquisition of implicit learning (Desmottes, Meulemans, & Maillart, 2016a; Desmottes, Maillart, & Meulemans, 2017). Initial procedural learning was intact for both children with developmental language disorder and controls, but the developmental language disorder group were impaired during a second attempt at the same task. Similar results were also reported on an alternating serial reaction time task in dyslexic children (Hedenius et al., 2013). However, Gabay et al. (2012a) found the opposite, with dyslexic adults performing comparably with controls during later learning stages, but showing impaired initial learning, while Henderson & Warmington (2017) found no implicit learning in either group across initial and consolidation sessions of their task. Once again, therefore, the answer is not clear-cut.

3.3.2   Hebb serial order learning task

The most widely used measure of verbal implicit learning is the Hebb serial order learning task (Hebb, 1961). In this task participants are asked to recall sequences of items in order. Unbeknownst to them a covert repeating sequence is introduced. Better recall for the repeated, as opposed to the random sequences, is considered evidence of implicit learning. Just as in the serial reaction time task, memory for serial order is critical for performance of this task. The cognitive processes required to demonstrate a Hebb learning effect have been compared to those involved in word-form learning (Page & Norris, 2008), reflecting the learning of sequences of phonemes or syllables as a single unit. Such processes are thought to underlie infants' learning of new words, as they extract the statistical regularities from verbal input, in order to segment the speech stream going on around them (Saffran, Aslin, & Newport, 1996). Research in

typically developing participants has demonstrated Hebb learning of nonsense syllables, combined with inhibited rejection of non-words composed of sequences of these syllables, compared to randomly composed non-words in a subsequent lexical decision task (Szmalec, Duyck, Vandierendonck, Mata, & Page, 2009). The authors claimed that the Hebb repetition effect served as a laboratory analogue of naturalistic vocabulary acquisition, with nonsense syllables forming phonological lexical representations during the Hebb learning process.

There is a question mark over the extent to which the Hebb repetition effect relies solely on implicit learning processes, as participants can become aware that a sequence is repeating. Although it has been claimed that being aware of the presence of a repetition does not influence the extent of the Hebb repetition learning (McKelvie, 1987; Stadler, 1993), Weitz, O'Shea, Zook, and Needham (2011) found that participants who were aware of a repetition out-performed those who were not.

Poor performance on verbal Hebb learning tasks has been found in children with developmental language disorder (Hsu & Bishop, 2014) and in dyslexic adults (Szmalec, Loncke, Page, & Duyck, 2011; Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015). Szmalec et al (2011) also found the same dyslexic adults to be impaired on a non-verbal visuo-spatial Hebb task using sequences of dot locations, suggestive of a domain-general impairment. The domain-general finding is in line with research by Page, Cumming, Norris, Hitch, and McNeil (2006) that the Hebb effect operates across modalities. However, Staels & Van den Broek (2015) found no evidence of impaired serial order learning on a verbal Hebb task in dyslexic adolescents or children and Majerus et al. (2008) found no link between developmental language disorder status in children and performance on a Hebb task using familiar digits. So, it can be seen that, just as with the serial reaction time task, the results of research investigating procedural learning in language disorder using the Hebb serial order learning paradigm is mixed and firm conclusions are difficult to draw.

3.3.3 Artificial grammar and statistical learning

A detailed review of artificial grammar and statistical learning tasks is beyond the scope of this thesis as these tasks will not be used in the experimental studies reported in later chapters. Nevertheless a brief overview is important, as the results of previous research has a direct bearing on the claims of the procedural deficit hypothesis.

Reber developed artificial grammar learning (AGL: Reber, 1967) to examine the implicit learning processes that underpin language learning. He was the first to coin the term "implicit learning" and his focus was on the cognitive processes underlying predominantly verbal statistical learning, rather than the visuo-motor learning targeted by the serial reaction time task. The artificial grammar learning paradigm is a predominantly verbal task. It first presents participants with strings of stimuli that conform to an undisclosed set of combinatory rules (a finite state grammar). During a subsequent testing session participants are asked to judge whether new strings conform to or violate this grammar. The measure of implicit learning is the number of correct judgements made, with higher than chance performance related to the presence of implicit learning. Normal participants perform above chance (Reber, 1967) and artificial grammar learning has been shown to endure over time periods of up to two years (Allen & Reber, 1980), which is representative of the lasting nature of implicit knowledge. In one case performance on an artificial grammar task was still above chance (68%) two years after participants had been given only 10 – 15 minutes exposure (Allen & Reber, 1980).

The rules governing the grammars in artificial grammar tasks are usually complex and they are infrequently attempted with children. However, the statistical learning paradigm (Arciuli & Simpson, 2011) provides a less complex version of artificial grammar-type learning. Learning and testing phases are structured in the same way as in artificial grammar learning tasks, but the strings of stimuli conform to a simpler base triplet structure. Even babies have been shown to attend preferentially to legal sequences in this task (Saffran, Aslin & Newport, 1996)

However, not everyone agrees that above chance performance represents implicit learning of the grammatical rules of the task (Perruchet & Pacteau, 1990; Pothos, 2007). Johnstone, Shanks, Neely, and James (1999) have pointed out that learning on many artificial grammar learning tasks in humans can be attributed to the learning of surface features of the strings rather than underlying rules. A recent review of artificial grammar learning in animals has also shown grammatical test strings in auditory tasks are more acoustically similar to strings heard during the training phase than non-grammatical test strings, with the training string either completely contained within the test string, sharing chunks or the same beginning or simply more similar in sequence to the training strings (Beckers, Berwick, Okanoya, & Bolhuis, 2017). Since these studies typically use the same grammars that are used in studies with human participants, acoustic similarity should be considered a possible alternative explanation for implicit learning in auditory artificial grammar learning tasks in humans too. Chan (1992, unpublished thesis, as cited in Berry & Dienes, 1993) postulated that implicit learning happened to a greater degree when the stimuli in artificial grammar learning tasks were "unnameable" symbols, rather than ones that could be verbally labelled, such as letters, since verbal stimuli encouraged explicit learning. In the same vein, Andrade and Baddeley (2011) demonstrated that allowing verbal rehearsal during the encoding phase in artificial grammar learning improved legality judgements, as it engaged phonological short-term memory processes (which depend on language) to aid performance. Even if explicit learning in artificial grammar learning tasks is limited, it may be enough to support legality discrimination (Perruchet & Pacteau, 1990).

The majority of published research using artificial grammar and statistical learning tasks to investigate implicit learning in language disorders have found such learning to be impaired in both developmental language disorder and dyslexia. Artificial grammar learning and statistical learning tasks in language-disordered group designs are a mixture of verbal (mainly auditory) and non-verbal (mainly visual) tasks. Impaired artificial grammar learning has been demonstrated in children on auditory verbal tasks (Lukacs & Kemeny, 2014) and on non-verbal visual tasks using geometric shapes (Pavlidou, et al., 2009; 2010; Pavlidou & Williams, 2010). Although, Nigro,

70

Jiménez-Fernández, Simpson, & Defior (2016) found learning on a visual verbal and a non-verbal task to be equivalent for dyslexic and control participants, a more fine-grained analysis showed that the dyslexic participants did not perform as well as controls on transfer measures that required the learning of the abstract rules governing the task. Only one study in children found equivalent levels of learning in both groups (Plante, Bahl, Vance, & Gerken, 2010). This study examined the implicit learning of word-level stress patterns in children with and without developmental language disorder. However, it should be noted that normal adults were better at the same stress pattern task compared to adults with language disorder (Bahl, Plante, & Gerken, 2009). Studies using artificial grammar learning tasks in adults are more mixed. Significant impairment on visual verbal tasks have been found in language disorder (Plante, Gomez, & Gerken, 2002) and in dyslexia (Kahta & Schiff, 2016). However, it should be noted that dyslexic participants may struggle with verbal stimuli presented in a visual format, which adds a possible confound to visual tasks presenting letter strings. Even so, two other studies found equivalent learning for visually presented symbol strings, using a simple CVC grammar structure in adults with language disorder (Aguilar & Plante, 2014) and a more complex finite state Markovian rule system in adults with dyslexia (Rüsseler et al., 2006). Pothos and Kirk (2004) also found no difference between dyslexic and normal adults, using an embedded shapes artificial grammar that encouraged perception of the stimuli as a whole, but they did find a small difference on a task that presented symbols sequentially. They interpreted this as evidence for an implicit serial order learning impairment in dyslexia.

Results using statistical learning tasks predominantly find a significant impairment for groups with developmental language disorder or dyslexia. Evans, Saffran, and Robe-Torres (2009) have shown impaired statistical learning in children with developmental language disorder on a statistical learning task using the tri-syllabic speech stream stimuli from Saffran, Newport, Aslin, Tunick, and Barrueco (1997), as well as on a non-verbal version of the task using tone triads. Impaired statistical learning in children with developmental language disorder in tasks using tri-syllabic speech stream stimuli has also been found by others (Mainela-Arnold & Evans, 2014;

Mayor-Dubois et al., 2014), as well as impaired learning in dyslexic adults using both verbal and non-verbal stimuli (Gabay, Theissen, & Holt, 2015).

Correlational studies also point to a relationship between statistical learning and language. While Saffran, Aslin, & Newport's (1996) study of statistical learning in infants showed that babies as young as 8 months use transitional properties to segment speech, relatively few correlational studies use traditional artificial grammar or statistical learning tasks to examine the relationship between statistical learning and language ability in older participants. Those that do are in children and use statistical learning tasks, rather than artificial grammar tasks. These studies concluded that there was evidence of a relationship between statistical learning and syntax acquisition (Kidd & Arciuli, 2016); reading ability (Arciuli & Simpson, 2012) and language comprehension (Misyak & Christiansen, 2012). It should be noted that other studies exist that examine statistical learning using novel tasks that are based on sequences that use an artificial grammar structure. For example, see Spencer, Kashak, Jones, & Lonigen (2015) for a significant relationship in children between statistical learning and language using the Simon task or Misyak, Christiansen, & Tomblin (2010) for similar findings in adults using a novel combined serial reaction time and artificial grammar task.

In conclusion, evidence for impaired performance on statistical learning tasks in both children and adults with language disorder seems more clear-cut than for studies using the serial reaction time task. The picture is similar in studies using more complex artificial grammar in children, but is more mixed in adults. However, the general pattern in the published literature does seem to point to a domain-general implicit learning deficit for sequentially presented information in groups with both developmental language disorder and dyslexia.

3.3.4 Probabilistic category learning

Probabilistic category learning is evident during language acquisition, for example, as infants track statistical regularities in a continuous speech stream and discover words (Saffran, 2003). The most frequently used task of this genre is the

dichotic, decision-making weather prediction task (Knowlton, Squire, & Gluck, 1994), in which participants classify combinations of stimuli into one of two possible outcomes. Each stimulus is given a fixed probability of a certain outcome. Although participants are unaware of the probabilistic nature of the task, feedback on the accuracy of their judgment after each trial enables incremental implicit learning of the probabilities over time. A trial is scored correct if it accords with the conditional probabilities of the stimuli shown, regardless of the feedback given and above chance performance is taken as evidence of implicit learning.

Research using fMRI has shown that the task recruits the striatum, the caudate nucleus in particular (Poldrack et al., 1999). Other evidence for the implicit nature of learning on this task is that post-task self-report of even simple strategies for performance correspond poorly with actual performance (Gluck, Shohamy, & Myers, 2002), which suggests that probabilistic information is acquired in an unconscious way that is difficult to verbalize. However, others have claimed that the task requires explicit learning as participants test hypotheses about outcomes (Fotiadis, 2013). Certainly experimental manipulations of the weather prediction task that would favour explicit processing result in higher accuracy.

Reber (2013) suggests this type of task might pit implicit and explicit learning mechanisms against one another, as the task requires "aggregation of outcomes over multiple trials" (p.2032). The probabilistic nature of the task means that occasionally the outcome is the opposite of the one predicted by the stimuli and over-reliance on explicit memory could mean that these outcomes are over-weighted thus impairing performance overall (Shoshamy, Myers, Kalnithi, & Gluck, 2008)

There is a small body of research into the procedural deficit hypothesis using this task, which has shown that children with developmental language disorder are less accurate and learning develops later compared to controls (Kemény & Lukács, 2010; Lee et al., 2016). The same pattern has been found in adults with dyslexia (Gabay et al., 2015).

3.3.5 Contextual cueing

While research into the relationship between implicit serial order learning and language is comparatively plentiful, there is little research investigating how language ability relates to other aspects of implicit learning, such as visual search efficiency. Although dyslexia is a linguistic, not a visual impairment (Ramus, Pidgeon, & Frith, 2003; Vellutino, 1979), impaired visuo-spatial attention has been posited as a potential cause of the reading deficits seen in dyslexia (Vidyasagar & Pammer, 2010), has been found in pre-schoolers at risk of dyslexia (Facoetti et al., 2010) and impaired serial visual search and spatial cueing facilitation have been found in preschoolers who went on be poor readers (Franceschini, Gori, Ruffino, Pedrolli, & Facoetti, 2012).

Moving away from the implicit learning of sequenced or probabilistically structured information, the contextual cueing paradigm (Chun & Jiang, 1998) investigates the implicit statistical learning that develops to detect contextual regularities during visual search (Goujon, Didierjean, & Thorpe, 2015). Participants identify the location of a target stimulus within matrices of distractor stimuli. Unbeknownst to participants the target position in some matrices is predictable and faster response times to these compared to unpredictable matrices is considered evidence of implicit learning.

There are several explanations for the mechanisms involved in contextual cueing. One limits the learning in contextual cueing to the extraction and formation of small chunks of information as perceptual units (Chase & Simon, 1973). Certainly the same learning on contextual cueing tasks has been found when matrices repeat only the locations surrounding the target, as when the entire matrix is repeated (Brady & Chun, 2007). Another explanation claims contextual cueing relies on the learning of repeated global configurations of all items. This would imply associative learning between the configurations and target location to create an integrated representation to guide attention in a top-down fashion (Goujon, 2015). In support of this explanation, search efficiency improves when the task is made more difficult (Kunar, Flusberg, & Wolfe, 2006). It is thought entirely possible that both mechanisms are at work in contextual cueing (Goujon, 2015).

Much of the performance on contextual cueing appears to be under an objective level of awareness (Chun & Jiang, 2003). Indeed, giving participants explicit instructions to try to learn the regularities within the task does not tend to help performance. Additionally, the learning has also been shown to be extremely robust (Zellin, von Mühlenen, Müller, & Conci, 2014). It is, therefore, broadly agreed that learning on contextual cueing tasks is implicit learning. However, "contamination" by explicit knowledge has also been shown to occur (Smyth & Shanks, 2008), while damage to the medial temporal lobe system has been found to impair contextual cueing performance (Chun & Phelps, 1999; Manns & Squire, 2001), so once again, it seems likely that implicit learning on the task is not process pure.

Studies using contextual cueing to investigate the procedural deficit hypothesis have so far not found impaired performance in dyslexic adults (Howard et al., 2006; Bennett, Romano, Howard, & Howard, 2008) or children (Jiménez-Fernández et al., 2011), although impaired implicit sequence learning using the serial reaction time task was found in these same participants.

To sum up, there does seem to be evidence of an implicit sequence learning deficit in participants with language disorders. This impairment appears to be found across tasks that index non-verbal learning (mainly serial reaction time tasks and some statistical learning and artificial grammar learning tasks) and in verbal tasks (mainly Hebb serial order learning and auditory statistical learning and artificial grammar tasks), which would suggest it is domain-general in nature. This impairment is also found in probabilistic categorization tasks, but does not seem to extend beyond sequence specific or probabilistically structured tasks, as reflected by non-significant findings in contextual cueing, as well as in an embedded shapes version of an artificial grammar task. However, by no means all studies find a significant difference in performance between disordered and normal participants. Inconsistent results seem to be a particular problem in studies using the serial reaction time task, but it is not known whether this points to specific issues with this task or other factors such as perhaps publication bias for studies using the other paradigms.

3.3.6 Different metrics to measure implicit learning

Different paradigms measure implicit learning in different ways and these measurements may shape the conclusions that are drawn about the nature of implicit learning. Each task discussed above is associated with its own set of methodological limitations and two such limitations are briefly discussed here. Firstly, use of an explicit offline "testing phase" measure of implicit learning, such as that used by artificial grammar and statistical learning tasks requires a level of meta-cognitive awareness to index prior implicit learning. This introduces the possibility that declarative processes are included within the learning measure. Additionally, the offline nature of the test phase (ie: it takes place after learning) may have an effect on the learning itself, as well as making it impossible to measure the temporal trajectory of the learning over the course of the task. The matrix below (Figure 3.2) arranges the tasks by implicit or explicit, as well as online or offline measure of learning with the arguably preferable measurement (implicit and online) in the top left hand quadrant. As a final note about the offline measure for artificial grammar and statistical learning tasks in particular, the two alternate forced choice test (2AFC) that provides the learning measure for these tasks has been criticized as being potentially insensitive to individual differences (Siegelman, Bogaerts, & Frost, 2017).

	Implicit test of implicit learning	Explicit test of implicit learning
Online test of implicit learning	Serial reaction time Hebb serial learning Contextual cueing[1] Weather prediction	
Offline test of implicit learning		Artificial Grammar Statistical learning

Figure 3.2 Categorisation of tasks based on methodology: Online and offline testing

The second potential limitation relates to the way the implicit learning is represented, either as a direct or an indirect measure. The former measures are total correct scores, such as the correct or incorrect judgment of outcome accorded to each trial in the weather prediction task, but the latter type of task uses a difference score between performance on two conditions as the measure of implicit learning. The first condition does not require implicit learning and serves as a baseline score. Performance on the second condition arguably does benefit from implicit learning. The difference between the conditions is, therefore, taken as a representation of implicit learning. However, difference scores are known to be unreliable (Lord, 1958).

---

[1] Learning in the contextual cueing task can be measured continuously or can use a separate testing phase. The latter format was selected in the first study (Chapter 5), which would site the contextual cueing task in this thesis in the bottom left, rather than the top left, quadrant of the matrix.

This concern will be re-visited in later experimental chapters. The matrix below (Figure 3.3) arranges the paradigms according to this methodological limitation.

	Implicit test of implicit learning	Explicit test of implicit learning
Total correct Score to represent implicit learning	Weather prediction	Artificial Grammar Statistical learning
Difference Score to represent implicit learning	Serial reaction time Hebb serial learning Contextual cueing	

Figure 3.3 Categorisation of tasks based on methodology: Scoring methods

It can be noted that both matrices site artificial grammar learning and statistical learning tasks on the right hand side of the matrices, as a result of the explicit nature of the implicit learning testing phase. In this thesis paradigms that assess implicit learning without the use of explicit recognition measures are preferred for the reasons cited above.

To reiterate, there is a growing body of behavioural research that suggests that impaired procedural learning may be an underlying cognitive cause of the symptoms seen in both developmental language disorder and dyslexia. These studies employ a range of implicit learning paradigms from the serial reaction time task to the weather

prediction task. However, the results of this literature are inconsistent, with a substantial number of studies finding no link between procedural learning and performance on these tasks. Methodological differences between studies, such as sequence complexity, for example, may explain some of this inconsistency.

3.4 Declarative memory and the distinction between short-term and long-term memory

It is also important to examine the other main claim made by the procedural deficit hypothesis that declarative memory remains intact in the disorders. To recap, declarative memory is the recollection for personal events or for facts that can be intentionally retrieved (Baddeley, Eysenck, & Anderson, 2009). Declarative learning in language-related research has frequently been measured using free recall and serial recall tasks (Baddeley, 2003; Vellutino & Scanlon, 1985). Impaired free recall performance has been found in dyslexia (e.g., Vellutino & Scanlon, 1985; Menghini, Carlesimo, Marotta, Finzi, & Vicari, 2010; Kibby & Cohen, 2008), as well as impaired explicit serial recall performance in dyslexic adults and children (Brosnan et al., 2002; Perez et al., 2012). Menghini et al. and Kibby & Cohen (op cit) further distinguish between worse verbal, but not non-verbal free recall performance in poor readers compared to controls, while Kibby (2009) distinguishes between impaired verbal short-term, but not long-term memory in dyslexia. Lum et al. (2012) argue that many of the findings of impaired explicit learning in language-disordered children may be due to their phonological processing problems and the verbal nature of the tasks themselves, rather than explicit memory per se. A meta-analysis of 53 studies investigating verbal short-term memory in children with dyslexia (Melby-Lervag, Lyster, & Hulme, 2012) found a large significant mean effect size ($d = -0.71$, 95% CI [-0.83, -0.60]), with dyslexic children performing more poorly on tasks than typically developing age-matched children. Oral language and phonemic awareness accounted for 48% of this difference. The meta-analysis also included 24 studies comparing dyslexic children with reading-level control groups, showing a non-significant overall mean effect size ($d = -0.09$, 95% CI [-0.28, 0.10]). The review concluded that verbal short-term memory and efficient reading both depend on establishing phonemically structured phonological representations. In other words, verbal short-term memory

correlates with reading ability, because they both rely on access to the same kind of phonological information (McDougall, Hulme, Ellis, & Monk, 1994; Hulme & Roodenrys, 1995).

The picture is similar in studies of short-term memory and developmental language disorder. Archibald and Gathercole (2006) report impaired verbal, but not non-verbal short-term memory in children with developmental language disorder. Much of the evidence for poor verbal short-term memory comes from tasks using non-word repetition (e.g., Gray, 2003; Conti-Ramsden, Botting, & Faragher, 2001; Norbury, Bishop, & Briscoe, 2002), but recall and serial recall tasks have also been used. Children with developmental language disorder perform more poorly than controls on verbal serial recall tasks (e.g., Hick, Botting, & Conti-Ramsden, 2005; Mainela-Arnold & Evans, 2005), and similar results have also been found on verbal free recall tasks (e.g., Kail, Hale, Leonard, & Nippold, 1984; Kirchner & Klatzky, 1985; Nichols et al., 2004) Results for non-verbal short-term memory in developmental language disorder are scarce, but significant impairment in developmental language disorder on non-verbal free recall tasks for patterns have been found (e.g., Hick, Botting, & Conti-Ramsden, 2005). Mainela-Arnold and Evans (2005) related the poorer performance on the complex verbal serial order recall task in their study (CLPT: Gaulin & Campbell, 1994) to degraded linguistic representations in developmental language disorder.

Avoiding possible task-related confounds, several studies have shown intact performance on explicit versions, but impaired performance on implicit versions of the same sequence learning tasks in dyslexic participants (Jiminez-Fernández et al., 2011; Vicari et al., 2003). However, Staels and Van den Broek (2017) did find their dyslexic children performed worse on an explicit version of the serial reaction time task, while learning on an implicit version was only poorer for the dyslexic group during the initial learning on the task and was equivalent for the principal measures of implicit learning on the task (an RT increase on transfer to a random sequence or an RT rebound on returning to the sequence). As such, the role explicit memory skills play in language disorder is still under debate, although the bulk of evidence does suggest poorer declarative learning in the disorders.

The procedural deficit hypothesis explains findings of impaired verbal declarative memory in the disorders using traditional declarative memory tasks, by relating verbal declarative deficits to impaired verbal working memory, which they distinguish from declarative memory and site within the procedural memory system (Ullman, 2004; Lum et al., 2012; Conti-Ramsden, Ullman & Lum, 2015). For example, in Conti-Ramsden et al's (2015) study, children with developmental language disorder displayed poorer performance on verbal declarative tasks from the Children's Memory Scale (Cohen, 1997: subtests indexing recall and recognition of semantically unrelated word pairs) than a typically developing group. However, once their verbal working memory capacity had been controlled for, using listening, counting and backwards digits recall scores, this difference disappeared.

The division of memory along temporal lines is another aspect of the multiple memory systems view, which partitions working memory separately from declarative memory and sites it outside of the medial temporal lobe in their taxonomy. Working memory is the ability to maintain and manipulate information in the mind to support other cognitive functions such as learning and reasoning (Baddeley, 2003). Visual working memory is very limited, typically limited to maintenance of three or four visual objects at a time (Fukuda, Awh, & Vogel, 2010). Verbal working memory is a little more expansive, possibly due to the potential for verbal rehearsal to maintain items (Squire & Dede, 2015). However, it is debatable whether the short-term memory recall and recognition measures from the Children's Memory Scale and the digit recall scores used to control for working memory in Conti-Ramsden et al. (2015) actually index separable memory mechanisms and using one to control for the other is a controversial practice at the least.

3.4.1 Neural evidence

There is evidence that areas of the brain active during procedural learning are also active during tasks designed to index working memory. Cabeza and Nyberg's (2000a) extensive review of the PET and fMRI literature implicates prefrontal and parietal areas of the brain as important for performance on working memory tasks. Links to the cerebellum have been suggested in some research, showing activation in the

81

cerebellum on phonological processing tasks that also recruit Broca's area, while working memory for sequences involves basal ganglia, thalamic and cerebellar regions among others. However, brain imaging studies have also shown that working memory tasks cause activation in medial temporal lobe (for reviews, see Graham, Barense, & Lee, 2010; Ranganath & Blumenfeld, 2005). For example, Olson, Page, Moore, Chatterjee, and Verfaellie (2006) demonstrated the hippocampus was required for working memory of object-location conjunctions, just as it is for long-term memory. Brain imaging has also revealed a shared neural substrate for both episodic and working memory in the fronto-parietal-cerebellar network (Cabeza & Nyberg, 2000b), with differential activity during working memory tasks in two parietal subregions: Broca's area in the left hemisphere and posterior/ventral parietal areas bilaterally. These authors speculated that the former reflected phonological aspects of working memory operations specifically, with the latter allocated to more general working memory operations.

3.4.2 Behavioural evidence

At a behavioural level, although there is evidence that a better working memory capacity may make implicit sequence learning easier (Howard & Howard, 1997) most studies have not found a link between working memory and implicit sequence learning (Janacsek & Nemeth, 2013). For example, they found no correlation between working memory and serial reaction time task performance (Bo, Jennett, & Seidler, 2012; Feldman et al., 1995), while Kaufman et al. (2010) went a step further, using a correlational design and structural equation modelling to demonstrate no link between procedural learning on a probabilistic serial reaction time task and working memory. Working memory has been linked to performance on explicit tasks, however, possibly acting to guide attentional focus and cognitive control (Jiménez, 2003; Kaufman et al., 2010). This would suggest that verbal working memory is more closely aligned with declarative than procedural learning processes.

3.4.3   The processing view of short-term and long-term memory

The debate surrounding the nature of working memory goes a step further than this, arguing that verbal working memory and verbal declarative memory are not distinct constructs and there is no justification for allocating them to neurally distinct stores in the brain. By extension, tasks reputed to index verbal working memory and verbal declarative memory respectively do not index distinct and different processes in the brain (Belleville, Caza, & Peretz, 2003). These arguments mirror the processing-based positions in the debate surrounding the nature of declarative and procedural learning reviewed in Chapter 2. For example, Belleville, Caza, and Peretz (2003) provide a process-based explanation for the apparent dissociations between working memory and declarative memory found in brain-damaged patients that relies instead on a distinction between the processing of verbal and non-verbal modalities. They studied a patient with extensive neural damage that appeared to cause an isolated short-term memory deficit, leaving long-term memory intact. Memory testing with traditional verbal and non-verbal short-term immediate serial recall and long-term supraspan tasks supported this conventional interpretation. However, closer examination revealed an alternative interpretation of the impairment was possible. This more fine-grained investigation linked the apparent short-term memory impairment to a deficit in phonological coding that was evident for both short-term and long-term memory, alongside preserved non-phonological semantic processing, again evident in both short- and long-term memory. The impairment was, therefore, better described by a distinction between the type of processing required for phonological and non-phonological semantic information respectively, rather than by the traditional short-term long-term memory system distinction. The apparent dissociations between short- and long-term memory, therefore, could be explained by the task demands of the different tasks used for each type of memory. This interpretation is consistent with claims that verbal short-term memory and language (perception and production) representations are closely related (Allen & Hulme, 2006; Hulme et al., 1997; Martin & Saffran, 1997; Walker & Hulme 1999) and this is the reason that deficits in short-term memory parallel deficits in language processing.

In summary, the first claim of the procedural deficit hypothesis is that deficits in the procedural memory system are behind the pattern of impairments seen in both developmental language disorder and dyslexia. As we have seen, research investigating this claim has reported inconsistent results to date. The other main claim of the hypothesis is that declarative learning in developmental language disorder and dyslexia remains intact. However, there is much previous research that demonstrates impaired declarative learning in these disorders, which is found on free recall and serial recall tasks, as well as on non-word repetition tasks. These impairments are found predominantly on verbal declarative learning tasks. The procedural deficit hypothesis explains this apparent contradiction by linking deficits on declarative learning tasks to impaired domain-general working memory, which it separates from declarative memory, and sites within the procedural memory system. However, it is not clear to what extent working memory processes can be separated from short-term declarative memory processes; or to what extent working memory can be allocated to either declarative or procedural circuits in the brain. Importantly, there is also a question about the extent to which short-term verbal declarative processes can be separated from language itself.

The range of populations sampled and paradigms employed, as well as methodological differences across studies makes it difficult to draw any definitive conclusions about the veracity of the procedural deficit hypothesis. We shall now proceed to a more data-driven analysis of the existing literature, using meta-analysis to shed light on the claim that impaired procedural learning is the underlying cause of developmental disorders of language learning.

Chapter 4   Meta-analyses

4.1   Previous reviews of the procedural deficit hypothesis

The following chapter reports the results of a series of meta-analyses of previous studies of the procedural deficit hypothesis of language learning disorders. Five relevant meta-analyses currently exist. All limit their purview to a single language disorder and exclude paradigms that are relevant to the debate. Lum and Conti-Ramsden (2013) conducted a meta-analysis on studies of both declarative and procedural learning in children with developmental language disorder, including published peer-reviewed studies up to December 2012. Procedural learning was assessed with two studies of auditory statistical learning and two studies of probabilistic category learning. The review concluded that there was robust evidence for a deficit in children with developmental language disorder compared to age-matched controls on statistical learning tasks (mean effect size $d = .834$, 95% CI [.42, 1.25], $p < .001$), but less conclusive, evidence for a deficit on probabilistic category learning (mean effect size $d = .502$, 95% CI [.08, .93], $p = .02$). However, given the small number of implicit learning studies ($n = 4$) included in the analysis, these results must be interpreted with caution.

Two subsequent meta-analyses by the same research group focused only on the most frequently used paradigm in the literature, the serial reaction time task. Each of the analyses restricted itself to papers with participants categorized as having one specific language disorder diagnosis only, either dyslexia (Lum, Ullman, & Conti-Ramsden, 2013) or developmental language disorder (Lum, Conti-Ramsden, Morgan, & Ullman, 2014). Only papers using independent group designs were included. The first of these meta-analyses (Lum, Ullman, & Conti-Ramsden, 2013) included 14 studies investigating implicit learning in serial reaction time tasks in developmental dyslexia, finding significantly impaired performance in dyslexic compared to age-matched control groups (mean effect size $d = .449$, 95% CI [.204, .693], $p < .001$). The analysis examined the effect of sequence complexity (ie: use of first order conditional or second order conditional sequences), number of sequence repetitions on the serial reaction time task and age. It found smaller effect sizes with complex sequences or

more repetitions in studies with adults, but not with children. The second meta-analysis of eight serial reaction time studies (Lum, Conti-Ramsden, Morgan, & Ullman, 2014) also found implicit learning deficits in children with developmental language disorder (mean effect size $d = .328$, 95% CI [.071, .584], $p = .01$). Once again there were smaller effect sizes for more complex or longer tasks, but only in older children. However, both meta-analyses only quantified sequences as second order conditional if every first order transition occurred equally often (1,2; 1,3; 1,4; 2,1; 2,3; 2,4; 3,1; 3,2; 3,4; 4,1; 4,2; 4,3). Ten item sequences that inevitably contained two locations that could only be followed by two out of the three remaining locations did not qualify for a second order conditional label.

A recent meta-analysis (Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016) examined procedural learning across several implicit learning paradigms in participants with developmental language disorder, as well as those with autism. In the developmental language disorder meta-analysis eleven serial reaction time tasks, two artificial grammar and statistical learning tasks and two probabilistic category learning tasks were included in a single meta-analysis, in spite of many task differences. No Hebb serial learning paradigms were included. The meta-analysis concluded that procedural learning is significantly impaired in developmental language disorder compared to age-matched controls (mean effect size $d = .498$, 95% CI [0.28, 0.72], $p < .001$), but not autism. It also found that procedural learning impairments in developmental language disorder were not age-dependent, contrary to the findings of the previous meta-analysis of developmental language disorder and procedural learning in serial reaction time tasks (Lum et al., 2014).

A final meta-analysis by Schmalz, Altoe, and Mulatti (2016) examined statistical learning in dyslexia, including studies using artificial grammar learning tasks. They included eight studies in the full analysis, but expressed concern about lab effects in the four studies with children and about under-powered studies generally. Although the effect size for this meta-analysis was $d = .47$, 95% CI [.04, .90], they concluded that the true effect size was likely to be small. An attempted meta-analysis for studies using SRT tasks was abandoned, owing to the variety of dependent variables used in

different studies and the scarcity of reported condition averages from which to gauge effect sizes.

The series of meta-analyses reported here takes a wider view of implicit learning and language than has been attempted in previous meta-analyses. In particular, each meta-analysis in the series includes studies of participants with developmental language disorder, as well as those with dyslexia. While previous meta-analyses have confined themselves to a single disorder, this exclusivity is questionable, given the heterogeneity of symptoms in language disorders, their frequent comorbidity, the fluid nature of language development itself and the range of diagnostic tests employed to classify participants into groups of either disorder for experimental purposes (see Chapter 1). For example, in one study (Hedenius et al., 2013) participants categorized as dyslexic, displayed scores on a test of receptive grammar (TROG: Bishop, 1982) that were on average 17 points lower than the typically developing group. The TROG test is frequently used diagnostically for developmental language disorder. Importantly, the procedural deficit hypothesis claims that deficits in the procedural memory system are the basis of the impairments seen in both disorders. Examining the moderating influence of disorder type would help to clarify the extent to which the symptoms of dyslexia and developmental language disorder may emerge from common procedural mechanisms.

The series of meta-analyses include the full range of implicit learning tasks most commonly used in this area of research: serial reaction time tasks, Hebb serial order learning tasks, artificial grammar and statistical learning tasks and probabilistic category learning tasks. Unfortunately, there were not enough studies with contextual cueing tasks ($n = 3$) to warrant a meta-analysis. However, no differences between language-disordered and control participants have been found on contextual cueing tasks in any of the studies. Each paradigm is allocated to a separate meta-analysis to take account of the variability arising from the use of different tasks. Where possible, the analyses systematically examine the effect of moderating factors, ranging from participant variables, such as age, to task variables such as stimuli modality. The aim of this process is to gain a more accurate picture of any relationship between a

procedural learning system and language disorder, within one consistently conducted review.

In particular, the series of meta-analyses aims to address the following questions:

- Whether there is evidence for a group deficit on procedural learning tasks in groups with developmental language disorder and groups with developmental dyslexia
- Whether any procedural learning deficit is domain-general, as has been asserted.
- To what extent the relationship between implicit learning and language disorders might be moderated by age.
- Whether developmental language disorder and dyslexia appear to be differentially impaired on measures of procedural learning
- Whether there is any variability in the severity of the impairment in language disordered groups across studies and, if so, whether moderators of the size of the deficit can be identified
- To what extent any relationship between procedural learning and language learning disorders is confined to tasks that measure implicit sequence learning, rather than, for example, probabilistic categorisation.
- To examine the effect that within-paradigm differences, such as task length, may have on experimental results.

4.2   Methodology and inclusion criteria

To be included in the meta-analyses reported here, studies needed to be primary studies reporting on either group or correlational experimental designs and published in a peer reviewed journal. Group designs needed to compare implicit learning performance of children or adults with developmental language disorder or dyslexia with performance of a control group(s). Correlational designs needed to measure the relationship between performance on implicit learning tasks and language measures in individual participants. Given the differences inherent between group and correlational designs, these two types of study were entered into separate analyses.

Eligible studies needed to report data on at least one implicit learning task (serial reaction time; Hebb serial order learning; artificial grammar or statistical learning; probabilistic categorization; contextual cueing). Group studies needed to include means and standard deviations for performance on the tasks to enable a measure of difference between groups to be calculated. Alternatively, correlational studies had to include a measure of effect size (r) for the relationship between implicit learning and language. However, in practice many of the eligible studies did not include sufficient information. For serial reaction time tasks, for example, means and standard deviations for the task were usually reported in figure format only. Whenever the relevant information was implied but not reported, the study was considered eligible in the first instance and the authors were contacted and additional data requested. Tables in each meta-analysis include all eligible studies, although only studies for which sufficient data for an effect size was forthcoming were included in the final meta-analyses and this is indicated in the table that accompanies each meta-analysis.

4.3 Search Strategy

Studies were collected using a variety of approaches (see flow diagram in Figure 4.1 for the search criteria). Primarily, studies were located using the following electronic bibliographic databases: Medline, PsychInfo, Web of Science, ERIC ProQuest and Google Scholar. The search strategy combined terms relating to implicit learning with terms relating to language and language disorder and was developed in collaboration with subject specialist librarians at UCL, London. The search terms for Medline and PsychInfo is available in Appendix A. This search was adapted for use with the other bibliographic databases as necessary, but the search terms themselves were the same across all databases. Additional measures were taken to ensure all eligible studies were found. All previous reviews were checked (Lum et al., 2013; Lum et al., 2014; Obeid et al., 2016; Schmalz et al., 2016). A manual review of the table of contents for four key journals was conducted (Developmental Science; Journal of Experimental Psychology: Learning, Memory & Cognition; Research in Developmental Disabilities; Annals of Dyslexia). Finally, reference lists in short-listed papers were also checked.

Results were imported from their respective databases to Endnote. Duplicate references were detected and removed using Endnote's de-duplication search. An initial screening of titles and abstracts of the remaining references checked that the study was a primary experimental study; that it included measurement of participants on one of the candidate implicit learning tasks; and that it included an experimental group of language-disordered participants or related task performance to measures of language ability. A second stage screening evaluated the relevance of the full texts according to the inclusion criteria. The electronic database searches were re-run just before the final analyses (1st April, 2017) and three further studies published in the interim were retrieved for inclusion.

Search

Search Features:
- Electronic databases (ERIC, Medline, PsychInfo, Web of Science, ProQuest Theses and Dissertations with keywords related to implicit learning and language (see appendix x for search terms)
- Google scholar
- Search of prior systematic reviews and meta-analyses
- Scanning of reference lists
- Citation search of authors' names

Screening

Records after duplicates removed: (n = 1374)

Inclusion Criteria

Studies must include:
- One or more of the following implicit learning paradigms: Serial reaction time; Hebb serial learning; artificial grammar learning; statistical learning; contextual cueing; probabilistic category learning.
- Measures of language ability (language and / or literacy)
- Either child or adult participants.
- Data that enables an effect size to be calculated for the relationship between implicit learning and language.

Eligibility

Abstracts screened: (n = 1374)

Abstracts excluded for not meeting research criteria: (n = 1287)
- Focus on different disorders (eg: aphasia; autism; dementia; depression; Down syndrome; Parkinson's disease; schizophrenia)
- Focus on 2nd language learning
- Case studies
- Different implicit learning tasks (eg: lexical decision; fast mapping; grapho-motor symbol acquisition)

Full text articles assessed for eligibility: (n = 86)

Screening re-run before final analysis (n = 3)

Full text articles excluded (n = 7)
- Wrong participant groups (Cherry & Stadler, 1995; De Visscher et al., 2015; Folia et al., 2008; Ise et al., 2012)
- Paradigm sufficiently different to warrant exclusion (eg: Roodenrys & Dunn, 2008; Sperling et al., 2004)
- Focus on IQ, not language ability (eg: Gebauer et al., 2007)

Included

Studies included in meta-analyses (n = 82*)

*Some group design studies include more than 1 independent group comparison and multiple paradigms

Figure 4.1 Flow diagram for the search and inclusion criteria for studies in this series or meta-analyses. Adapted from "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement", by Moher, Liberati, Tetzlaff, Altman, and the PRISMA Group (2009).

4.4   Procedure

Six separate meta-analyses were conducted (Table 4.1). Data for meta-analyses on each of the main implicit learning paradigms were coded and analysed using the Comprehensive Meta-Analysis programme (CMA: Borenstein, Hedge, Higgins, & Rothstein, 2005). Unless otherwise highlighted other meta-analyses followed the same procedures. The standardised mean difference between groups was coded for group designs and the correlation between implicit learning and language measures was coded for correlational studies. These were entered in random effects models.

Studies that shared authors, had equal number of participants, reported the same results or use the identical task measures were further investigated to limit the risk of coding the same data twice. Duplicate reports of the same study were treated as one collective report.

Table 4.1 Series of meta-analyses with numbers of eligible studies and final inclusion numbers

Meta-analysis	Experimental Design	Eligible studies*
Serial reaction time task	Group	46 (21)
Serial reaction time task	Correlational	6 (5)
Artificial grammar learning and statistical learning tasks	Group	17 (14)
Artificial grammar learning and statistical learning tasks	Correlational	4 (3)
Hebb serial learning task	Group	9 (8)
Probabilistic category learning tasks	Group	6 (5)
Contextual cueing task	Group	Insufficient studies[1]

*Number of studies considered eligible. The final number included in the meta-analysis is noted in brackets. [1]Three studies were eligible, but only two contained sufficient data for inclusion.

4.4.1.1   *Effect size calculation for meta-analyses of group design studies*

The standardised mean difference in procedural learning between groups was coded for group designs, using Hedges' *g* (Hedges, 1981). When Hedge's *g* is

negative, the language disordered group is performing more poorly than the control group.

Effect sizes for the group difference in procedural learning on serial reaction time and Hebb task meta-analyses were calculated from the means and standard deviations of the component conditions on the tasks (SRT task: means and standard deviations in millisecond for sequenced trials and for random trials; Hebb task: means and standard deviations as a percentage score for repeated and for non-repeated trials). This method of calculating an effect size is directly analogous to the way effect sizes are calculated for randomised control trials, otherwise known as pre- post-test control group designs (Lund, 1988; Morris, 2008; Ray & Shadish, 1996). This is the experimental design that these extreme groups studies most closely resemble structurally (see Figure 4.2).
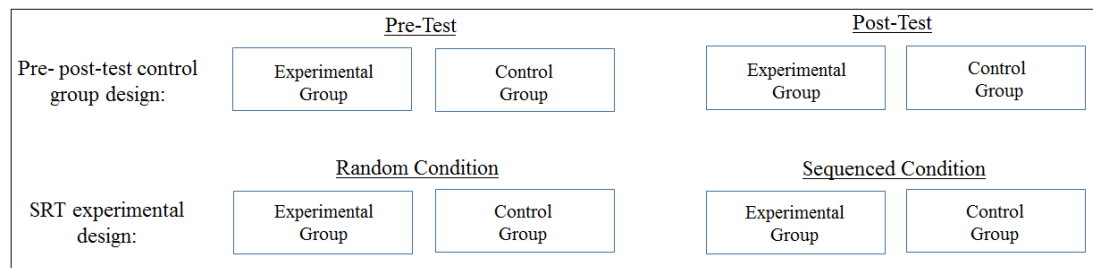
Figure 4.2 Analagous design structure for randomised control trials and implicit learning tasks based on difference scores.

Effect sizes were calculated using standard deviations for the random trial condition for serial reaction time tasks and the unrepeated sequence condition for the Hebb tasks, as this is equivalent to using the pre-test standard deviations recommended by Morris (2008). The standard deviation for the control group only was used to standardize effect sizes, as this gives a change score that relates directly to the size of the improvement seen, compared to control group performance; this decision will tend to increase the effect sizes obtained slightly compared to using the random (or unrepeated) condition standard deviations for both groups, as the clinical group standard deviations tend to be larger than those in the control group. For this reason, it should be noted that study effect sizes are more likely to reflect an upwards bias, rather than a downwards one.

These Cohen's *d* estimates were then entered into CMA using inverse variance weights to calculate effect sizes. In cases where a single effect size was calculated from condition means for each group across several blocks of a task, a pooled standard deviation for each group condition mean was calculated using the following formula (see Equation 1) in order to take account of the variance between the block means, as well as the variance within them.

$$S_{pGC} = \sqrt{n_{B1}(\sigma_{B1}^2 + \delta_{B1}^2) + n_{B2}(\sigma_{B2}^2 + \delta_{B2}^2) \dots + n_{Bn}(\sigma_{Bn}^2 + \delta_{Bn}^2) \div (n_{B1} + n_{B2} \dots + n_{Bn})} \quad (1)$$

Where S_{pGC} is the pooled standard deviation for a group condition mean, δ is the difference when subtracting the Grand mean from the Block mean for the group condition, σ denotes the block standard deviation for the group condition and $_B$ denotes the task block (1 to *n*).

The correlation between random and sequenced conditions for serial reaction time tasks and unrepeated and Hebb sequences for Hebb serial order learning tasks were not reported in any of the papers included in the meta-analyses. The meta-analyses were, therefore, estimated including this correlation at varying levels (0.0, 0.5, 0.7, 0.9) to assess the impact this might have on results. Inclusion of any of these correlations did not change the effect size, nor did it impact on the between study variance estimates. Therefore, since actual correlations for each study were unknown, the final meta-analyses were based on a zero correlation between conditions.

For artificial grammar and statistical learning tasks, group scores for statistical learning on the tasks (both correctly-identified recognition and generalization measures, as well as any scores for violations) were entered directly into CMA taking account of the direction of the effect. The mean of these estimates formed the effect size for the comparison. Where only one overall score per group was reported this formed the effect size for the comparison.

For weather prediction tasks, the proportion of correct responses per group was entered directly into CMA per task total or per block. In studies that reported

proportions per block, the mean of all block estimates formed the effect size for the comparison.

Several studies in the artificial grammar and statistical learning and weather prediction meta-analyses did not report scores by group, but reported *t*-tests statistics or *F* ratios that enabled an effect size to be calculated using the effect size calculator on the Campbell Collaboration website (http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php). These studies are identified in the tables accompanying each meta-analysis.

4.4.1.2   Effect size calculation for meta-analyses of correlational studies

The correlations between procedural learning and language and /or decoding measures (Pearson's *r*) were coded directly into CMA where the calculations are done with Fishers *z* and then transferred back to Pearson's *r* to ease interpretation.

4.4.1.3   Mean effect size and heterogeneity

For all meta-analyses, random effects models in CMA were then used to calculate weighted averages of individual comparison effect sizes, in order to estimate an overall effect size for each meta-analysis. 95% confidence intervals are given for each pooled effect size. The impact of any potential outliers was examined using sensitivity analyses, which give an adjusted overall effect size after removing studies one at a time. The variation in effects sizes between studies was examined, using the *Q*-test of homogeneity (Hedges & Olkin, 1985) and I^2 was used to examine the degree of any true heterogeneity that was not attributable to random error (Borenstein et al., 2009).

4.4.1.4   Moderator Coding

Where possible, moderator analyses were performed to examine the extent to which certain variables were able to explain any heterogeneity between study effect sizes. In this way, the moderator variables attempted to account for potential sources of systematic difference.

Potential moderators are detailed in the tables accompanying each meta-analysis. They included implicit learning task features. For the serial reaction time task, the following moderators were coded: task type (deterministic, alternating or probabilistic); sequence type (first or second order conditional structure); sequence length, as well as the number of sequence repetitions prior to the introduction of the random sequence (deterministic tasks) or repetitions of the sequence across the task (alternating and probabilistic versions). For the Hebb serial order learning tasks, the following moderators were coded: modality (verbal or non-verbal) and number of repetitions of the Hebb sequence. In addition verbal tasks were further subdivided into auditory-verbal or visual-verbal tasks, although insufficient numbers of studies meant this could not be examined as a moderator. For artificial grammar and statistical learning tasks, modality (verbal or non-verbal, including a further distinction between visual-verbal and auditory-verbal tasks) and complexity (finite artificial grammar or simple triplet structure) were coded. For probabilistic classification tasks, modality (verbal or non-verbal), number of trials and variations in cue probabilities were coded.

Information on relevant participant features was also coded, including age (adults or children) and language disorder classification (developmental language disorder or dyslexic). The moderating effect of the severity of language disorder in group designs was also examined. This involved coding the scores for the language measures that studies used to classify participants into groups or to support a prior diagnosis of language disorder. Group designs were, therefore, preferably supported by at least one language measure related to either (1) language ability or (2) literacy or (3) both. Tests of language ability were limited to those that indexed grammar (e.g., TROG-2) and / or vocabulary knowledge (e.g., BPVT). Literacy measures were confined to standardized tests that indexed reading ability (e.g., TOWRE). However, group design studies were still included in the meta-analyses in cases where language data was unavailable. Data for equivalent language measures had to be available for correlational studies.

The majority of group design studies, but not all, reported no significant differences between groups in NVIQ. However, this was not the case across all studies

and varying small differences between groups remained in others. For this reason the extent to which variation in the difference in NVIQ between groups reflected variation in the effect size for implicit learning was also investigated, as was the relationship between NVIQ and implicit learning in correlational studies.

4.5 Serial reaction time task: Meta-analysis of comparisons of language-disordered groups and age-matched controls

The serial reaction time task measures procedural learning by examining the speed with which participants are able to press buttons that correspond to four possible locations shown on a screen. The locations are either random or conform to a covert sequence. Faster RTs to trials that follow the sequence compared to random trials is taken as evidence of implicit learning. This measure of procedural learning is, therefore, instantiated in the difference between response times to random and sequenced trials.

Forty six eligible studies were found for the meta-analysis of group design studies including deterministic, alternating and probabilistic serial reaction time tasks (see Table 4.2). A triplet frequency learning task (Bennett et al., 2008), similar to the alternating serial reaction time task, was also included, as were two studies that used deterministic serial search tasks (Desmottes et al., 2016b; Gabay et al., 2012b), where participants pressed keys corresponding to an auditory signal (tones, letters or words dependent on the task) that followed a covert sequence. Although most tasks reported results for accuracy as well as response time, analysis was confined to the latter, as this is the principal indicator of implicit learning on serial reaction time tasks.

Table 4.2 Characteristics of the 46 group design studies eligible for the meta-analysis using the SRT task

Study	Task	Diagnosis	Age	Sample Size*	Sequence Complexity	Sequence Length	Sequence Repetitions	Additional information
Bennett, Romano, Howard Jr, & Howard, 2008[1]	Alt.	DD	Adult	16; 18	SOC	3	-	high vs low frequency triplets
Bussy, Krifi-Papoz, Vieville, Frenay, Curie, Rouselle, Rougeot, Des Portes, & Herbillon, 2011[3]	Det.	DD	Child	24; 18	SOC[5]	10	36	
Clark & Lum, 2017[8]	Det.	DLD	Child	25; 25	FOC/SOC	10	18	FOC & SOC tasks
Conti-Ramsden, Ullman & Lum, 2015[2]	Det.	DLD	Adult	45; 46	Not stated	10	36	
Deroost, Zeischka, Coomans, Bouazza, Depessemier, & Soetens, 2010[1,2]	Det.	DD	Child	28; 28	FOC/SOC	12	108	FOC & SOC tasks
Desmottes, Meulemans, & Maillart, 2016a[2]	Alt.	DLD	Child	21; 21	SOC	10	50	
Desmottes, Meulemans, & Maillart, 2016b[2]	Det. (SST)	DLD	Child	24; 24	-	6	40	Motor & verbal tasks
Desmottes, Maillart & Meulemans, 2017[2]	Alt.	DLD	Child	18; 17 & 17; 17	SOC	10	50	2 comparisons
Du & Kelly, 2013[4]	Det.	DD	Adult	12; 12	FOC/SOC[6]	12	64	
Gabay, Schiff, & Vakil, 2012b[1]	Det. (SST)	DD	Adult	14; 14	SOC	8	60	Motor & verbal tasks
Gabay, Schiff, & Vakil, 2012a[1]	Det.	DD	Adult	12; 12	SOC	12	36	
Gabriel, Maillart, Guillaume, Stefaniak, & Meulemans, 2011[2]	Prob.	DLD	Child	16; 16	SOC	8	96	
Gabriel, Maillart, Stefaniak, Lejeune, Desmottes, & Meulemans, 2013[1,2]	Det.	DLD	Child	21; 25	SOC	12	48	

Gabriel, Meulemans, Parisse, & Maillart, 2015[2]	Det.	DLD	Child	14; 14	SOC	8	48	Visual & auditory tasks
Gabriel, Stefaniak, Maillart, Schmitz, & Meulemans, 2012[1,2]	Det.	DLD	Child	15; 15	SOC	8	48	2 tasks
Hedenius, Persson, Alm, Ullman, Howard, Howard, & Jennische, 2013[2]	Alt.	DD	Child	12; 17	FOC	8	250	Across 3 sessions
Hedenius, Persson, Tremblay, Adi-Japha, Verissimo, Dye, Alm, Jennische, Tomblin, and Ullman, 2011[2]	Alt.	DLD	Child	19; 29	FOC	8	250	Across 3 sessions
Henderson & Warmington, 2017[1,2]	Alt.	DD	Adult	30; 29	FOC	8	45	Across 3 sessions
Howard, Howard, Japikse, & Eden, 2006[7]	Alt	DD	Adult	23,23		8	400	
Hsu & Bishop, 2014[1]	Det.	DLD	Child	48; 20	FOC	10	20	
Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011[3]	Det.	DD	Child	14; 14	SOC	6	74	
Kelly, Griffiths, & Frith, 2002[3]	Det.	DD	Adult	14; 14	SOC[5]	9	64	
Laasonen, Vare, Oksanen-Hennah, Leppamaki, Tani, Harno, Hokkanen, Pothos, & Cleeremans, 2014[8]	AG.	DD	Adult	36; 35	-	-	-	Sequence follows an AGL-type grammar
Lee & Tomblin, 2015[1]	Det.	DLD	Adult	23; 25	SOC	12	18	Alternating sequence & random blocks
Lee, Mueller, & Tomblin, 2016[2]	Det.	DLD	Adult	22; 19	SOC	12	18	
Lukacs & Kemeny, 2014[1,2]	Det.	DLD	Child	28; 87	SOC	12	55	
Lum & Bleses, 2012[3]	Det.	DLD	Child	13; 20	SOC	10	24	
Lum, Conti-Ramsden, Page, & Ullman, 2012[3]	Det.	DLD	Child	51; 51	-	10	36	
Lum, Gelgic, & Conti-Ramsden, 2010[2]	Det.	DLD	Child	15;15	SOC	10	36	
Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014[3]	Det.	DLD	Child	18; 65	SOC	10	20	
Menghini, Finzi, Benassi, Bolzani, Facoetti, Giovagnoli, Ruffino, & Vicari, 2010[2]	Det.	DD	Child	60; 65	FOC	9	30	

Study								
Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006[1]	Det.	DD	Adult	14;14	SOC[5]	9	30	
Menghini, Hagberg, Petrosini, Bozzali, Macaluso, Caltagirone, & Vicari, 2008[1]	Det.	DD	Adult	10; 10	SOC	9	30	
Perlant & Largy, 2011[4]	Det.	DD	Child	25; 20	-	6	25	
Rüsseler, Gerth, & Munte, 2006[2]	Det.	DD	Adult	12; 12	SOC	12	80	
Sengottuvel & Rao, 2013[1,2]	Det.	DLD	Child	17; 23	SOC	10	20	SLavg1, 2 & 3 incl.
Sengottuvel & Rao, 2014[1,2]	Det.	DLD	Child	22; 34	SOC	10	20	SLavg1, 2 & 3 incl.
Sengottuvel, Rao, & Bishop, 2016[1,2]	Det.	DLD	Child	30; 30	FOC	12	40	Only 2nd half measure
Stoodley, Harrison, & Stein, 2006[1]	Det.	DD	Adult	19; 21	SOC[5]	10	10	
Stoodley, Ray, Jack, & Stein, 2008[1]	Det.	DD	Child	45; 44	SOC[5]	6	14	
Tomblin, Mainela-Arnold & Zhang, 2007[3]	Det.	DLD	Child	38; 47	SOC	10	20	
Vakil, Lowe, & Goldfus, 2015[4]	Det.	DD	Child	23; 30	SOC[5]	12	54	
Vicari, Finzi, Menghini, Marotta, Baldi, & Petrosini, 2005[3]	Det.	DD	Child	16; 16	SOC[5]	5	60	
Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003[3]	Det.	DD	Child	18; 18	FOC	9	24	
Yang, Bi, Long, & Tao, 2013[1]	Det.	DD	Child	9; 12	SOC[5]	8	18	
Yang & Hong-Yan, 2011[1]	Det.	DD	Child	27; 27	SOC	6	20	2 tasks (each hand)

Comparisons in bold are included in final meta-analysis; [*] = Disordered group first; [1] = Means and SDs for both sequenced and random trials per group; [2] = Means and SDs for difference between sequenced and random trials only; [3] = Effect size from previous meta-analysis based on difference scores; [4] = insufficient data for any effect size measure; [5] = Categorized differently in Lum et al's meta-analysis; [6] = Both conditional properties within one sequence; [7] = Sufficient data reported for correlational analysis only; [8] = Data normalized with z-score transformation, removing between subjects variance; Det. = Deterministic SRT sequence structure; Alt. = Alternating SRT sequence structure; SST = Serial Search Task; Prob. = Probabilistic SRT sequence structure; AG. = Artificial Grammar SRT sequence structure

Although 46 studies were eligible for the meta-analysis, only 21 of these reported or were able to supply data as means and standard deviations by group for each of the sequence types separately. Two of these were excluded, as means and standard deviations had been normalized using a Z-score transformation referenced to the median to control for individual differences (Clark & Lum, 2017; Laasonen et al. 2014). Transforming data in this way reduces between subjects variance, artificially inflating any effect size estimates.

For this reason only the 19 independent comparisons of serial reaction time with language-disordered groups and age-matched controls studies that reported or were able to supply untransformed data per sequence type and group were included in the final meta-analysis (see Figure 4.3). In total these studies included 428 participants with language disorder (mean sample size = 22.53, SD = 10.77, range 9 to 48) and 488 control participants (mean sample size = 25.68, SD = 17.21, range 10 to 87). The overall mean effect size was small, but significant, g = -0.28, 95% CI [-0.42, -0.15]. However the variation in effect sizes was not significant sizes, Q (18) = 17.89, p = .46, I^2 = 0.00%, k = 19, Tau^2 = .00. With an I^2 of zero, it was not possible to examine the effect of age, diagnosis and severity of language or decoding impairment in the disordered group, or the task variables of sequence length and number of repetitions.

| Study name | Statistics for each study | | | | | Hedges's g and 95% CI |
	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	
Sengottuvel & Rao, 2013	-1.262	0.346	-1.940	-0.583	0.000	
Gabay et al., 2012b	-0.643	0.387	-1.402	0.116	0.097	
Sengottuvel & Rao, 2014	-0.615	0.276	-1.157	-0.074	0.026	
Stoodley et al., 2006	-0.572	0.317	-1.193	0.049	0.071	
Gabay et al., 2012a	-0.557	0.403	-1.347	0.234	0.168	
Menghini et al., 2006	-0.453	0.372	-1.182	0.276	0.223	
Hsu & Bishop, 2014	-0.421	0.267	-0.945	0.102	0.115	
Yang et al., 2013	-0.406	0.459	-1.306	0.493	0.376	
Gabriel et al., 2013	-0.360	0.293	-0.935	0.215	0.220	
Menghini et al., 2008	-0.314	0.431	-1.159	0.531	0.467	
Gabriel et al., 2012	-0.234	0.358	-0.936	0.467	0.513	
Yang & Hong-Yan, 2011	-0.194	0.269	-0.721	0.333	0.471	
Sengottuvel et al., 2016	-0.180	0.255	-0.681	0.320	0.480	
Lukacs & Kemeny, 2014	-0.124	0.216	-0.547	0.300	0.567	
Stoodley et al., 2008	-0.116	0.210	-0.528	0.296	0.581	
Lee & Tomblin, 2015	-0.105	0.284	-0.663	0.452	0.712	
Deroost et al., 2010	-0.046	0.264	-0.562	0.471	0.863	
Henderson & Warmington, 2017	0.020	0.257	-0.483	0.523	0.938	
Bennett et al., 2008	0.144	0.336	-0.514	0.802	0.668	
	-0.283	0.068	-0.415	-0.150	0.000	

-2.00   -1.00   0.00   1.00   2.00

Figure 4.3 Forest plot showing effect sizes for the group difference in performance on the SRT task in 19 studies, with the effect sizes calculated using the control group's standard deviations for random trials. Although Lukács & Kemény (2012) reported *z*-transformed scores in their paper, means and SDs supplied in milliseconds were used in the meta-analysis for consistency with the majority of other studies included.

4.5.1 Comparing methods of calculating effect sizes

It should be noted that the method used to calculate effect sizes in this meta-analysis is different to the method used in the previous meta-analyses of serial reaction time tasks discussed earlier (Lum et al., 2013; 2014; Obeid et al., 2016). These previous meta-analyses base their effect size calculations on a method set out in an early meta-analysis of serial reaction time tasks in Parkinson's disease patients by Siegert, Taylor, Weatherall, and Abernethy (2006). At first glance this method looks identical to the one we have used (see Equation 2), but the pooled standard deviation that forms the denominator of the equation in Siegert et al.'s method only uses a single standard deviation for each group for the difference between the conditions (see Equation 3), rather than standard deviations for raw scores.

$$d = \left(M_{Control\ group} - M_{LD\ group}\right) \div S_p \tag{2}$$

Where S_p is calculated as follows:

$$S_p = \sqrt{\left(S^2_{control\ mean\ difference} + S^2_{LD\ mean\ difference}\right) \div 2} \tag{3}$$

This method is questionable, since the denominator that represents variance in the effect size equation will be underestimated as a result of using only variance of difference scores (not the variance of component raw scores). This will inflate the estimate of effect size obtained (Morris & Deshon, 2002), as has been previously demonstrated (Lund, 1988; Ray & Schadish, 1996).

In order to illustrate the important point about the impact of different methods of effect size calculation, the above effect size can be compared to the moderate and significant overall effect size (see Figure 4.4) that resulted from an analysis of the 22 comparisons reporting or supplying data as the difference between sequences, $g = -0.44$, 95% CI [-0.60, -0.27]. This analysis calculated the effect sizes using only the standard deviations of the difference between sequences in the same way as previous meta-analyses in this literature. The overall effect size showed significant variation in

effect sizes $Q(21) = 39.19$, $p = .01$, $I^2 = 46.41\%$, $k = 22$, $Tau^2 = .07$. This result is similar to those in the meta-analyses by Lum et al. (2013; 2014), but will not be considered further for the reasons outlined above.

| Study name | Statistics for each study | | | | | Hedges's g and 95% CI |
	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	
Sengottuvel & Rao, 2013	-1.706	0.367	-2.425	-0.987	0.000	
Lum et al., 2010	-0.938	0.375	-1.673	-0.202	0.012	
Russeler et al., 2006	-0.924	0.416	-1.740	-0.109	0.026	
Sengottuvel & Rao, 2014	-0.904	0.283	-1.458	-0.349	0.001	
Conti-Ramsden et al., 2015	-0.681	0.214	-1.101	-0.262	0.001	
Gabriel et al., 2013	-0.652	0.299	-1.237	-0.066	0.029	
Desmottes et al., 2016b	-0.605	0.291	-1.174	-0.035	0.037	
Sengottuvel et al., 2016	-0.602	0.261	-1.113	-0.091	0.021	
Desmottes et al., 2017 - Exp 1	-0.559	0.337	-1.220	0.101	0.097	
Lukacs & Kemeny, 2014	-0.524	0.219	-0.952	-0.095	0.017	
Hedenius et al., 2013	-0.519	0.379	-1.262	0.224	0.171	
Desmottes et al., 2016a	-0.488	0.307	-1.091	0.114	0.112	
Clark & Lum, 2017	-0.412	0.282	-0.965	0.141	0.144	
Desmottes et al., 2017 - Exp 2	-0.393	0.338	-1.056	0.270	0.246	
Gabriel et al., 2012	-0.277	0.359	-0.980	0.427	0.441	
Menghini et al., 2010	-0.162	0.178	-0.511	0.188	0.364	
Deroost et al., 2010	-0.092	0.264	-0.608	0.425	0.728	
Gabriel et al., 2011	-0.050	0.345	-0.725	0.626	0.886	
Lee et al., 2016	-0.040	0.300	-0.629	0.549	0.893	
Hedenius et al., 2011	0.017	0.290	-0.552	0.586	0.953	
Henderson & Warmington, 2017	0.132	0.255	-0.368	0.632	0.605	
Gabriel et al., 2015	0.236	0.346	-0.442	0.914	0.495	
	-0.437	0.085	-0.604	-0.269	0.000	

-2.00   -1.00   0.00   1.00   2.00
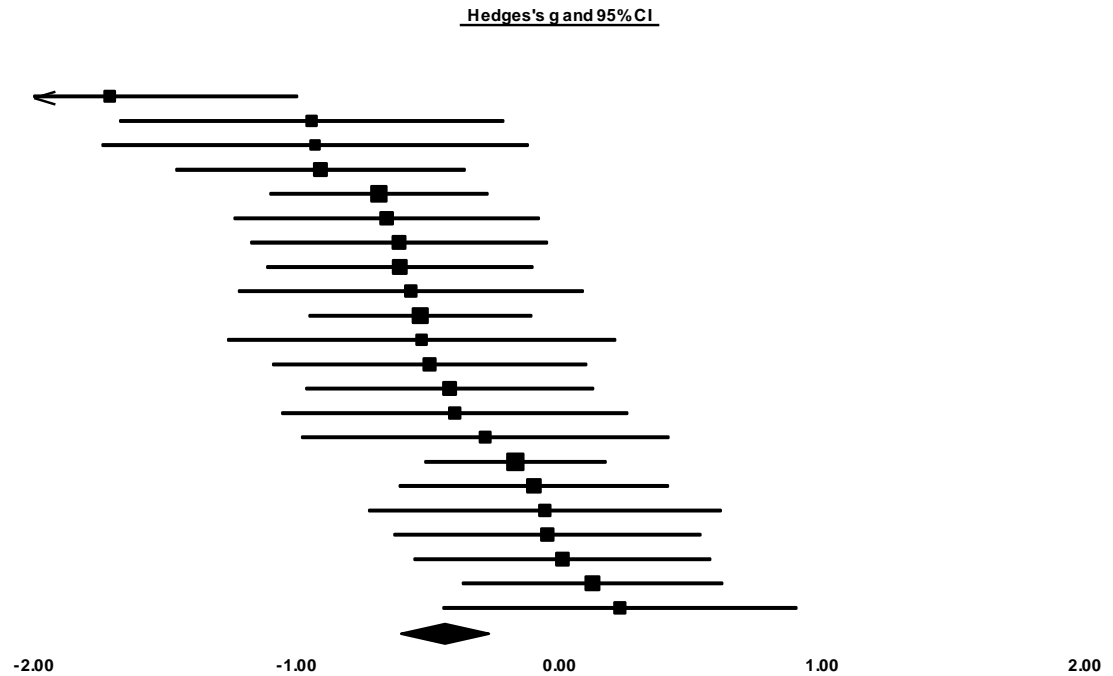
Figure 4.4 Forest plot showing effect sizes for the group difference in performance on the SRT task in 22 studies using the mean and standard deviation for the difference between sequenced and random blocks per group. The analysis included 22 separate comparisons in total.

105

To further emphasise the very real impact of the different calculation methods, effect sizes were calculated using both methods (see Figure 4.5) for all studies that were able to provide information in both formats ($n = 8$). Using the difference score methodology, followed by the previous meta-analyses in the literature, the effect size was moderate and significant, $g = -0.55$, 95% CI [-0.90, -0.21], showing language disordered groups performing more poorly on serial reaction time tasks compared to age-matched controls. This method also showed significant variation in effect sizes too $Q\ (7) = 22.18$, $p = .002$, $I^2 = 68.44\%$, $k = 8$, $Tau^2 = .17$.

However, since the variance in these effect sizes was under-estimated, the pooled effect size was larger than it should have been. The preferred raw score methodology, using the standard deviations for the control group's random trials, gave a different picture of the data. The effect size for the eight studies was lower, $g = -0.31$, 95% CI [-0.57, -0.05], as was the heterogeneity estimate $Q\ (7) = 12.43$, $p = 0.09$, $I^2 = 43.69\%$, $k = 8$, $Tau^2 = 0.06$. This comparison of methods demonstrates the importance of using the optimal raw score method of calculating effect sizes in group design studies using tasks that rely on the difference between experimental conditions as their dependent variable.

Study name | | Statistics for each study | | | | Hedges's g and 95% CI

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Sengottuvel & Rao, 2013	-1.706	0.367	-2.425	-0.987	0.000
Sengottuvel & Rao, 2014	-0.904	0.283	-1.458	-0.349	0.001
Gabriel et al., 2013	-0.652	0.299	-1.237	-0.066	0.029
Sengottuvel et al., 2016	-0.602	0.261	-1.113	-0.091	0.021
Lukacs & Kemeny, 2014	-0.524	0.219	-0.952	-0.095	0.017
Gabriel et al., 2012	-0.277	0.359	-0.980	0.427	0.441
Deroost et al., 2010	-0.092	0.264	-0.608	0.425	0.728
Henderson & Warmington, 2017	0.132	0.255	-0.368	0.632	0.605
	-0.549	0.177	-0.896	-0.202	0.002
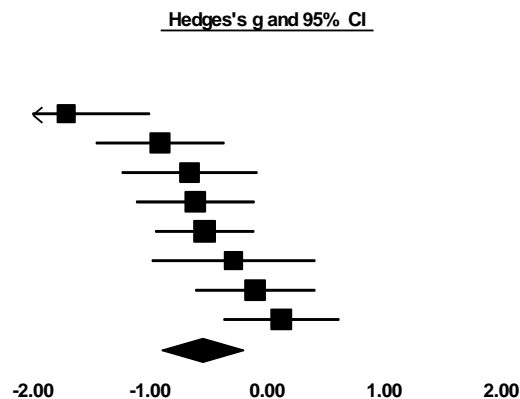
Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Sengottuvel & Rao, 2013	-1.262	0.346	-1.940	-0.583	0.000
Sengottuvel & Rao, 2014	-0.615	0.276	-1.157	-0.074	0.026
Gabriel et al., 2013	-0.360	0.293	-0.935	0.215	0.220
Gabriel et al., 2012	-0.234	0.358	-0.936	0.467	0.513
Sengottuvel et al., 2016	-0.180	0.255	-0.681	0.320	0.480
Lukacs & Kemeny, 2014	-0.124	0.216	-0.547	0.300	0.567
Deroost et al., 2010	-0.046	0.264	-0.562	0.471	0.863
Henderson & Warmington, 2017	0.020	0.257	-0.483	0.523	0.938
	-0.310	0.131	-0.566	-0.054	0.018
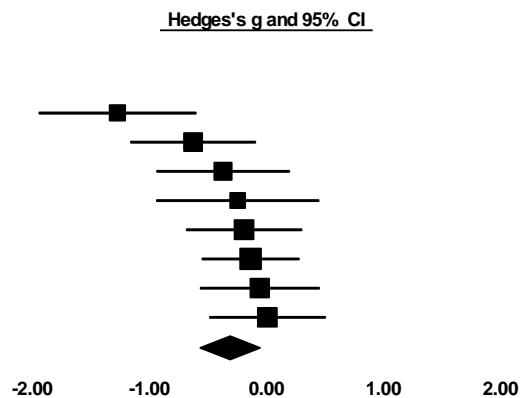
Figure 4.5 Meta-analysis of 8 studies: The top forest plot uses the mean and SDs in milliseconds for the difference between sequenced and random blocks per group. The bottom forest plot uses means and SDs in milliseconds for both sequenced and random blocks per group.

107

4.5.2 Examining variability in group differences in severity of language disorder and NVIQ

In spite of the nonsignificant heterogeneity estimate in the meta-analysis of 19 studies using the serial reaction time task, variability in group differences in language skills and decoding in all eligible studies was also analysed, as a possible explanation for the inconsistency in the literature. Studies using decoding measures all investigated the procedural deficit hypothesis in dyslexia ($n = 20$). There was a large variation between the degree of difference between the disordered and comparison groups, mean difference was $g = -2.30$, 95% CI [-2.73, -1.87], ranging from $g = -7.48$ to $g = -0.63$, showing significant heterogeneity between these studies $Q (19) = 114.68$, $p < .001$, $I^2 = 83.43\%$, $k = 20$, $Tau^2 = 0.74$. One study (Jiménez-Fernández et al., 2011) reported a very large difference between groups with an effect size of $g = -7.48$. A sensitivity analysis showed that after removing outliers, the overall effect size was in the range of $g = -2.14$, 95% CI [-2.73, -1.87] to $g = -2.38$, 95% CI [-2.79, -1.97].

Twenty four studies reported language measures (grammar and / or vocabulary) by group, of which 21 were studies investigating the procedural deficit hypothesis in developmental language disorder. Once again, there was a large amount of variability in the level of language ability in the language-disordered groups, mean difference was $g = -1.81$, 95% CI [-2.16, -1.45], ranging from $g = -4.29$ to $d = -0.27$, showing significant heterogeneity between the studies $Q (23) = 139.56$, $p < .01$, $I^2 = 83.52\%$, $k = 23$, $Tau^2 = 0.63$.

Many studies endeavoured to match groups for NVIQ. However, the extent of the difference between groups still differed widely across studies. NVIQ scores for groups were, therefore, coded where possible ($n = 31$) and this showed significant variation across studies, mean difference was $g = -0.35$, 95% CI [-0.51, -0.20], ranging from $g = 0.51$ to $g = -.66$. This variation in effect sizes between studies was significant, $Q (30) = 61.46$, $p = .001$, $I^2 = 51.19\%$, $k = 31$, $Tau^2 = 0.09$.

It is, therefore possible that, inspite of the nonsignificant heterogeneity estimate for the group deficit in procedural learning on the serial reaction time task, variation in the severity of decoding or language impairment or level of NVIQ in the disordered groups across the literature is involved in the pattern of inconsistent results.

4.5.3 Publication bias

A funnel plot was not considered as a measure of publication bias, given the subset of studies contained in the meta-analysis. Instead, a p-curve (Simonsohn et al., 2013; 2014) was estimated in order to identify whether the body of published work (all 46 eligible studies) was subject to publication bias. Publication bias occurs when studies with significant findings are published while those with null findings are not. This can lead to results that appear to be replicated in press being treated as true effects, when no such trust is warranted. The p-curve examines the distribution of significant results, with the shape of the curve determining the evidential value of the studies it contains. It does this by calculating the probability of observing a p-value as extreme if the null were true for each significant p-value. It then aggregates these to give a chi square test for skew, such that only right-skewed curves with more low than high p values show evidential value. The gradient of right-skewed curves increases as power increases, so steeper curves are suggestive of higher powered findings. This is the case even when findings are highly heterogenous. By contrast, a left-skewed curve indicates possible p-hacking, where any combination of decisions on data collection, screening, transformation, introduction of covariates, etc. may have been taken, in order to obtain a significant result. Such a curve would contain a disproportionate number of high p values (close to the alpha of .05), which points to authors stopping analysis once they obtain a significant result. It should be noted that p-hacking for evidence of true effects happens too and this is particularly the case when studies are underpowered with small sample sizes. In these cases the curve will incorporate the right skew from the true effect and the left skew from the p-hacking. Finally, p-curves with a uniform horizontal line suggests the studied effect does not exist (ie: every p-value below .05 should appear equally often) or at the least that there are not enough p-values to infer evidential value.

The procedural deficit hypothesis at the centre of these meta-analyses, claims that language-disordered groups will display poorer implicit learning on the implicit memory tasks than control groups with normal language. Therefore, a single statistic that related to the principal measure of implicit learning on the serial reaction time task was coded for each study (see Table 4.3). This was the statistic that referred to the difference in RTs between sequenced and random trials. For deterministic tasks, the statistic typically related to the difference between the last sequenced block and a subsequent block of random trials. For alternating or probabilistic tasks, this measure was sometimes taken across the whole of the task. For the majority of the 46 studies eligible for the meta-analysis, this principal measure of implicit learning was the ANOVA interaction between group x sequence x block. Where studies used an alternative analysis, the equivalent statistic was selected. Where studies contained two comparisons, a statistic was coded for each one and p-curves were run twice, each time including only the first or the second comparisons from the study, as recommended by Simonsohn et al. (2013). The results for the two p-curves were equivalent, so only the first one is reported here.

Of the 46 studies, 23 reported significant results for a difference between groups on the principal measures of implicit learning and 23 studies reported null results (see Table 4.3), underlining the inconsistency of results in the field. Several of these null results came from studies claiming support for the procedural deficit hypothesis, in the light of significant secondary findings, so the full extent of non-significant findings on the principal implicit learning measure for the serial reaction time task were not immediately apparent from the literature. For example, Bennett et al. (2008) reported a null result, but claimed support for the procedural deficit hypothesis in light of a positive correlation between implicit learning scores and reading ability. Desmottes et al. (2016a; 2017) reported initial null results, but impaired consolidation of procedural learning in children with developmental language disorder, with poorer performance during a second attempt at the task. Similar results were also reported on an alternating

serial reaction time task in dyslexic children (Hedenius et al., 2013)[2]. Implicit learning impairments in language disorder have also been linked to task-specific differences too. Gabriel et al. (2014) reported equivalent learning for groups with regards to response times, but suggested that children with developmental language disorder might be more error prone than typically developing children during an auditory, but not a motor, version of the serial reaction time task. Only seven studies (Bussy et al., 2011; Gabriel et al., 2011; Kelly et al., 2002; Laasonen et al., 2014; Lum & Bleses, 2012; Rüsseler et al., 2006; Vakil et al., 2015) stood firmly behind their null result on the serial reaction time task.

Three of the 23 studies with significant results reported statistics in a format that could not be included in the p-curve, failing to report the F-ratio and including only the p value (Menghini et al., 2008; Stoodley et al., 2006; 2008). One study reported no between group difference during a first training session, but a significant difference over subsequent sessions in two separate experiments (Desmottes et al., 2017). Only one experiment was included in each analysis, with no significant difference to results. In addition, three studies reported results that approximated the test of interest (significant group differences in the difference in RTs between random and sequenced trials), but with minor variations. The first of these reported significant results for differences in the growth curve of the sequenced phase of the task, without reference to the random phases (Tomblin et al., 2007). Two others reported the group x block difference across all blocks in the task, sequenced and random (Vicari et al., 2003; 2005). As recommended by Simonsohn et al. (2013) the p-curve analysis was run with and without these three studies, but found equivalent results both times. Therefore, p-curve results are reported for all 20 studies with significant results. Figure 4.6 shows a right-skewed p-curve which demonstrates evidential value for the 46 studies eligible for the serial reaction time task extreme groups meta-analysis ($Z = 4.11$, $p < .001$).

---

[2] A finding of impaired consolidation of implicit learning should be put in context at this point as contradictory results have also been reported. Gabay et al. (2012a) found the opposite, with dyslexic adults performing comparably with controls during later learning stages, while showing impaired learning during initial acquisition.

There is also no reliable evidence that the studies' evidential value is inadequate due to low power (power estimate = 53%, 90% CI: [27%, 75%].
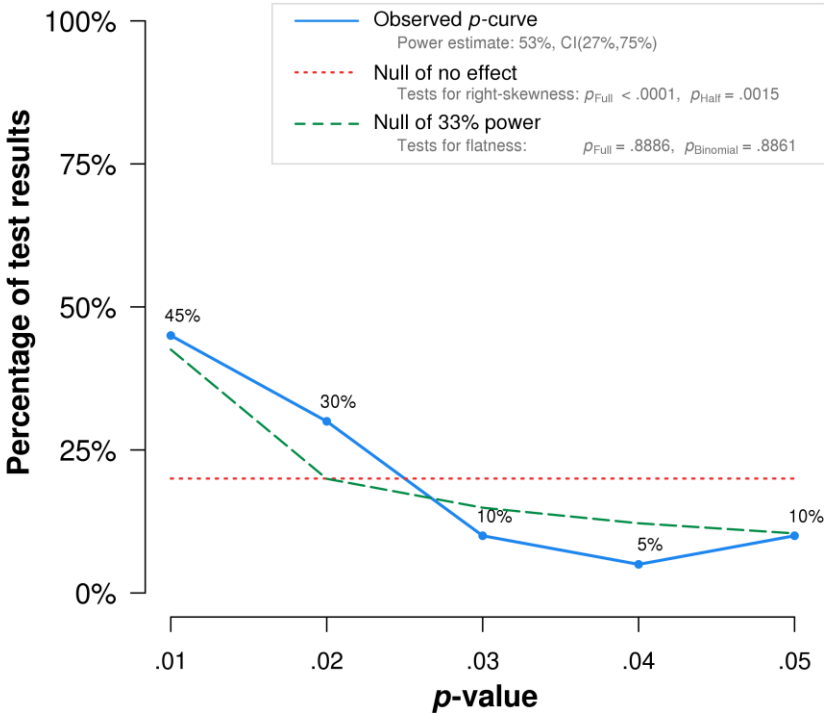
Figure 4.6 P-curve examining publishing bias in extreme groups studies using the serial reaction time task to investigate the procedural deficit hypothesis.

Table 4.3 Disclosure table for 46 group design studies eligible for the meta-analysis using the SRT task.

Study name	Analysis	Quoted test from paper with statistical results	Significance*
Bennett, Romano, Howard Jr, & Howard, 2008	RT difference between high and low frequency triplets by group	"Group x triplet type and Group x triplet type x epoch interactions were not significant, P's > .10, indicating that we did not detect group differences in sequence learning." (p. 190)	Null
Bussy et al., 2011	2 (group) x 2 (sequence) x 6 (blocks) ANOVA.	"Premierement, L'effet principal du facteur groupe n'est pas significatif ($F(2,40) = 1.43$; p > 0.10) [...].La difference de temps de reaction entre le dernier bloc sequential et le dernier bloc aleatoire (le cinqieme bloc) est egalement significative pour CG (F(2,40) = 32.55, $p < .001$), pour DP (F(2,40) = 14.26, $p < .001$), et pour DS (F(2,40)= 20.39, $p < .001$)." (p. 144)	Null
Clark & Lum, 2017	FOC: RTs for random block were compared to mean RT for sequenced blocks 3 & 5. A 2 (group) x 2 sequence type: Block 4 vs mean of Blocks 3 & 5)	"However, a significant Group x Block interaction with a medium to large effect size was observed, $F(1,50) = 4.785$, $p = .033$, $\pi^2p = .087$." (p. 154)	Significant
Clark & Lum, 2017	SOC: RTs for random block were compared to mean RT for sequenced blocks 3 & 5. A 2 (group) x 2 sequence type: Block 4 vs mean of Blocks 3 & 5)	"Neither the main effect of group [...], nor the interaction between block and group was significant, $F(1,50) = .725$, $p = .399$, $\pi^2p = .014$." (p. 154)	Null
Conti-Ramsden, Ullman & Lum, 2015	Difference Z score between block 4 and 5. T-test difference between groups.	"Children with DLD had significantly lower scores on all predictor variables." (p. 6). $t(89) = 3.00$, $p = .003$ (Table 2, p.7)	Significant
Deroost, Zeischka, Coomans, Bouazza, Depessemier, & Soetens, 2010	RT difference between B14 (random) and mean of Sequence blocks 13 & 15. A 2 (Group) x 2 (task) x 2(sequence type). NB: P therefore includes 2 tasks (FOC & SOC)	"Critically, no interaction of Group x Sequence learning, nor an interactio of Group x Sequence x Sequence Learning could be observed, both $F < 1$." (p. 566)	Null

Desmottes, Meulemans, & Maillart, 2016a	Effect of group overall on RT difference, as part of a 2 (Group) x 4 (Difference score on Epoch 1, 5, 6, 7) ANOVA. NB: Epoch 1 and 5 are start and end of Day 1, Epoch 6 is 24 hrs later and Epoch 7 is 1 week later. Single mean is for overall difference over epochs.	"This analysis showed a marginal effect of Group ($F(1,40) = 3.46$, $p = .066$, $\pi^2 p = .08$), indicating a (slightly) better sequence knowledge in children with TD (M = 0.14, SD = 0.11) than in children with DLD (M = 0.09, SD = 0.09)." (p. 60)	Null
Desmottes, Meulemans, & Maillart, 2016b	RT difference between B5 (random) and mean of Sequence blocks 4 & 6. A 2 (Group) x 2 (task) x 3 (Block 4-6)	"Interestingly, the interaction between block and group showed that these differences in RT's differed between groups ($F(2,92) = 3.22$, $p = .044$) [...] Indeed the difference between the random and both surrounding sequence blocks was significant in TD children ($F(1,46) = 23.197$, $p < .001$), but not for children with DLD ($F(1,46) = 2.525$, $p = .140$)". (p 525)	Significant
Desmottes, Maillart, & Meulemans, 2017 - Experiment 1	RT difference between B5 (random) and mean of Sequence blocks 4 & 6. A 2 (Group) x 3 (Epoch 1 - 3 difference scores)	"Finally, there was no interaction between group and epoch, $F(2,66) = .237$, $p = .789$, $\pi^2 p = .007$, indicating that a similar improvement in sequence knowledge with practice could be observed in both DLD and TD groups" (p. 8)	Null
Desmottes, Maillart, & Meulemans, 2017 - Experiment 2	RT difference between B5 (random) and mean of Sequence blocks 4 & 6. A 2 (Group) x 3 (Epoch 3 - 5 difference scores).	[The ANOVA] "...showed no main effect of group...or epoch...Nevertheless, the interaction between the two variables was statistically significant, $F(2,64) = 5.85$, $p = .004$, $\pi^2 p = .155$. This indicated that the evolution of the sequence knowledge differed between the groups over the post-training sessions." (p. 12)	Significant over several sessions
Du & Kelly, 2013	The difference between block 9 (random) and mean of blocks 8 & 10 (sequence). 2 (Group) x 2 (Block 9 vs mean of Blocks 8 & 10)	"...no significant effect of group [...] and no significant interaction of group x block, $F(1,22) < 1$. These results indicate that both dyslexic and control groups demonstrated significant and comparable learning." (p. 162)	Null
Gabay, Schiff, & Vakil, 2012a	Transfer measure of difference between Block 4 & 5. 2 (Group) x 2 (Block 4 (Sequence) to 5 (Random)) ANOVA	"The interaction between these variables did not reach significance, $F(1,22) = 1.648$, MSE = 682, $p > .05$." (p. 284)	Null

Gabay, Schiff, & Vakil, 2012a	Recovery measure of difference between Block 5 & 6. 2 (Group) x 2 (Block 6 (Sequence) to 5 (Random)) ANOVA	"The interaction between those variables was also significant, $F(1,22) = 7.458$, MSE = 680, $p < .05$, $\pi^2 p = .25$. This pattern indicates that the DD group needs a longer time in order to recover from learning of a different sequence than does the control group." (p. 284)	Significant
Gabay, Schiff, & Vakil, 2012b	For 1st ratio: 2 (Group) x 2 (sequence transfer - Block 3 to 4) x 2 (task: motor vs letters); second ratio is the same but task specific	"The group by transfer interaction was marginally significant, $F(1,26) = 3.53$, $p = .07$ [….] In order to analyse this interaction, separate 2 (transfer) x 2 (group) ANOVAs were computed for each sequence type. For the motor sequence, the group by transfer interaction was far from significance $F<1$, suggesting that both groups learned the specific motor sequence [...] For the letter names sequence, the group by transfer interaction was significant, $F(1,26) = 7.89$, $p < .01$." (p. 2438) (NB: F-ratio for Letters SST is entered into p-curve.)	Significant
Gabriel, Maillart, Stefaniak, Lejeune, Demottes, & Meulemans, 2013	Difference in RTs between last sequenced and random block. 2 (Group) x 2 (Block 6S vs 7R) ANOVA	"However, the Group by Block interaction was not significant $F(1,40) = 2.87$, MSE = 1642, $p = .09$, $\pi^2 p = .06$, [...] suggesting that the magnitude of the RT difference between blocks 6 and 7 does not differ significantly between groups." (p. 268)	Null
Gabriel, Maillart, Guillaume, Stefaniak & Meulemans, 2011	Probability by Group interaction on last block	"…the Probability by Group interaction was non-significant, $F(1,28) = .039$, MSE = 2970, $p = .84$, $\pi^2 p = .0014$, p = .84, suggesting that all children (DLD vs. NL) responded faster for probable than improbable locations." (p. 340)	Null
Gabriel, Meulemans, Parisse, & Maillart, 2015	Auditory modality: Difference in RTs between last sequenced and random block. 2 (Group) x 2 (Block 6S vs 7R) ANOVA	"We first performed and ANOVA in the auditory modality […] The results showed no group effect…, a block effect..., and no interaction effect, $F(1,26) = 1.05$, $p = .31$, $\pi^2 p = .039$." (p. 14)	Null
Gabriel, Meulemans, Parisse, & Maillart, 2015	Visual modality: Difference in RTs between last sequenced and random block. 2 (Group) x 2 (Block 6S vs 7R) ANOVA	"We then performed the same analysis in the viusal modality and found comparable results: no group effect… a Block effect… and no interaction effect, $F(1,26) = 0.46$, $p = .503$, $\pi^2 p = .017$)..." (p. 14)	Null

Gabriel, Stefaniak, Maillart, Schmitz, & Meulemans, 2012	2 (Group) x 2 (Block 6seq to block 7Rand & Group) ANOVA	"However, the interaction was not significant, $\pi^2 p$ (1,28) = .0005, MSE = 12172, p = .98, $\pi^2 p$ < .001, suggesting that both groups demonstrated a significant increase in their RTs from Block 6 to Block 7." (p. 334)	Null
Gabriel, Stefaniak, Maillart, Schmitz, & Meulemans, 2012	2 (Group) x 2 (Block 6seq to block 7Rand & Group) ANOVA	"...the Block x Group interaction was nonsignificant, $F(1,28)$ = 2.59. MSE = 12172, p = .11, $\pi^2 p$ < .08." (p. 335)	Null
Hedenius, Persson, Tremblay, Adi-Japha, Verissimo, Dye, Alm, Jennische, Tomblin, and Ullman, 2011	RT difference between high and low frequency triplets per group by epoch. A 2 (Group) x 5 (Epoch difference score) ANCOVA controlling for NVIQ	"…, though this was qualified by a significant Group x Epoch interaction, also with a medium to large effect size ($F(1,45)$ = 6.56, p = .014, $\pi^2 p$ = .127)." (p. 10)	Significant
Hedenius, Persson, Alm, Ullman, Howard, Howard, & Jennische, 2013	RT difference between high and low frequency triplets per group by epoch. 2 (Group) x 2 (trial-type interaction) x 3 (learning stage)	"Of particular interest here, the two groups did not differ with respect to sequence learning effects on RT (group x trial type interaction: $F(1, 27)$ < 1; group x trial type x learning stage interaction: $F(2,54)$ = 1.51, p = .230, $\pi^2 p$ = .053." (p. 3928)	Null
Henderson & Warmington, 2017	RT difference between sequenced and random trials across task. 2 (Group) x 2 (sequence type) x 5 (Block).	"There were no significant interactions: […] Condition x Block x Group F < 1." (p. 204) (NB: This is for Day 1 only, but results are also null for consolidation sessions too).	Null
Howard, Howard, Japikse, & Eden, 2006[1]	2 x 2 (Group x sequence)	"Although both groups show sequence learning, the dyslexics show significantly less learning than controls on both measures. This is supported by significant Trial Type x Group interactions for [...] speed $F(1,21)$ = 4.61, MSE = 226.58." (p. 1135)	Significant
Hsu & Bishop, 2014	Group diffs compared with growth curve analysis (as in Tomblin et al, 2007)	"…we examined changes in the RTs when the task proceeded from the pattern phase to the subsequent random phase […] There was a significant effect of group ($F(2,41.76)$ = 9.51, p < .0001), with a greater reboundin RTs in the age-matched group than the other two groups" (p. 359)	Significant
Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011	2 (Group) x 2 (Sequence type) ANOVA	The Group x Type of Block interaction also reached significance ($F(1,26)$ = 13.49, p = .002). (p 96)	Significant

116

Kelly, Griffiths, & Frith, 2002	2 (Group) x 2 (Sequence type) ANOVA	"*F*<1. The lack of significance for these interactions suggests that the amount of learning shown by the two groups is not significantly different from each other…" (p. 49)	Null
Laasonen, Vare, Oksanen-Hennah, Leppamaki, Tani, Harno, Hokkanen, Pothos, & Cleeremans, 2014	Difference in RTs between last random block 12 and mean of sequence blocks 11 & 13. 3 (Group: control, dyslexia, ADHD) x 2 (sequence type).	"The group x block type interaction did not reach significance, $F(2,82) = .308$, $p = .736$, $\pi^2p = .007$, observed power = 0.097." (p. 18)	Null
Lee & Tomblin, 2015	RT difference between interleaved Random and Sequence blocks. 2 (Group) x 2 (sequence type) ANOVA.	"However, the interaction effect was not significant, $F(1,46 = .39$, $p = .54$, $\pi^2p = .01$." (p. 224)	Null
Lee, Mueller, & Tomblin, 2016	RT difference between Random and Sequence blocks. T-test difference between groups for learning score.	Independent samples t-test showed that the learning effect was not significantly different between the two groups in our study, $t(39) = .13$, $p = .90$. (p. 1105)	Null
Lukacs & Kemeny, 2014	Difference between sequenced block 11 and random block 12. Univariate ANOVA (Group) on transformed difference scores to take account of participant variability.	"Next, the difference between the mean of z-transformed Block 11 (the last sequence block) RTs were extracted from the mean of the z-transformed Block 12 (random block) RTs. This difference reflecting the size of sequence learning was compared by group, revealing a significant group main effect, $F(1,113) = 5.888$, $p < .05$, $\pi^2p = .050$, with bigger learning effect in the control than in the SLI group." (p. 478)	Significant

Lum & Bleses, 2012	Normalised RT difference between sequence block (Block 4) and random block (Block 5). Analysis on difference score conducted by group and difference in effect sizes compared for significance.	"The first analysis revealed that the TD group had significantly slower RTs in Block 5 compared to Block 4 ($F(1,19) = 42.194$, $p < .001$, $\pi^2 p = .690$). The second analysis indicated that the SLI group also had significantly slower RTs in Block 5 compared to Block 4 ($F(1,12) = 6.354$, $p = .027$, $\pi^2 p = .389$). While both groups were found to have slower RTs in Block 5, it is interesting to note that th eff3ect size for the RD group is larger in comparison to the SLI group. However, the difference in effect sizes was not found to be statistically significant ($z = 1.15$, $p = .25$)." (p 54) — Null
Lum, Conti-Ramsden, Page, & Ullman, 2012	Normalised RT difference between sequence block (Block 4) and random block (Block 5). One way ANOVA on this difference.	"One-way repeated-measures ANOVA revealed a significant effect of group [$F(1,102) = 5.17$, $p = .026$, $\pi^2 p = .58$], with an approximately medium effect size, indicating a larger RT difference between blocks 4 and 5 for the TD children than the children with SLI." (p. 1148) — Significant
Lum, Gelgic, & Conti-Ramsden, 2010	Normalized RT difference between sequence block (Block 4) and random block (Block 5). T test of this difference between groups, controlling for motor speed.	"Analysis of these standardised residuals indicated the magnitude of difference between the fourth and fifth Blocks was significantly larger for the TD than the SLI group ($t(27) = 2.545$, $p = .017$, $r^2 = .193$)." (p. 104) — Significant
Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014	2 (Groups) x 2 (Sequence type) x 5 (Block)	"The groups (SLI versus C) differed in their performance in the Blocks, Groups x blocks,... but not in the sequence, Groups x sequence, $F(1,80) = .614$, ns. No triple interaction, Blocks x sequences x Group, $F(4,77) = .369$, ns), indicating an absence of statistical differences in motor learning between both groups." (p. 18) — Null
Menghini, Finzi, Benassi, Bolzani, Facoetti, Giovagnoli, Ruffino, & Vicari, 2010	Difference in RTs between last sequenced (Block 6) and random block (Block 7) as z scores, with controls at mean = 0 (SD = 1) . MANCOVA with Age as covariate, Group as between subjects factor and cognitive task measures as DVs.	"Finally, in the GLM procedure, no significant difference was found in the SRTT between children with DD and NR children, considering the difference between RTs of the last pseudo-random block (R2) and the last sequenced block (S4) as an index of viusal-motor sequence learning (in DD mean z-score +/- SD: SRTT: -.17 +/- 1.09)." (p. 867) — Null

Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006	Difference in RTs between last sequenced (Block 6) and random block (Block 7). 2 (Groups) x 2 (Sequence type) ANOVA.	"The block effect […] and the group by block interaction ($F(1,26) = 6.5$, $p < .05$) were significant, while the group effect […] did not reach significance." (p. 4)	Significant
Menghini, Hagberg, Petrosini, Bozzali, Macaluso, Caltagirone, & Vicari, 2008	Difference in RTs between last sequenced (Block 6) and random block (Block 7). One way ANOVA in the two groups comparing RTs in the two relevant blocks.	…the group of 10 subjects with DD selected for the current study were impaired in IL, showing no SRTT changes between S5 and R2 (DD means; one-way ANOVA: $p > .1$). In contrast, the subgroup of NRs showed an IL effect (NR means; one way ANOVA: $p > .05$. (p. 216) (NB: No F-ratio given)	Significant
Perlant & Largy, 2011	Experiment 2 only: Difference between interleaved sequenced and random trials over blocks. A (Group) x 2 Sequence type) x 5 (/block) x 2 (item: linguistic and nonlinguistic) was done and no Group interactions were reported. Separate analyses for each group were then done.	"In typical readers […] analysis also shows the significance of condition x block interaction, principal indicator of sequence learning, $F(4,76) = 4.03$, $p <.001$ [...] In children with dyslexia […] The analysis also reaveals the presence of significant condition x block interaction, principal indicator of sequence learning ($F(4,96) = 4.49$, $p <.01$)." (p. 309) (NB: No Group interactions for main ANOVA were reported, indicating a null result. Both groups separately show a significant learning effect. However, the three way interaction result in each of these is different and this is claimed as a difference between groups.)	Null
Rüsseler, Gerth, & Munte, 2006	Difference in RTs between Block 10 (random) and mean of Blocks 9 and 11 (sequence). A 2 (Group) x 2 (sequence) ANOVA.	"A post-hoc F test indicates that the amount of learning did not differ reliably between the two groups (GROUP by BLOCK: $F(1,22) = 2.8$, $p < .1085$)." (p. 817)	Null
Sengottuvel & Rao, 2013	ISL = mean of final 30 trials of random - mean of final 30 trials of sequence	"Children with DSLI performed significantly poorer compared to TD children on sequence learning skill (see Table 3)." $F(1,40) = 29.61$, $p < .001$ (p. 3323)	Significant

Sengottuvel & Rao, 2014	Difference between sequenced and random RTs. ANOVA structure unclear.	"Even though, the SLavg1 of SLI was not significantly lower than TD, ISL value of the SLI group (ie: RLavg - SLavg1) was significantly lower than that of the TD group, thereby suggesting obvious slow RTs for the SLI group even in initial learning trials (see Table 2)." $F(1,54) = 10.72$, $p < .001$ (p. 58)	Significant
Sengottuvel, Rao, & Bishop, 2016	Mean untransformed difference btw random and sequence blocks. ANCOVA controls for NVIQ and age.	"This showed that children with SLI were significantly poorer than TD children, $F(1,52) = 5.76$, $p = .02$." (p. 10)	Significant
Stoodley, Harrison, & Stein, 2006	Difference between RTs on random and repeated blocks (split into 1st and 2nd half of task in paper).	"A repeated measures ANOVA showed a significatn group by condition interaction during the random and repeated sequence blocks ($p = .03$)." (p. 796) (NB: No F-ratio given).	Significant
Stoodley, Ray, Jack, & Stein, 2008	Percent decrease in RTs during the sequence condition compared to 1st random condition	"In the repeated measures analysis, there was a significant effect of block type [...] and a significant block by group interaction ($p = .001$)." (p. 178) (NB: No F-ratio given)	Significant
Tomblin, Mainela-Arnold & Zhang, 2007	Difference between groups on the 2 types of sequence given separately as Group differences in intercept for Pattern and for Random 2 trials as part of several growth curve models. The growth curve analysis measure of highlighted as the measure of interest in the paper is for pattern trials, so this is the F ratio we have selected.	Pattern Phases: [...] This model showed that the SLI group was significantly slower than the NL group at the third trial block which represents the intercept [group difference in intercept = -39.94 ($SD = 14.49$), $F(1,602) = 7.59$, $p = .018$]. (p. 281)	Significant
Vakil, Lowe, & Goldfus, 2015	Difference between last sequenced (Block 6) and final random block (Block 7). 2 (Group) x 2 (Sequence) ANOVA	In this case as well, an interaction effect was not found between the group and the influence of training, $F(1,50) = .432$, $p > .05$, as no significant difference was identified between individuals with or without DD in the increase in RT to the random sequence. (p. 475)	Null

Vicari, Finzi, Menghini, Marotta, Baldi, & Petrosini, 2005	Although implicit learning is the difference between the last sequence and final random block, the ANOVA is a 2 (Group) x 6 (block) model, so the interaction F ratio does not specifically reference implicit learning, so much as group differences over the whole task.	"…the group x block interaction ($F(5,150) = 2.8$, $p = .02$) were significant, demonstrating a different patterns of RT changes in the two groups across blocks. Critcally, for the aims of this study, the two groups RTs differed significantly (Tukey's test) passing from the fifth to the sixth block [...] controls ($p = .0002$) [...] dyslexic children ($p = 1$)." (p. 1394)	Significant
Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003	The ANOVA is a 2 (Group) x 6 (block) model, so the interaction F ratio does not specifically reference implicit learning, so much as group differences over the whole task. Control group differed significantly on difference between 5th and 6th block (p < .001), but the dyslexics did not (ns).	The group x block interaction was also significant $F(5,170) = 5.95$, $p < .0001$, thus demosntrating a different pattern of RT changes in the two groups across blocks…Critically, for the aims of the study, the RTs of the two groups strongly differed passing from the fifth to the sixth block. (p. 110)	Significant
Yang & Hong-Yan, 2011	Left and right hand tasks. Measures are the difference between sequenced block 3 and random block 4. 2 (group) x 5 (block) ANOVA, so the interaction F ratio does not specifically reference implicit learning, so much as group differences over the whole task. Left hand: control group differed significantly on difference between 3rd and 4th block (p < .05), but the dyslexics did not (ns). Right hand: both groups showed significant differences ($p < .05$)	Left hand: "The interaction between block and group was not significant, $F(4,49) = 1.16$, $p = .34$." (p. 4). Right Hand: "The interaction between block and group was not significant, $F(4,49) = .21$, $p = .93$." (p. 5)	Null

| Yang, Bi, Long, & Tao, 2013 | Difference between sequenced block 4 and random block 5. 2 (group) x 5 (block) ANOVA, so the interaction F ratio does not specifically reference implicit learning, so much as group differences over the whole task. The differnce in learning rate between groups is quantified with a *t* test statistic, however, and this is used here. | "… the interaction of group and block were not significant, […] $F(1,14) = 1.222$, $p = 0.345$, ES = 0.259 [...]The mean learning rate of RT of dyslexic group ([Block 5 - Block 4] / [Block 4 + Block 5]...) was 0.06 and control group was 0,095. But, the difference of learning rate did not reach statistic significance [$t(18) = -1.188$, $p = 0.25$]" (p. 303) | Null |

*Significance of principal indicator of implicit learning; [1]Study is a group design study, but provides sufficient data for inclusion in correlational analysis only.

4.6   Serial reaction time tasks: Meta-analysis of correlational studies

Five studies examined the relationship between serial reaction time performance and language ability using correlational designs (see Table 4.4). Only one of these studies (Waber et al., 2003) did not include sufficient information to enable calculation of an effect size between serial reaction time task performance and language ability. This study of 422 children found no evidence of a relationship between impaired sequential procedural learning and reading. The remaining five studies, including 376 participants (mean sample size = 94, SD = 22.49, range = 58 to 120), were entered into a meta-analysis that calculated the effect size *(r)* for the relationship between implicit learning on the serial reaction time task and measures of language and decoding (see Figure 4.7). The pooled effect size in this meta-analysis was very small and non-significant (r = 0.07, 95% CI [-0.03, 0.17]), with nonsignificant variability between samples (Q (3) = 2.66, p = 0.45, I^2 = 0.00%, Tau = 0.00) (r = .08, 95% CI [-0.02, 0.18]).

Three of these studies also contained sufficient information to calculate an effect size for the relationship between NVIQ and serial reaction time task implicit learning performance (see Figure 4.8). The overall effect size was also small, but it was significant (r = .12, 95% CI [0.004, 0.22], p = .04), with non-significant variability between samples (Q (2) = 97, p = 0.62). This indicated that lower NVIQ did reduce the level of learning on the serial reaction time task, but this factor did not account for any systematic variation in effect size between studies.

Table 4.4 Characteristics of the 6 studies eligible for the meta-analysis investigating correlational studies using the SRT task.

Study	Task(s)	Diagnosis	Age	Sample Size	Sequence length	Sequence Repetitions
Kidd, 2012	Det.		Child	100	10	24
Kidd & Kirjavainen, 2011	Det.		Child	120	10	24
Lum & Kidd, 2012	Det.		Child	58	10	24
Waber, Marcus, Forbes, Bellinger, Weiler, Sorensen, & Curran, 2003	Det.	Incl. DD	Child	422	6	50
West, Vadillo, Shanks, & Hulme, 2017	Prob.		Child	98	12	c.45

Studyname	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Kidd & Kirjavainen, 2011	-0.035	-0.213	0.145	-0.379	0.705	
Lum & Kidd, 2012	0.050	-0.211	0.305	0.375	0.708	
West, Vadillo, Shanks & Hulme, 2017	0.098	-0.107	0.295	0.933	0.351	
Kidd, 2012	0.183	-0.014	0.366	1.823	0.068	
	0.071	-0.032	0.173	1.346	0.178	
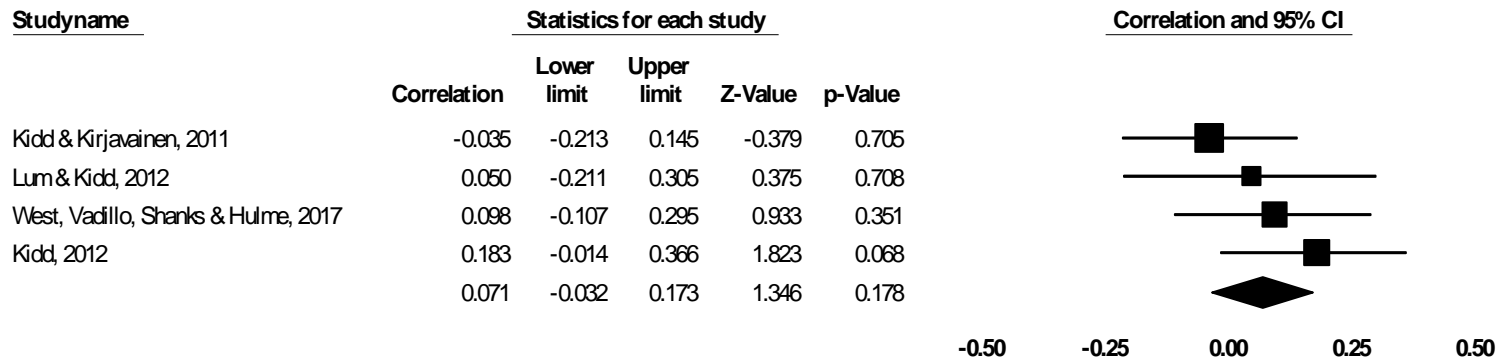
Figure 4.7 Forest plot showing effect sizes for the correlation between implicit learning on the serial reaction time task and language and decoding measure scores.

Studyname	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Kidd & Kirjavainen, 2011	0.069	-0.109	0.242	0.760	0.447	
West, Vadillo, Shanks & Hulme, 2017	0.091	-0.114	0.288	0.865	0.387	
Kidd, 2012	0.195	-0.001	0.377	1.945	0.052	
	0.115	0.004	0.224	2.035	0.042	
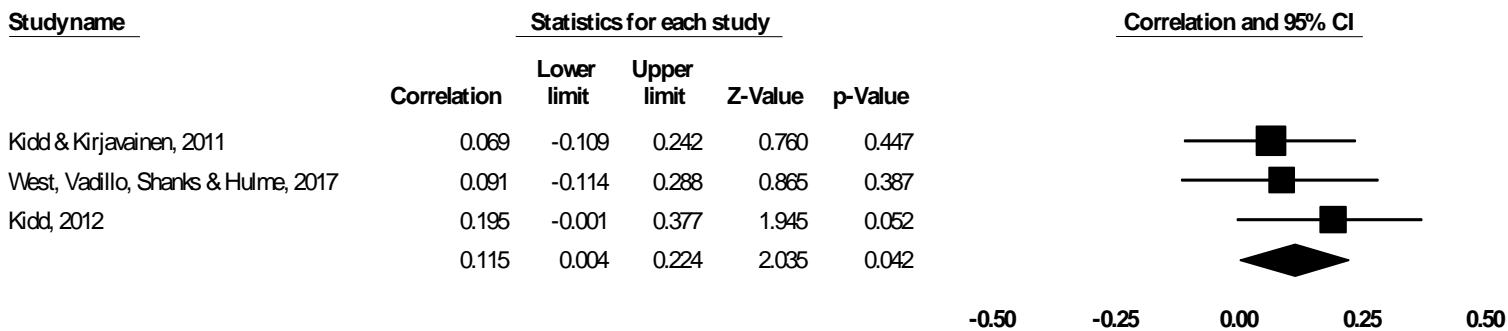
Figure 4.8 Forest plot showing effect sizes for the correlation between implicit learning on the serial reaction time task and NVIQ.

125

4.7 Discussion: Meta-analyses of serial reaction time task group and correlational designs

The meta-analysis of 19 comparisons of performance on serial reaction time tasks between language-disordered and age-matched control groups found a small but significant difference between groups. However, with nonsignificant variation in effect sizes and less than half of the eligible studies included in the meta-analysis, it was not possible to answer any questions about the extent of the relationship between procedural learning and language disorder. In spite of this, the meta-analysis has highlighted some important points.

The significant effect sizes in support of the procedural deficit hypothesis reported in previous meta-analyses are likely to be too high. Previous meta-analyses examining procedural learning on serial reaction time tasks by language-disordered and control groups by Lum et al. (2013; 2014) and Obeid et al. (2016) calculated their effect sizes using a single standard deviation for each group for the difference between sequenced and random trials, as set out in the method used by Siegert et al. (2006). The current results showed that using Siegert et al's (2006) method resulted in an overall effect size in line with previous findings. However, this calculation method is problematic for the reasons set out in the results section. Serial reaction time task group design studies most closely resemble pre- and post-test control group designs. Each group undertaking a serial reaction time task is measured on their response times to two types of trial (sequenced and random). These two trial types can be regarded in the same way as the two measures per group (pre- and post-test) in randomized controlled trials. The optimal method for calculating effect sizes for this experimental design divides the numerator of the effect size equation by the pooled standard deviations for the raw trials themselves not the pooled standard deviations for the difference between trial types (Morris, 2008), either pooling the pre-test standard deviations for the groups or both the pre- and post-test standard deviations (ie: either the sequenced trial standard deviations or the random and sequenced trial standard deviations for both groups).

1) Support for the selection of this method is convincing. Two papers caution against using standard deviations of differences to calculate effect sizes for pre- and post-test control group designs (Lund, 1988; Ray & Shadish, 1996), both demonstrating that it results in an overly small and underestimated denominator and an overestimated final effect size, as a result. Lund (1988) illustrated this by calculating effect sizes for several datasets using both methods (the pooled standard deviations for the within group difference and pooled 'post-test' standard deviations). The resulting effect sizes were very different and were far larger for the method using the difference standard deviations. Ray and Shadish (1996) also found that effect size calculations using difference standard deviations performed poorly, even when they incorporated pre- and post-test correlations. They substantially overestimated effect size, compared to using pooled standard deviations for raw scores. This was particularly the case when the two scores forming the difference standard deviation were highly correlated, as is the case in experiments using serial reaction time tasks. These findings were replicated by the two current meta-analyses of eight extreme group studies, each using data supplied in one of the above formats.

2) The moderating effects of variables such as participant age, diagnosis or task differences remains unclear. Although the final group design meta-analysis did not show any heterogeneity in effect sizes, this cannot be accepted at face value. The large variance estimates that formed part of the effect size calculation may serve to mask underlying true heterogeneity between the effect sizes in the model. The review table which shows the result of the principal measure of implicit learning for serial reaction time tasks for each of the 46 eligible extreme groups studies may be more useful than the meta-analysis in sign-posting the possible influence of moderators. For example, twenty three studies report a null result, but twenty three did not. Of those twenty three significant results seventeen were studies of children, which represents 55% of the total number of 31 studies with children and only 5 are studies of adults, representing 33% of the 15 studies in adults. An interesting question is whether the higher proportion of significant results in studies with children reflects a real underlying difference in procedural learning or whether it simply reflects lower reliability of the

tasks in this young age group and the confounding effect of noise in the data. It is also still possible that the significant variability in level of NVIQ and severity of language disorder between studies may explain some of the inconsistency in results.

Additionally, task differences may also explain some of the inconsistency in results. The majority of tasks used to examine the procedural deficit hypothesis are deterministic tasks ($n = 34$), with just a few studies selecting alternating versions (Desmottes et al., 2016a; 2016b; 2017; Hedenius et al., 2011; 2013; Henderson & Warmington, 2017; Howard et al., 2006); or probabilistic ones (Gabriel et al., 2011) in an effort to minimize explicit learning of the sequence. Deterministic tasks may be more open to explicit learning of the sequence (Shanks & Johnstone, 1999). The length of sequence used may also influence the extent of explicit learning on the tasks. Sequence length ranged from 5 to 12 items on deterministic serial reaction time tasks used to examine the procedural deficit hypothesis, although over half included sequences of 10 or 12 items and only 5 studies contained sequences of 6 items or less (Desmottes et al., 2016b; Jiménez-Fernández et al., 2011; Perlant & Largy, 2011; Stoodley et al., 2008; Vicari et al., 2003; Yang & Hong-Yan, 2011). However, all short sequences were in studies of children. Another factor that might affect the extent of implicit sequence learning in deterministic tasks is the number of repetitions of the sequence before the introduction of the random sequence that forms the baseline against which to measure the difference in response times. These ranged hugely from 10 (Stoodley et al., 2006) to 108 repetitions (Deroost et al., 2010). Although the mean number of repetitions was 40 ($SD = 22$), 11 of the studies included 20 or fewer repetitions of the sequence, with the vast majority of these administered to children ($n = 9$).

Finally, significant variability between studies was found in the severity of language disorder, as well as NVIQ discrepancy between groups. Once again, the influence of these differences could not be investigated, but it is still possible that they contribute to an explanation of the pattern of significant and non-significant results in the literature.

3) The failure to replicate group deficits in procedural learning on the serial reaction time task in correlational designs undermines the claims of the procedural deficit hypothesis, demonstrating that procedural learning is not a reliable correlate of language-related ability in unselected samples. Taken together, the results of both the group design and correlational meta-analyses suggest that studies using the serial reaction time task may not provide strong evidence for a procedural learning deficit in developmental disorders of language.

To conclude, despite the inclusion of only half of the eligible studies, the meta-analysis of group design studies should still be considered a valuable exercise. It indicates that the effect size estimates in previous meta-analyses are likely to be too high, as well as highlighting possible methodological issues with the task in a group design setting.

4.8 Artifical grammar and statistical learning tasks: Meta-analysis of comparisons of language-disordered groups and age-matched controls

Artificial grammar learning and statistical learning tasks present participants with strings of stimuli that conform to an undisclosed set of combinatory rules. Participants are subsequently asked to judge whether new strings conform to or violate these rules. The measure of implicit learning is the number of correct judgments made. Better than chance performance is taken to reflect implicit learning of the underlying combinatory rules.

Nineteen group design studies were eligible for this meta-analysis (see Table 4.5). Two studies were excluded as insufficient data were available to calculate an effect size for the tasks (Bahl et al., 2009; Plante et al., 2010). Results from two comparisons in Pavlidou and Williams (2014) were excluded as a duplicate of results in Pavlidou and Williams (2010). This meant that 16 out of 19 eligible studies were entered into the meta-analysis, which included 21 independent comparisons of artificial grammar learning and statistical learning tasks with language-disordered groups and age-matched controls. The studies included 477 participants with language disorder (mean

sample size 222.71, *SD* = 14.10, range = 12 to 77) and 695 control participants (mean sample size 33.10, *SD* = 33.69 range = 12 to 146).

The studies all used a separate offline testing phase to measure implicit learning in one of two ways. The first type of measure asked participants to judge whether they recognized sequences of items that they had seen during an earlier learning phase (seen items). The other type of measure asked them to judge whether sequences of items they had not seen before were consistent with the sequential rules followed during the learning phase (transfer items). The offline format for both types of test either used a two alternate forced choice (2AFC) structure, or presented test stimuli that were either correct or incorrect one at a time (50% of each type). Where overall group differences on the test phase of the artificial grammar or statistical learning task were reported, these were coded as a single measure for the study in CMA. Where several measures relating to different types of grammaticality performance on the test phase were reported separately, these were all coded and the mean effect size for all of them was taken.

Effect sizes with confidence intervals for the different studies are shown in Figure 4.9. The overall mean effect size was moderate and significant, *g* = -0.534, 95% CI [-0.79, -0.28] confirming that overall language disordered groups performed more poorly on artificial grammar learning and statistical learning tasks than age-matched controls without difficulties. The variation in effect sizes between studies was large and significant, Q (20) = 84.25, $p < .001$, $I^2 = 76.26\%$, $k = 21$, $Tau^2 = 0.26$. A sensitivity analysis showed that after removing outliers, the overall effect size was in the range of *g* = -0.50, 95% CI [-0.77, -0.24] to *g* = -0.55, 95% CI [-.82, -0.28].

Table 4.5 Characteristics of the 19 group design studies eligible for the meta-analysis using artificial grammar and statistical learning tasks.

Study	Task	Domain	Modality	Diagnosis	Age	Sample Sizes*	Additional Information
Aguilar & Plante, 2014[1]	SL	Verbal	Visual	DLD	Adult	12; 12 & 28; 28	2 separate comparisons
Bahl, Plante, & Gerken, 2009[2]	SL	Non-verbal	Auditory	DLD	Adult	15; 15 & 13; 13	2 separate comparisons
Evans, Saffran, & Robe-Torres, 2009[1]	SL	Experiment 1: Verbal; Experiment 2: Verbal & non-verbal	Auditory	DLD	Child	35; 78 & 15; 15	2 separate comparisons
Gabay, Theissen & Holt, 2015[1]	SL	Verbal & non-verbal	Auditory	DD	Adult	16; 16	
Hsu, Tomblin, & Christiansen, 2014[1]	SL	Verbal	Auditory	DLD	Child	20; 20 (in each comparison)	3 separate comparisons
Kahta & Schiff, 2016[1]	AGL	Verbal	Visual	DD	Adult	14; 15	
Laasonen, Vare, Oksanen-Hennah, Leppamaki, Tani, Harno, Hokkanen, Pothos, & Cleeremans, 2014[1]	AGL	Non-verbal	Visual	DD	Adult	36; 35	
Lukacs & Kemeny, 2014[1]	AGL	Verbal	Auditory	DLD	Child	28; 87	
Mainela-Arnold & Evans, 2014[1]	SL	Verbal	Auditory	DLD	Child	20; 20	
Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014[1]	SL	Verbal	Auditory	DLD	Child	18; 65	
Nigro, Jiminez-Fernández, Simpson, & Defior, 2016[1]	AGL	Experiment 1: Non-verbal; Experiment 2: Verbal	Visual	DD	Child	21; 21 & 21; 21	2 separate comparisons
Pavlidou, Kelly, & Williams, 2010[1]	AGL	Non-verbal	Visual	DD	Child	16; 16	
Pavlidou & Williams, 2010[1]	AGL	Non-verbal	Visual	DD	Child	16; 16	
Pavlidou & Williams, 2014[3]	AGL	Non-verbal	Visual	DD	Child	16; 16	

Pavlidou, Williams, & Kelly, 2009[1]	AGL	Non-verbal	Visual	DD	Child	16; 16	
Plante, Bahl, Vance, & Gerken, 2010[2]	AGL	Non-verbal	Auditory	DLD	Child	29; 29 & 16; 16	2 separate comparisons
Plante, Gomez, & Gerken, 2002[1]	SL	Verbal	Auditory	DD / DLD	Adult	16; 16	
Pothos & Kirk, 2004[1]	AGL	Non-verbal	Visual	DD	Adult	77; 146	
Rüsseler, Gerth, & Munte, 2006[1]	AGL	Verbal	Visual	DD	Adult	12; 12	

[*]Sample size, disordered group first;[1] = included in meta-analysis; [2] = insufficient data for inclusion in meta-analysis; [3] = Duplicate data

Study name	Statistics for each study					Hedges's g and 95% CI
	Hedges's g	Standard error	Lower limit	Upper limit	p-Value	
Pavlidou & Williams, 2010	-1.338	0.383	-2.088	-0.587	0.000	
Kahta & Schiff, 2016	-1.246	0.397	-2.023	-0.469	0.002	
Mainela-Arnold & Evans, 2014	-1.126	0.335	-1.782	-0.471	0.001	
Evans et al., 2009 (Expt2)	-0.979	0.380	-1.723	-0.234	0.010	
Plante et al., 2002	-0.929	0.364	-1.642	-0.216	0.011	
Pavlidou et al., 2009	-0.855	0.361	-1.562	-0.148	0.018	
Mayor-Dubois et al., 2014	-0.828	0.268	-1.352	-0.303	0.002	
Gabay et al., 2015	-0.809	0.359	-1.513	-0.105	0.024	
Pavlidou et al., 2010	-0.681	0.355	-1.376	0.015	0.055	
Aguilar & Plante, 2014 (Expt1)	-0.601	0.405	-1.395	0.192	0.137	
Aguilar & Plante, 2014 (Expt2)	-0.558	0.269	-1.086	-0.030	0.038	
Lukacs & Kemeny, 2014	-0.556	0.219	-0.985	-0.127	0.011	
Evans et al., 2009 (Expt1)	-0.480	0.205	-0.881	-0.079	0.019	
Nigro et al., 2016 (Expt2)	-0.439	0.307	-1.039	0.162	0.153	
Laasonen et al., 2014	-0.425	0.236	-0.887	0.036	0.071	
Russeler et al., 2006	-0.237	0.396	-1.012	0.539	0.549	
Nigro et al., 2016 (Expt1)	-0.204	0.304	-0.799	0.391	0.502	
Hsu et al., 2014 - LV task	-0.196	0.311	-0.806	0.414	0.529	
Hsu et al., 2014 - HV task	-0.093	0.315	-0.710	0.525	0.769	
Hsu et al., 2014 - MV task	0.044	0.310	-0.564	0.652	0.887	
Pothos & Kirk, 2004	0.658	0.144	0.375	0.940	0.000	
	-0.534	0.131	-0.792	-0.277	0.000	

Figure 4.9 Forest plot showing effect sizes for group difference in performance on artificial grammar learning and statistical learning tasks (displayed by ♦) with 95% confidence interval for each study.

The analysis of categorical moderator variables (see Table 4.5) showed that the difference between samples with dyslexia and samples with developmental language disorder was not significant, Q (1) = 0.07, p = .80, g (dyslexia) = -0.498, k = 10, 95% CI [-0.86, -0.13], g (language disorder) = -0.563, k = 11, 95% CI [-0.91. -0.22]. Although studies with adults showed a smaller effect size than studies with children, this moderator was not significant either, Q (1) = 0.23, p = .63, g (Adults) = -0.453, k = 8, 95% CI [-0.85, -0.05], g (Children) = -0.578, k = 13, 95% CI [-0.89, -0. 27]. The difference between studies using artificial grammar learning or statistical learning tasks was also not significant, Q (1) = 0.056, p = .81, g (AGL tasks) = -0.565, k = 11, 95% CI [-0.93, -0.20], g (SL tasks) = -0.502, k = 10, 95% CI [-0.88, -0. 13]. Finally, the difference between studies using verbal or non-verbal stimuli was examined. This required the exclusion of two studies that had administered tasks of more than one modality to the same participants (Evans et al., 2009; Gabay et al., 2015). Once again, the difference was not significant Q (1) = 0.298, p = .59, g (verbal) = -0.556, k = 12, 95% CI [-0.89, -0.22], g (non-verbal) = -0.395, k = 6, 95% CI [-0.86, -0. 07].

The extent to which group differences in language skills and decoding, respectively, related to the group difference in artificial grammar or statistical learning task performance was also analysed. For language skills, there was variation between the degree of difference between the language-disordered groups and the comparison group. The mean difference was g = -2.146, ranging from g = -2.46 to g = -1.306, but heterogeneity between the studies was not significant Q (7) = 6.90, p = .44.

For decoding there was a large variation between the degree of difference between the disordered and comparison group (these were predominantly studies investigating dyslexia), mean difference was g = -2.37, ranging from g = -7.85 to g = -0.63. For decoding ability there was significantly heterogeneity between the studies Q (9) = 88.76, p < .01, I^2 = 89.86%, k = 10, Tau^2 = 1.53. However, a meta-regression showed that the degree of severity of disorder in the disordered group did not explain significant variation in the relationship between decoding ability and implicit learning on the task, β = -.581, p = .77, k = 9, R^2 = 0.00. One comparison (Nigro et al., 2016:

Experiment 2) contained extremely large effect sizes for the difference between groups on measures of word and non-word reading accuracy ($g = -7.85$). However, excluding this study did not significantly change the results of the meta-regression.

Finally, there was variation between the disordered and comparison groups for the measures of NVIQ used in the studies, mean difference was $g = -0.387$, ranging from $g = -1.26$ to $g = -0.35$. This variation in effect sizes between studies was significant, Q (12) = 24.69, $p = 0.02$, $I^2 = 51.40\%$, $k = 13$, $Tau^2 = 0.10$. Once again, in spite of this variation, a meta-regression showed that the degree of disparity in NVIQ between groups did not explain significant variation in the relationship between decoding ability and implicit learning on the task, $\beta = 0.468$, $p = .20$, $k = 13$, $R^2 = 0.11$.

4.8.1 Publication bias

A funnel plot was used to determine the presence of publication bias in the studies included in the meta-analysis (see Figure 4.10). The x-axis of a funnel plot represents the magnitiude of the effect size, while the y-axis plots the dependent statistic from the meta-analysis. In the absence of publication bias the plot should form an inverted symmetrical funnel, while lack of symetricality denotes publication bias. The strength of this procedure is that it uses the identical data in the meta-analysis in order to investigate publication bias and is, therefore, entirely representative of the result of the meta-analysis itself. However, funnel plots for random effects models can be difficult to interpret visually (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). For this reason a trim and fill analysis (Duval & Tweedie, 2000) was used, which estimates the impact of publication bias, imputing the missing values that are needed in order to make the funnel plot symmetrical. In addition, this procedure calculates an adjusted overall effect size based on inclusion of these imputed studies. This indicated the presence of publication bias, suggesting that the true effect size in the meta-analysis should be much lower, adjusted point estimate $g = -.24$, 95% CI [-0.58, -0.11].
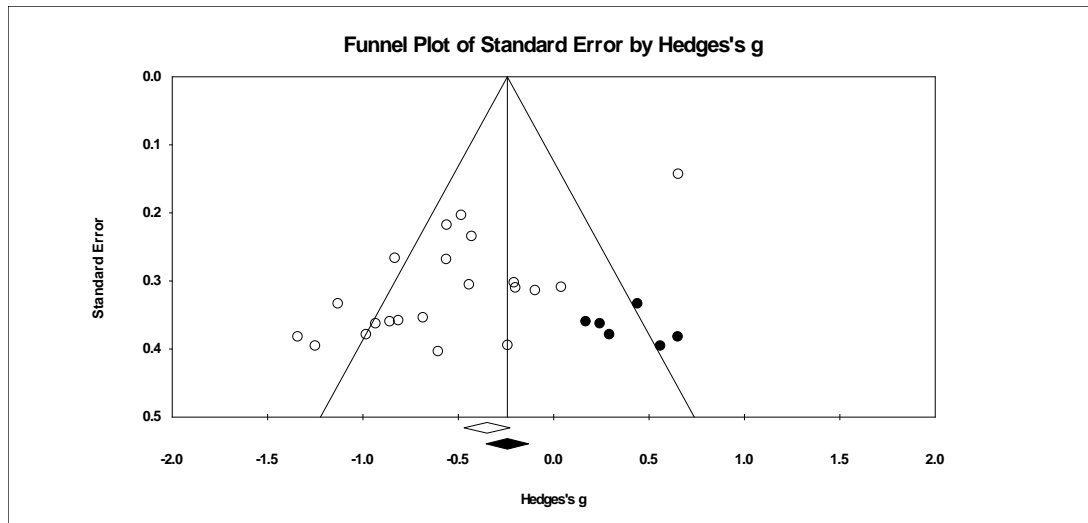
Figure 4.10 Funnel Plot showing evidence of publishing bias for artificial grammar and statistical learning studies in the meta-analysis. Open circles and diamond correspond to observed studies and point estimate. Filled circles and diamond correspond to imputed missing studies and adjusted point estimate, following Duval and Tweedie's (2000) Trim and Fill procedure.

A p-curve analysis was also undertaken to investigate whether the complete body of eligible studies was subject to publication bias. The p-curve focused only on overall group differences, since this is the effect size of interest in the meta-analysis. All 19 studies eligible for the meta-analysis were examined and a p-value for each study was coded that related specifically to this group difference (see Table 4.6). For the majority of studies this was an ANOVA main effect of group.

Three studies were categorized as non-significant for the purposes of the p-curve. These reported a non-significant main effect of group, but highlighted significant secondary group interactions: Aguilar and Plante (2014) reported differences in scores for correct and incorrect items; Kahta and Schiff (2016) reported similar differences; Nigro et al. (2016) reported differences in scores for transfer to unseen items. One study with significantly different group means was excluded because p-values related only to multiple regression analyses (Mainela-Arnold & Evans, 2014). Another study reported a significant effect, but in the opposite direction, with the dyslexic group performing better than controls (Pothos & Kirk, 2004). This study was, therefore,

categorized as a null result for the purposes of the p-curve analysis. Finally, two studies contained results on more than one task. Pavlidou and Williams (2010) reported a significant main effect for each of two tasks taken by the same participants. Evans et al. (2009) gave a second task to a subset of the same participants. As recommended by Simonsohn et al. (2013), a p-curve was run for the values from the first tasks and a second analysis was run that included the values for the second tasks. The results for the two p-curves were equivalent, so only the first one is reported here.

There were 11 significant values for the 19 studies eligible for the meta-analysis that could be entered into the p-curve. Figure 4.11 shows a right-skewed p-curve, demonstrating evidential value ($Z = 3.16$, $p = .0008$) and no reliable evidence that the studies' evidential value is inadequate due to low power (power estimate = 57%, 90% CI [21%, 83%].
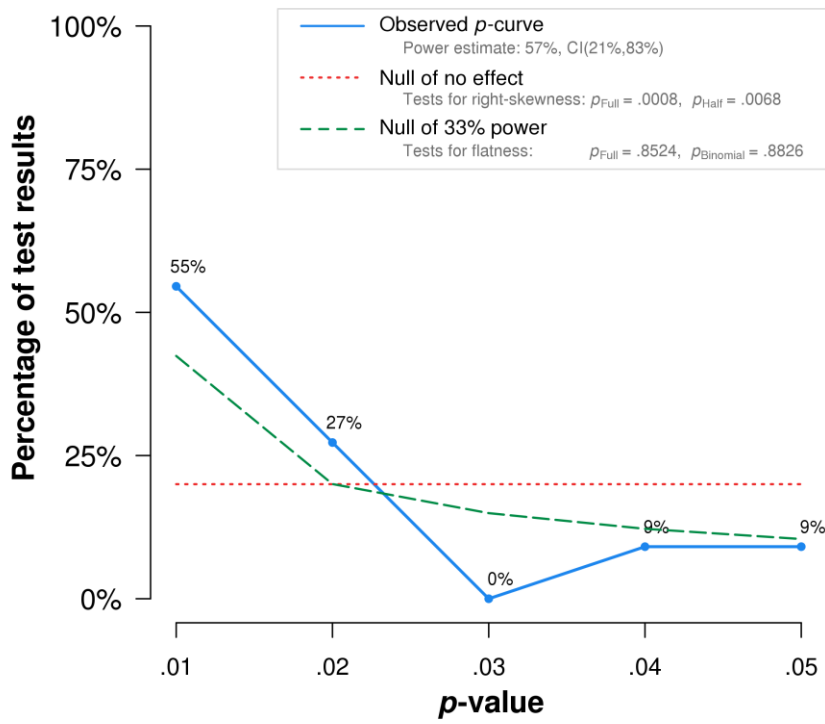
Note: The observed p-curve includes 11 statistically significant (p < .05) results, of which 9 are p < .025. There were no non-significant results entered.

Figure 4.11 P-curve examining p-hacking in extreme groups studies using artificial grammar learning or statistical learning tasks to investigate the procedural deficit hypothesis

Table 4.6 Disclosure table for the 19 group design studies eligible for the meta-analysis using artifical grammar and statistical learning tasks.

Study name	Model	Quoted test from paper with statistical results	Significance of main effect
Aguilar & Plante, 2014 (Expt 1)	Mixed ANOVA: 2 (Group) x 4 (item type: correct seen, correct generalization, co-occurence violation; linear order violation)	"The main effect for Group was not significant, $F(1,22) = .43$, $p = .5186$, $\pi^2 p = .02$, nor was the Group x Item Type interaction." (p. 1398)	Null.
Aguilar & Plante, 2014 (Expt 2)	Mixed ANOVA: 2 (Group) x 4 (item type: correct seen, correct generalization, co-occurence violation; linear order violation)	"The main effect of group was not significant, $F(1,54) = 2.49$, $p = .12$, $\pi^2 p = .04$. […] This was qualified by a significant Group x Item Type interaction, Wilk's $F(1,162) = 69.03$, $p = .0116$, $\pi^2 p = .07$. […] this reflected a general pattern for the NL group to accept more correct items than the LLD group, whereas the LLD group tended to accept more incorrect items than their NL counterpart." (p. 1400)	Null for main effect of group, significant for Group x item type interaction.
Bahl, Plante, & Gerken, 2009 (Expt 1)[1]	Mixed ANOVA: 2 (Group) x 2 (language A vs B) x 2 (item type (correct vs incorrect) x 2 (generalization type - pattern or principle) x 2 (item type - correct & incorrect)	"The ANOVA revealed a significant main effect of group, $F(1,25) = 9.16$, $p < .005$, $\pi^2 p = .276$, with hLLD group accepting more items overall than the NL group." (p. 317)	Significant. Insufficient data for meta-analysis.
Bahl, Plante, & Gerken, 2009 (Expt 2)	Mixed ANOVA: 2 (Group) x 2 (item type (correct vs incorrect) x 2 (generalization type: pattern or principle) x 2 (item type - correct & incorrect)	"There was no significant main effect for group, $F(1,24) = 1.39$, $p < .25$, or generalization type." (p. 319)	Null. Insufficient data for meta-analysis.

Evans, Saffran, & Robe-Torres, 2009 (Expt 1) [1]	ANCOVA: 2 (Group) with Age & NVIQ as covariates	"An analysis of covariance with age and nonverbal IQ as covariates revealed that the SLI group's ability to attend to transitional probabilities in the speech stream was significantly poorer than the NL group's, $F(1,109) = 5.6$, $p < .01$, $\pi^2 p = .05$." (p. 7)	Significant
Evans, Saffran, & Robe-Torres, 2009 (Expt 2)	Mixed ANCOVA: 2 (Group) x 2 (Task variant - Speech or Tone) with Age and NVIQ as covariates	"A repeated measures ANCOVA with age and nonverbal IQ as covariates revealed a main effect for group, $F(1,26) = 7.4$, $p = .003$, $\pi^2 p = .37$, across the speech and tone conditions, with overall performance for the children with SLI being poorer than that of their typical language peers. (p 9)	Significant
Gabay, Theissen & Holt, 2015 [1]	Mixed ANOVA: 2 (Group) x 2(SL task variant)	There was a main effect of group, $F(1,30) = 10.366$, $p = .003$, $\pi^2 p = .256$), indicating that the DD group performed significantly less accurately (M = 69%) than the control group (M = 85%). (p. 939)	Significant
Hsu, Tomblin, & Christiansen, 2014 [1]	Mixed ANOVA: 2 (Group) x 3 (variability condition) x 2 (grammaticality)	"There was a significant main effect of grammaticality [...] and Grammaticality x Language Group interaction, $F(1,114) = 6.34$, $p = 0.01$, $\pi^2 p = .05$." (p. 4)	Significant
Kahta & Schiff, 2016	Mixed ANOVA: 2 (Group) x 2 (grammaticality score (G vs NG)	"No significant main effect was found for group (F 1<). However, there was a significant interaction for grammaticality x group, $F(1, 27) = 11.86$, $p = .002$, $\pi^2 p = .3$." (p. 241)	Null for main effect of group. Significant for Group x Grammaticality interaction.
Laasonen, Vare et al., 2014	Mixed ANCOVA: 3 (Group) x 2 (answer type: Accuracy vs Similarity)	"A 3 x 2 mixed ANCOVA with Group as a between subjects factor, answer type as a within subjects factor and proportion of correct responses as the dependent variable resulted in a non-significant main effect of group ($F(2,84) = 2.416$, $p = .095$, $\pi^2 p = .054...$" (p. 22)	Null.

Study	Analysis	Result	Conclusion
Lukacs & Kemeny, 2014[1]	Univariate ANOVA (Group on performance difference score)	"The control group outperformed the clinical group, as revealed by a significant main effect of group, $F(1,113) = 6.645$, $p < .05$, $\pi^2 p = 0.056$." (p. 478)	Significant
Mainela-Arnold & Evans, 2014	Analyses relate to whether SL ability predicts performance on lexical gating and definition tasks: Multiple regression with age, NVIQ, SL, Group , Group x SL interaction	From table: predicting lexical phonology: Group x statistical learning interaction: ß = -.08, R^2 = .27, R^2 change = .01, F change = .52; predicting lexical-semantics: Group x statistical learning interaction: ß = .36, R^2 = .46, R^2 change = .00, F change = .15	N/A
Mayor-Dubois, Zesiger et al., 2014[1]	*T*-test for Group difference	"Significant difference in scores between the SLI and the Control groups, $t(77) = 3.137$, $p < .01$. The performance of the SLI group did not differ from chance level […], contrary to the Control Group who obtained scores above the chance level..." (p. 18)	Significant
Nigro, Jiminez-Fernández et al., 2016 (Expt 1)	*T*-test by Group against chance	"...participants from the TD group performed above chance level in all three cases […] $t(20) = 3.85$, $p = .001$, $r = .65$ […] Participants with DD also performed above chance level in the overall task […] $t(20) = 3.20$, $p = .005$, $r = .58$." (p. 208)	Null for overall difference, but significant difference with transfer to unseen items.
Nigro, Jiminez-Fernández et al., 2016 (Expt 2)	*T*-test by Group against chance	"Results from single-sample t-tests showed that participants from the TD group again performed above chance level in all three cases (overall […]: $t(20) = 4.06$, $p = .001$, $r = .67$). […] Participants with DD also performed above chance level in the overall task (…$t(20) = 3.07$, $p = .006$, $r = .57$)." (p. 211)	Null for overall difference, but significant difference with transfer to unseen items.

Pavlidou, Kelly, & Williams, 2010[1]	Mixed ANOVA: 2 (Group) x 2 (Grammaticality) x 2 (Chunk strength)	"The between subjects ANOVA revealed a main effect of Participant type ($F(1,30) = 4.521$, $p < .05$, p-value reported two-tailed): the two types of children were performing significantly different…" (p. 152)	Significant
Pavlidou & Williams, 2010[1]	Both models: Mixed ANOVA: 2 (group) x 2(grammaticality) x 2 (chunk strength)	Non transfer task: "Between subjects ANOVA revealed an effect of group ($F(1,30) = 14.46$, $p = .001$): The typical group outperformed the dyslexic group." (p . 3292) Transfer task: "Between subjects tests showed a group effect ($F(1,30) = 4.63$, $p < .05$). The two groups of children were performing significantly different during the testing phase…" (p. 3294)	Both significant
Pavlidou & Williams, 2014	Mixed ANOVA: 2 (Group) x 2 (Grammaticality) x 2 (Chunk strength)	"A main effect of reader Group was obtained ($F(1,30) = 14.46$, $p = .0001$), with higher number correct for typically developing children […] than dyslexic children..." (p. 1462) Transfer task: "A main effect of reader Group was obtained ($F(1,30) = 4.63$, $p < .05$), such that grammaticality-decisions for the test items were more accurate for TD […] than DD children..." (p. 1465)	Both significant (Same experiment as Pavlidou and Williams (2010), so not included)
Pavlidou, Williams, & Kelly, 2009[1]	Mixed ANOVA: 2 (Group) x 2 (Grammaticality) x 2 (Chunk strength)	"The ANOVA revealed a main effect of group ($F(1.30) = 8.18$, $p < .01$)." (p. 63)	Significant
Plante, Bahl, Vance, & Gerken, 2010 (Expt 1)	Mixed ANOVA: 2 (Group) x 2 (generalization type - pattern or principle) x 2 (item type - correct & incorrect)	"No other main effect or interaction effect was significant." Significant effect were not considered relevant to implicit learning by authors (see p. 402)	Null. Not in meta-analysis.

Plante, Bahl, Vance, & Gerken, 2010 (Expt 2)	Mixed ANOVA: 2 (Group) x 2 (generalization type - pattern or principle) x 2 (item type - correct & incorrect)	"No other effect was significant […] the variance that contributed to the the three-way interaction occurred only because incorrect items were accepted more frequently than correct items under certain conditions. (p. 403)	Null. Not in meta-analysis.
Plante, Gomez, & Gerken, 2002[1]	*T*-Test for Group difference	"In contrast, the NLD average [...] was both above chance levels and significantly greater than the mean of the L/LD group ($t(30) = 2.75$, $p = .01$)." (p. 458)	Significant
Pothos & Kirk, 2004	Mixed ANOVA: 2 (Group) x 2 (Task variant)	"There was a main effect for the factor Dyslexia ($F(1,210) = 4.39$, $p = .04$), showing that dyslexic participants performed better than non-dyslexic ones…" (p. 71)	Effect in opposite direction
Rüsseler, Gerth, & Munte, 2006[1]	3 (Group) ANOVA on grammaticality judgements	"…both the normal and the dyslexic readers' classification scores exceeded that of the random comparison group [...] main effect GROUP: $F(2,33) = 23.94$, $p < .0001$…" (p. 819)	Significant

[1] = Included in p-curve

4.9   Statistical learning: Meta-analysis of correlational studies

Three studies examined the relationship between performance on statistical learning tasks and language ability using correlational designs, with a total of five independent samples (see Table 4.7). No eligible correlational studies using artificial grammar tasks were found.

The three studies, including 177 participants (mean sample size = 44.25, *SD* = 16.58, range = 30 to 68) were entered into a meta-analysis that calculated the effect size *(r)* for the relationship between implicit learning on the statistical learning task and measures of language and decoding (see Figure 4.12). The overall effect size in this meta-analysis was moderate and significant (*r* = .311, 95% CI [0.17, 0.44]). The variability across samples was not significant (Q (3) = 0.087, p = 0.99), indicating that there was no case for examining the relationship separately for language or decoding.

Two of these studies also contained sufficient information to calculate an effect size for the relationship between NVIQ and statistical learning task performance. The overall effect size was also was not significant (*r* = .157, 95% CI [-0.05, 0.35], p = .13). This indicated that NVIQ did not significantly relate to the level of learning on the statistical learning task.

Table 4.7 Characteristics of the 3 correlational studies eligible for the meta-analysis using the statistical learning task.

Study	Task	Domain	Modality	Age	Sample Size	Additional Information
Arciuli & Simpson, 2012	SL	Non-verbal	Visual	Expt 1: Child Expt 2: Adult	Expt 1: 42 Expt 2: 37	2 separate comparisons
Kidd & Arciuli, 2016	SL	Non-verbal	Visual	Child	68	
Misyak & Christiansen, 2012	SL	Verbal	Auditory	Adult	30	2 tasks (adjacent & non-adjacent dependencies)

Study name		Statistics for each study				Correlation and 95% CI
		Correlation	Lower limit	Upper limit	p-Value	
Misyak & Christiansen, 2012		0.278	-0.091	0.580	0.137	
Kidd & Arciuli, 2016		0.300	0.066	0.502	0.013	
Arciuli & Simpson, 2012 (Experiment 1)		0.327	0.026	0.574	0.034	
Arciuli & Simpson, 2012 (Experiment 2)		0.338	0.016	0.597	0.040	
		0.311	0.167	0.441	0.000	

-1.00   -0.50   0.00   0.50   1.00

Figure 4.12 Forest plot showing effect sizes for the correlation between statistical learning tasks performance and language and decoding measure.

4.10 Discussion: Artificial grammar and statistical learning meta-analyses

The meta-analysis of group design tasks using artificial grammar or statistical learning tasks returned a significant moderate overall effect size. The results also indicate that the moderate effect size in the group design meta-analysis may be the result of publication bias. Although the p-curve that examined publishing bias across all eligible studies suggested there was no evidence of bias in the data, the funnel plot did indicate publication bias and suggested that a far smaller overall effect size may be closer to reality. The funnel plot is based on the actual measures entered into the meta-analysis, rather than on the single reported statistic that best represents learning (such as a *F* ratio or *t* statistic), so while the p-curve suggests that the 20 studies accurately report a balanced and fully representative view, the funnel plot suggests that studies with non-significant results are likely to exist, but are missing from the published literature. This concurs with the recent meta-analysis by Schmalz et al. (2016), who conjectured that the true effect size in artificial grammar tasks was likely to be small. The result of the small meta-analysis of three correlational studies, showing a low to moderate overall effect size, lends further support to this suggestion. It is far closer to the estimate suggested by the funnel plot analysis than to the higher overall effect size in the group design analysis.

Why should the pooled effect size for this meta-analysis be so much larger than the pooled effect size from the meta-analysis of serial reaction time tasks? One possibility is that offline measures that do not require the derivation of a difference score are far less noisy and, therefore, allow a more precise estimation of effect size. An alternative view is that the tasks index an aspect of implicit statistical learning that is indeed impaired in language disorder, while the serial reaction time tasks do not. However, it is also possible that the significant results reflect the artifacts of a third variable, such as, attention (de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016). Additionally, these meta-analyses included results for both seen and transfer items on the tasks. It could be argued that scores for recognition of seen items are at least partially declarative measures and that only transfer items represent the implicit

146

statistical learning of abstract grammatical rules. If this is the case, then the overall mean effect size includes elements of both declarative and statistical learning.

4.11 Hebb serial order learning: Meta-analysis of language-disordered groups and age-matched controls

The Hebb serial order learning task asks participants to recall sequences of items in order. The task then introduces a covert repeated sequence. Better recall for the repeated as opposed to random sequences is considered evidence of implicit learning.

There were nine eligible studies for this meta-analysis. These studies analysed implicit learning on Hebb tasks in two different ways. The first method compared the gradient of the regression line for performance on Hebb trials to the gradient for random trials, while other studies chose to compare overall accuracy rates for the Hebb and random sequences. In order to include as many studies investigating Hebb performance and language disorder as possible using a consistent measure, the meta-analysis compared overall accuracy rates, rather than regression-based accuracy measures. Analysing data in this way meant excluding the reported measures for two experiments (Bogaerts et al., 2015), that administered tasks of differing lengths to participants dependent on the length of time taken to reach a criterion of two correct repetitions of the Hebb sequence. Instead, data kindly supplied by the study's authors was used, which gave a measure for the length of the tasks completed by all participants. Only one paper, which predated the procedural deficit hypothesis by 15 years, was excluded from the meta-analysis because it contained insufficient information to calculate an effect size (Gould & Glencross, 1990). This paper reported no significant differences in Hebb learning between groups of normal and poor readers. In this study differences in mean scores for Hebb and filler trials for groups of normal and poor readers were equivalent on a visuospatial Corsi blocks Hebb task, as were scores on the second half of verbal-visual Digits Hebb task. The only difference in performance between groups was on the first half of the Digits task, where normal readers showed better performance on early Hebb trials while poor readers did not.

The meta-analysis, therefore, included eight studies of Hebb serial order learning tasks with language-disordered groups and age-matched controls, which contained ten independent comparisons in total (see Table 4.8). All 10 comparisons calculated effect sizes using the standard deviation for the control group of the unrepeated condition, not standard deviations for the differences between conditions.

The studies included 200 participants with a diagnosis of language disorder (mean sample size = 22.22, SD = 5.74, range = 12 to 29) and 201 control participants (mean sample size = 22.33, SD = 6.58, range = 12 to 32). Effect sizes with confidence intervals for the different studies are shown in Figure 4.13. The overall mean effect size g = -0.32, 95% CI [-0.52, -0.12], $p < .01$, with language disordered groups in these studies showing less facilitation on repeated lists compared to age-matched controls without language difficulties. The variation in effect sizes between studies was not significant Q (9) = 10.36, $p = .32$, I^2 = 13.10%, Tau^2 = 0.01.

Table 4.8 Characteristics of the 9 group design studies eligible for the meta-analysis using the Hebb serial order learning task.

Study	Diagnosis	Age	Sample Size*	Modality	Total trials (Hebb trials)	Additional Information
Archibald & Joanisse, 2013	DLD	Child	23; 27	Verbal (visual & auditory)	84 (42)	3 sessions
Bogaerts, Szmalec, Hachmann, Page & Duyck, 2015	DD	Adult	Expt 1: 25; 23 Expt 2: 18; 18	Verbal-visual	Expt 1: 9 (3)** Expt 2: 18 (6)**	2 separate comparisons
Bogaerts, Szmalec, De Maeyer, Page & Duyck, 2016	DD	Child	23; 23	Verbal-visual; visuospatial		
Gould & Glencross, 1990	DD	Child	18; 18	Verbal–visual; Visuospatial	32 (10)	2 tasks
Henderson & Warmington, 2017	DD	Adult	29; 30	Verbal-auditory	26 (8)**	Main testing session only
Hsu & Bishop, 2014	DLD	Child	28; 20	Verbal-visual	13 (5)	
Majerus, Leclercq, Grossmann, Billard, Touzin, Van der Linden, & Poncelet, 2009	DLD	Child	12; 12	Verbal-auditory	24 (8)	Expt 2 only
Staels, Van der Broek, 2015	DD	Expt 1: Adult Expt 2: Child	26; 32	Verbal-visual; Verbal-auditory; Visuospatial	30 (10)	2 comparisons. 3 tasks in each
Szmalec, Loncke, Page, & Duyck, 2011	DD	Adult	16; 16	Verbal- visual; Verbal-auditory; Visuospatial	30 (10)	3 tasks

*= Sample size disordered group first; ** = Length of task taken by all participants

Study name	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Hsu & Bishop, 2014	-0.865	0.301	-1.455	-0.274	0.004
Szmalec et al., 2011	-0.715	0.357	-1.414	-0.016	0.045
Bogaerts et al., 2015 (Expt 2)	-0.568	0.337	-1.229	0.093	0.092
Bogaerts et al., 2016	-0.524	0.299	-1.111	0.062	0.080
Staels, & Van den Broeck, 2015 (Expt 1)	-0.327	0.262	-0.841	0.186	0.211
Bogaerts et al., 2015 (Expt 1)	-0.262	0.285	-0.821	0.298	0.359
Archibald & Joanisse, 2013	-0.135	0.280	-0.684	0.414	0.630
Henderson & Warmington, 2017	-0.129	0.259	-0.636	0.378	0.617
Staels, & Van den Broeck, 2015 (Expt 2)	-0.048	0.263	-0.565	0.468	0.854
Majerus et al., 2009	0.341	0.397	-0.438	1.119	0.391
	-0.319	0.101	-0.517	-0.122	0.002

Figure 4.13 Overall average effect size for the group difference in performance on Hebb tasks (displayed by ♦) with 95% confidence interval for each study.

150

4.11.1 Publication bias

A funnel plot for the random effects model of the 10 comparisons was used to examine whether there was evidence for the existence of publication bias in extreme groups studies of Hebb learning (see Figure 4.14). The Trim and Fill procedure (Duval & Tweedie, 2000) did not find evidence of publication bias. However, given the low number of studies entered into the plot it should be interpreted with caution. Lau, Ionnidis, Terrin, Schmid, and Olkin (2006) do not recommend the use of funnel plots in meta-analyses with less than 10 comparisons.
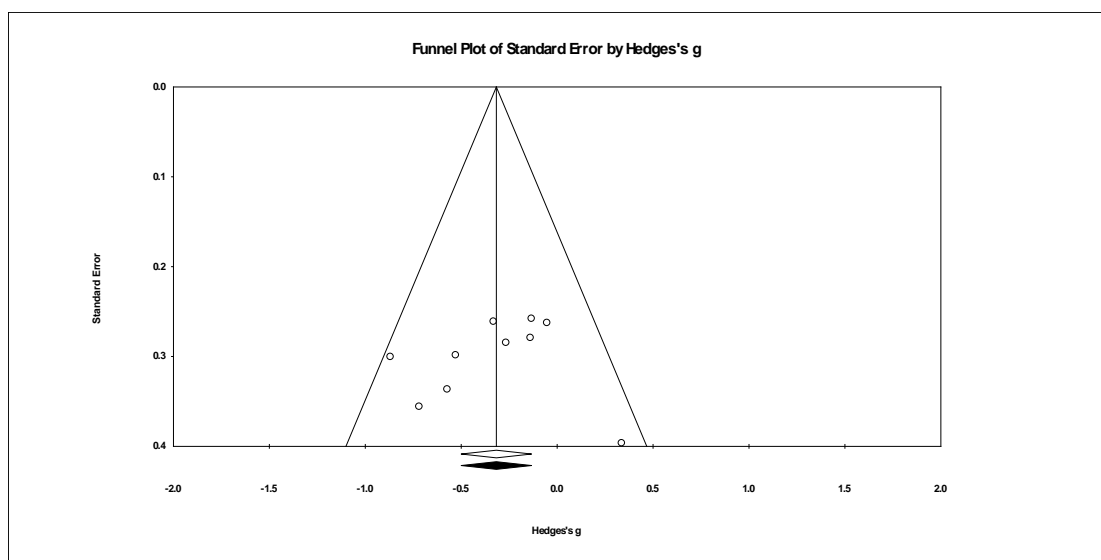
Figure 4.14 Funnel plot for the Hebb learning group design random effects model. Open circles and diamond correspond to observed studies and point estimate. Duval and Tweedie's Trim and Fill procedure shows no evidence of publication bias.

The p-curve for this set of studies coded the principal measure of learning on the Hebb task, according to each study (see Table 4.9). This included several regression-based measures that indicate improving recall for the Hebb sequence over time, as well as measures that related to an overall group difference in performance across the task. This enabled the includion of studies that only reported regression-based inferential statistics. However, the inclusion of both types of measure should be kept in mind when interpreting the result of the p-curve. The low number of studies is also a

151

concern. Eight studies contained sufficient data for the analysis, with three studies providing a significant statistic that represented a different gradient of improvement in implicit learning over the course of the task for the two groups and two studies indicating an overall difference in improvement. Figure 4.15 shows a right-skewed p-curve, demonstrating evidential value ($Z = -4.47$, $p = .0001$) and no reliable evidence that the studies' evidential value is inadequate due to low power (power estimate = 93%, 90% CI [68%, 99%].
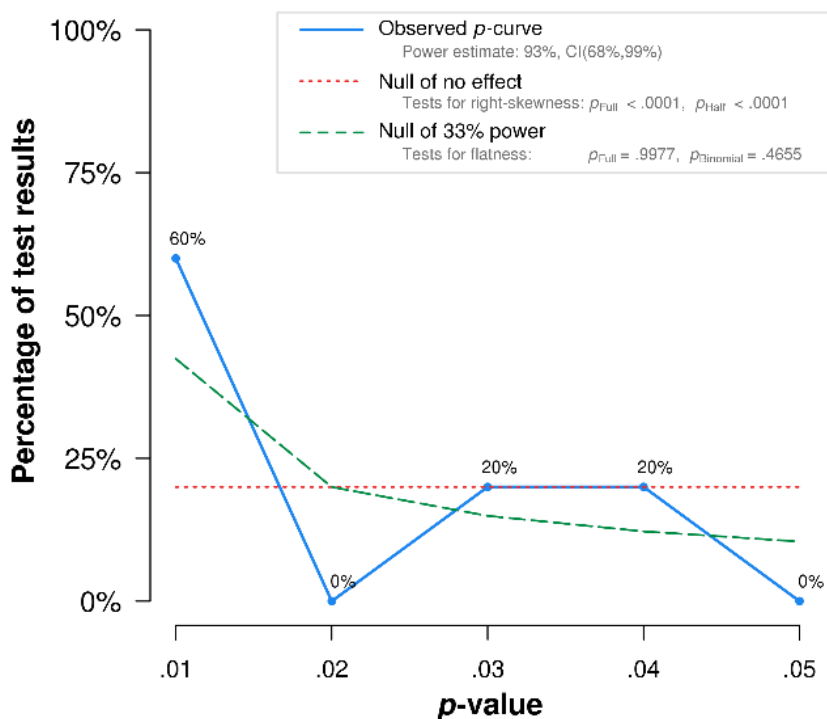
Note: The observed p-curve includes 5 statistically significant (p < .05) results, of which 3 are p < .025. There were no non-significant results entered.

Figure 4.15 P-curve examining publishing bias in extreme groups studies using Hebb tasks to investigate the procedural deficit hypothesis.

Table 4.9 Disclosure table for the 9 group design studies eligible for the meta-analysis using Hebb serial learning tasks

Study name	Analysis	Quoted test from paper with statistical results	Significance
Archibald & Joanisse, 2013[1]	ANCOVA: 2 (Group) x 2 (task modality) x 2 (sequence type) x 2 (Task half), with WM and NVIQ as covariates	"The results revealed two significant interactions with group: the interaction between modality and group […] .all remaining effects and interactions involving group were not significant […] Importantly this interaction was not differentiated by list types, indicating a general auditory retention difficulty rather than a specific deficit in carryover learning on the Hebb lists." (p. 274)	Null
Bogaerts et al., 2015 (Expt 1)[2]	Mixed ANOVA: 2 (Group) x 3 (Task) x 2 (Sequence type)	"Crucially, we found a significant interaction between Sequence type and Group, $F(1,46) = 4.73$, $p < .05$, $\pi^2 p = .09$. Planned comparisons indicate a HRL effect in both groups, however, HRL was significantly stronger for controls." (p. 111)	Significant for development of implicit learning over task
Bogaerts et al., 2015 (Expt 2)[2]	Mixed ANOVA: 2 (Group) x 3 (Task) x 2 (Sequence type)	"… a significant interaction was found between Sequence type and Group, $F(1,34) = 5.52$, $\pi^2 p = 0.14$, $p < .05$." (p. 115)	Significant for development of implicit learning over task
Bogaerts et al., 2016	Mixed logit models (Jaeger, 2008): Fixed vs = Group, Sequence type, task, block, NVIQ as control variable	"A group difference in the disadvantage of the poor readers would surface as a threeway interaction, Type x Presentation x Group, with a negative coefficient…A simple slopes analysis...suggesting that Hebb learning is present in both groups but to a lesser extent for the poor readers, $chi^2(2) = 56.04$, $p < .001$." (p. 146)	Significant for development of implicit learning over task

Gould & Glencross, 1990[1]	Verbal task Mixed ANOVA: 2 (Group) x 2 (sequence type) x 2 (early vs late trials)	"Table 2 shows that Normal Readers were more accurate on the repeated sequences in both the Early and Late Trials whereas the Disabled Readers did not show greater accuracy until the Late Trials." Table 2: Group x sequence interaction effect = ns; Group x sequence x trials: $F(1,18) = 8.6$, $p < .009$ (p. 275)	Null for consistent measure
	Visuospatial task Mixed ANOVA: 2 (Group) x 2 (sequence type) x 2 (early vs late trials)	"Table 3 shows that the pattern of results was very similar for both groups." (p. 275)	Null
Henderson & Warmington, 2017[1]	Mixed ANOVA: 2 (Group) x 2 (sequence type) x 2 (1st half vs 2nd half)	"...a marginally significant List x Half x Group interaction ($F(1,57) = 3.99$, $p = .051$, $\pi^2 p = .07$." NB: Group x sequence type = ns (p. 202)	Null for consistent measure
Hsu & Bishop, 2014[2]	3 (Group) ANCOVA, with Random gradient as covariate	"There was a significant effect of group, $F(2,76) = 3.68$, $p = .03$, $\pi^2 p = .09$. Pair-wise comparisons indicated that the age-matched group showed a steeper learning rate of word sequences than the SLI and the grammar-matched group." (pp. 357, 358)	Significant for development of implicit learning over task
Majerus et al., 2009[1]	Mixed ANOVA: 3 (Group) x 2 (Sequence type)	"This analysis revealed no significant group effect, $F(2,33) = 1.14$, ns [....] and no interaction effect, : $F(2,33) < 1$, ns." (p. 714)	Null
Staels, & Van der Broek, 2015 (Expt 1) [2]	Mixed ANOVA: 2 (Group) x 3 (Task) x 2 (Sequence type)	"Unlike Szmalec et al. (2011), however, the crucial Group x Sequence type interaction effect was not significant, $F(1,57) = .128$, $p = .722$, $\pi^2 p = .002$, indicating a similar Hebb effect for the control and the dyslexic group. Planned comparisons [...] confirmed the absence of a differential Hebb effect for [all 3 tasks]." (p. 6)	Null
Staels, & Van der Broek, 2015 (Expt 2) [2]	Mixed ANOVA: 2 (Group) x 3 (Task) x 2 (Sequence type)	The crucial Group x Sequence type interaction effect was also not significant, $F(1,55) = .087$, $p = .769$, $\pi^2 p = .002$, indicating a similar Hebb effect for the control and dyslexic group. (p. 13)	Null

| Szmalec et al., 2011[2] | Mixed ANOVA: 2 (Group) x 3 (Task) x 2 (Sequence type) | "The crucial interaction effect between Group and Sequence Type was significant, $F(1,30) = 23.22$, $p < .001$, $\pi^2 p = .44$, indicating a stronger Hebb effect for the control group. Further planned comparisons [...] demonstrate that the persons with dyslexia showed reduced Hebb learning for all stimulus and presentation modalities." (p. 12) | Significant for development of implicit learning over task |

[1] = Mean proportion of correct responses for Hebb vs Random; [2] = Repeated regression line compared to random one

4.12 Discussion

The pooled effect size for this meta-analysis was small but significant and was a similar size to the pooled effect size estimate from the meta-analysis of serial reaction time group design studies. Additionally, neither the funnel plot nor the p-curve analysis indicated any evidence of publishing bias in the literature. It would be useful to be able to compare the results of this meta-analysis with one that contains studies using correlational designs, as agreement between them would lend further support to this finding. However, such studies do not currently exist.

4.13 Weather prediction task: Meta-analysis of comparisons of language-disordered groups and age-matched controls

The weather prediction task asks participants to classify combinations of four possible stimuli into one of two possible outcomes. The stimuli each have a fixed probability of a certain outcome. A trial is scored correct if it accords with the conditional probabilities of the stimuli shown. Above chance performance is taken as evidence of implicit learning.

Five out of six eligible studies (see Table 4.10) were entered into this meta-analysis, resulting in 5 independent comparisons of weather prediction task performance with language-disordered groups and age-matched controls. One study was excluded, as there was not sufficient data to calculate an effect size (Lee, Mueller, & Tomblin, 2016). This study reported significantly poorer performance on the task for the language-disordered group compared to the control group. The studies in the meta-analysis included 101 participants with language disorder (mean sample size = 20.02, *SD* = 5.81, range = 15 to 29) and 208 control participants (mean sample size = 41.60, *SD* = 32.58, range = 15 to 87). Effect sizes with confidence intervals for the different studies are shown in Figure 4.16. The overall mean effect size was significant, *g* = -0.629, 95% CI [-1.07, -0.19] indicating that overall language disordered groups perform poorly on weather prediction tasks compared to age-

matched controls without language difficulties. The variation in effect sizes between studies was also significant, Q (4) = 11.79, $p < .02$, $I^2 = 66.09\%$, $k = 5$, $Tau^2 = 016$.

Only one study (Gabay et al., 2015) tested dyslexic participants, so participant diagnosis was not examined as a moderator variable. The moderating effect of participant age was examined, even though power in this analysis was low (Adults $k = 2$, Children $k = 3$). The difference between studies with adults and with children was not significant, Q (1) = 0.219, $p = .64$, g (Adults) = -0.79, 95% CI [-1.59, 0.01], g (Children) = -0.55 95% CI [-1.17, 0.07].

Only two studies reported data for language tests to accompany measures of effect size for the weather prediction task (with only one of these including decoding measures), so the relationship between the severity of language disorder and weather prediction task performance could not be examined. Three studies reported data for NVIQ measures, which showed that there was a large variation between the disordered and comparison groups for the measures of NVIQ used in the studies, the mean difference was $g = -1.159$, 95% CI [-1.99, -.33], ranging from $g = -0.36$, 95% CI [-1.06, 0.35] to $g = -1.27$, CI [-1.93, -0.61]. This variation in effect sizes between studies was significant, Q (2) = 9.07, $p = .01$, $I^2 = 77.94\%$, $k = 3$, $Tau^2 = 0.42$. However, there were not enough studies in the analysis to be able to examine the effect of this difference in NVIQ group disparity on weather prediction task performance in a meta-regression. There were also too few studies to investigate the moderating influence of any task related variables.

Table 4.10 Characteristics of the 9 group design studies eligible for the meta-analysis using the weather prediction task

Study name	Diagnosis	Age	Sample Size*	Task variant	Trial Total	Combinations	Stimuli	Probabilities (of Sun)
Gabay, Vakil, Schiff & Holt, 2015[1,3]	DD	Adult	15; 15	Holl et al. (2012)	150	14	Geometric	89%; 78%; 22%; 11%
Kemeny & Lucaks, 2010[1]	DLD	Child	16; 16[4]	Not stated	150	Not stated	Geometric	90%; 70%; 30%; 10%
Lee & Tomblin, 2015[1]	DLD	Adult	23; 25	Knowlton et al. (1994)	50	14	Not stated	75%; 57%; 43%; 25%
Lee, Mueller, & Tomblin, 2016[2]	DLD	Adult	22; 19	Knowlton et al. (1994)	50	14	Not stated	75%; 57%; 43%; 25%
Lukacs & Kemeny, 2014[1]	DLD	Child	29; 87	Knowlton et al. (1994)	200	13	Geometric	85.7%; 70%; 30%; 14.3%
Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014[1]	DLD	Child	18; 65	Shohamy et al. (2004)	200	14	Mr Potato Head	20%; 40%; 60%; 80%

[1]included in meta-analysis; [2]insufficient data for inclusion in meta-analysis; [3]included feedback and paired associate versions of the task; [4]3 groups took this task (16 DLD & 16 TD children & 16 normal adults - only age-matched groups are coded); *Language-disordered group first

Study name	Hedges's g	p-Value	Hedges's g and 95% CI
Gabay, Vakil, Schiff & Holt, 2015	-1.271	0.001	
Kemeny & Lucaks, 2010	-0.918	0.012	
Mayor-Dubois et al., 2014	-0.824	0.002	
Lee & Tomblin, 2015	-0.404	0.149	
Lukacs & Kemeny, 2014	-0.023	0.916	
	-0.629	0.005	
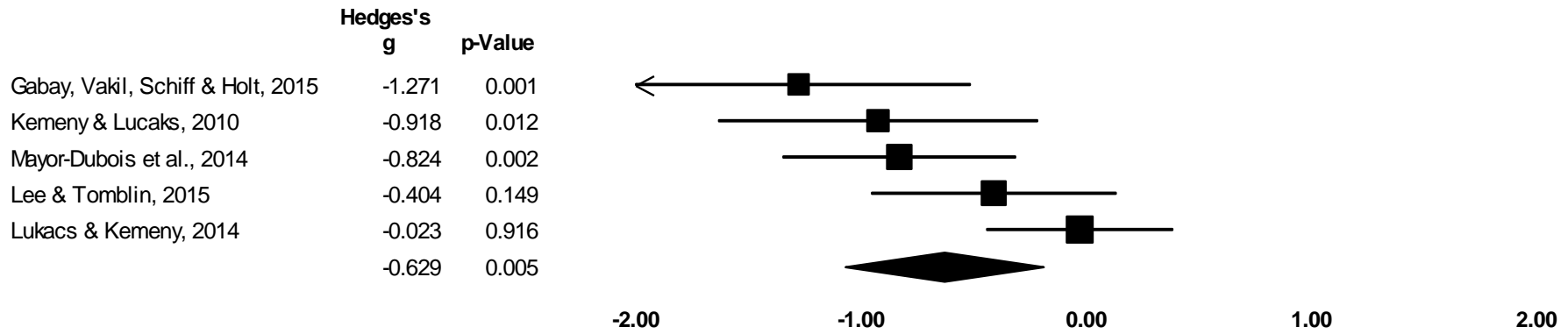
Figure 4.16 Overall average effect size for the group difference in performance on weather prediction tasks (displayed by ♦) with 95% confidence interval for each study.

4.13.1 Publication bias

A funnel plot was estimated for the studies entered into the weather prediction task group design studies meta-analysis, which suggested there was some evidence of publication bias (Figure 4.17). However, with so few studies in the meta-analysis, the funnel plot is not a reliable indicator of publication bias (Lau et al., 2006).
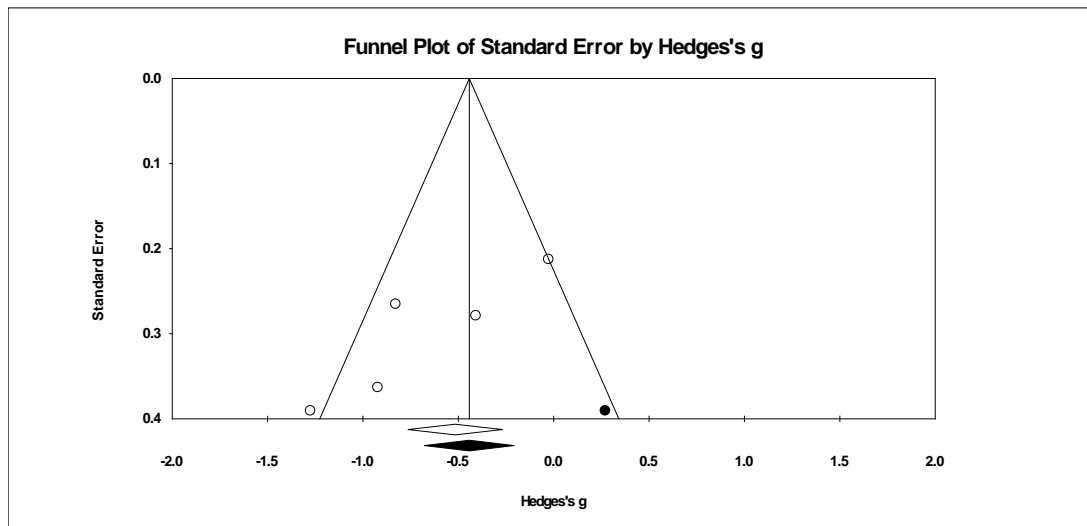
Figure 4.17 Funnel plot showing evidence of publishing bias for the weather prediction group design random effects model. Open circles and diamond correspond to observed studies and point estimate. Filled circles and diamond correspond to imputed missing studies and adjusted point estimate of d = 0.51, 95% CI [-0.08, 15.50], following Duval and Tweedie's Trim and Fill procedure.

A P-curve was also estimated for the six eligible studies, using the quoted test statistic for the principal measure of implicit learning in the studies (see Table 4.11). These test statistics related to the overall difference in learning between groups and were typically the main effect of group in a Group x Block ANOVA. Although there were four significant values for this statistic, it should be noted that no papers reported significant results for the difference in the rate of learning between groups over the task. Figure 4.18 shows a right-skewed p-curve, demonstrating evidential value ($Z = -3.48$, $p = .0003$) and no reliable evidence that the studies' evidential value is inadequate due to low power (power estimate = 87%, 90% CI [46%, 98%].
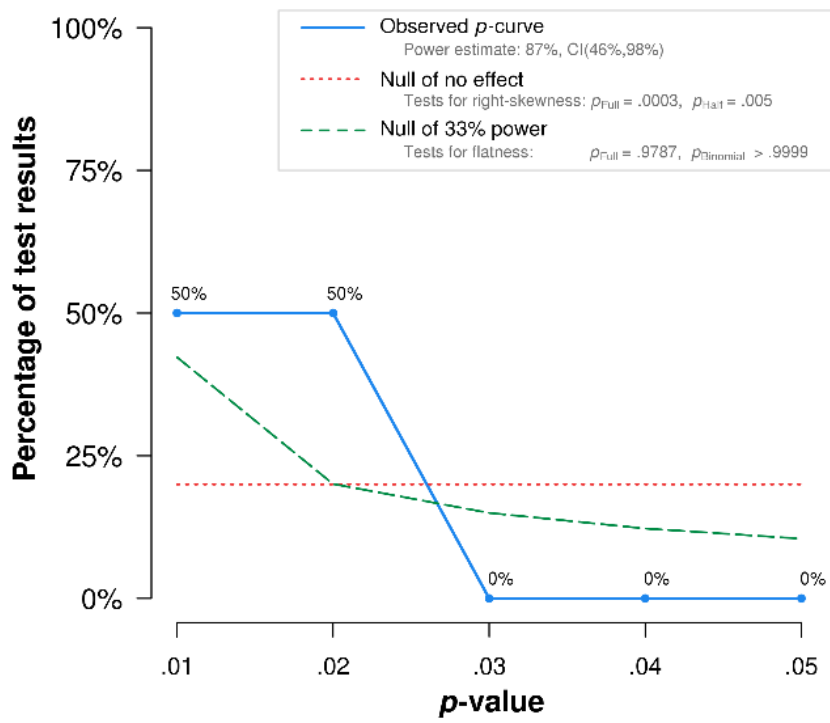
160

Figure 4.18 P-curve examining publishing bias in the six group design extreme group studies using weather prediction task tasks to investigate the procedural deficit hypothesis.

Table 4.11 Disclosure table for the 6 group design studies eligible for the meta-analysis using weather prediction tasks

Study name	Analysis	Quoted test from paper with statistical results	Significance
Gabay, Vakil, Schiff & Holt, 2015	Mixed ANOVA: 2 (Group) x 2 (Task: FB vs PA)	"The main effect of Group was significant, $F(1, 28) = 7.51$, $p = .011$, $\pi^2 p = .204$, indicating that test-phase accuracy of the dyslexia group […] was poorer than that of the control group." (p. 6)	Significant for overall difference for 2 tasks.
Kemeny & Lucaks, 2010	Mixed ANOVA: 3 (Group) x 3 (Block)	"There was a significant main effect of group ($F(2,46) = 15.584$, $p < 0.001$, $\pi^2 p = .409$) showing that there is a significant difference between the groups with adults giving the most correct answers, followed by typically developing children, and children with LI giving the least […] The group block interaction did not appear to be significant ($F(4,46) = .882$, $p = .478$, $\pi^2 p = .409$)…" (p. 18)	Significant for overall difference (P-curve 1); Null for performance over time.
Lee & Tomblin, 2015	Mixed ANOVA: 2 (Group) x 5 (Block)	"Figure 1 (d) shows the results of a significant main effect of Group, $F(1,46) = 6.72$, $p = .01$, $\pi^2 p = .13$ […] The interaction effect was not significant, $F(4,184) = .75$, $p = .56$, $\pi^2 p = .02$." (pp. 225, 226)	Significant for overall difference (P-curve 1); Null for performance over time.
Lee, Mueller, & Tomblin, 2016	Mixed ANOVA: 2 (Group) x 5 (Block)	"Results showed a significant Group effect, $F(1,39) = 11.54$, $p = .0021$ […] The interaction effect was not significant, $F(4,156) = .85$, $p = .50$." (p. 1106)	Significant for overall difference (P-curve); Null for performance over time.
Lukacs & Kemeny, 2014	Mixed ANOVA: 2 (Group) x 4 (Block)	"The Huyhh-Feldt corrected ANOVA revealed that neither the main effect of block ($p = .196$) nor the main effect of group ($p = .814$) was significant. The Block x Group interaction approached, but did not reach significance, $F(2.502, 285.197) = 2.302$, $p = .089$." (p. 478) NB: Main effect of Group supplied by authors: $F(1,114) = .56$, $p = .814$.	Null.
Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014	Mixed ANOVA: 2 (Group) x 4 (Block)	"...but no interaction between Blocks and Groups, $F(3,85) = 1.072$, ns, indicating a similar improvement of cognitive learning in both groups." (p. 19) NB: Effect of group not reported.	Null.

4.14 Discussion: Weather prediction task meta-analysis

The overall effect size in this meta-analysis is the largest in the series. Unfortunately, study numbers were too low to investigate reliably whether this reflected publication bias in the literature. Once again, the results are contradictory for the p-curve and the funnel plot. However, in spite of the substantial significant effect size in the meta-analysis for the group difference, none of the studies reported a significant effect for the Group x Block interaction that would have suggested different implicit learning trajectories for the groups across the tasks. This suggests that the language-disordered group were simply less good overall at learning the conditional probabilities of the task than the control groups. It is possible that lower NVIQ of language-disordered groups accounts for this difference, rather than differences in procedural learning. However, although three of the six studies did not have groups equated for NVIQ, only two of the three were included in the meta-analysis, which meant there were insufficient study numbers to conduct a meta-regression to address this question. It is also possible that poorer learning of language-disordered groups on the task reflected deficits in declarative learning, rather than procedural learning, but again there was insufficient reported data in the studies to investigate this.

4.15 Procedural learning impairment: Domain general or task-specific

Although impaired procedural learning in language disorders is found on both verbal and non-verbal tasks, a question still remains over the domain-generality of any implicit learning impairment. The procedural deficit hypothesis is embedded in the classic multiple systems model of memory. As such, it would reasonably expect implicit procedural memory to be domain-general, reflecting the common reliance of different procedural paradigms on a specific memory system in the brain. Impairments in procedural learning in language disorders would, therefore, also be domain-general. Impaired learning on verbal procedural tasks should then be found, not only on non-verbal analogue tasks, but across other procedural learning paradigms too. However, other memory accounts would not necessarily expect results from different experimental paradigms to relate to one another, as each task would lead to a different

manifestation of changes in the brain, depending on the associated brain areas involved in the performance of each specific task.

Most studies examining the procedural deficit hypothesis have used a single implicit learning task and are, therefore, unable to deliberate on the extent to which any impairments in a language-disordered group may generalize to other tasks. However, fourteen studies (see Table 4.12) have used multiple tasks and evidence to date suggests considerable task specificity, with impaired learning at a group level confined to a single task. One of these studies (Gebauer & Mackintosh, 2007) was not eligible for the meta-analyses, as it did not investigate language ability.

With the exception of Gebauer and Mackintosh (2007) these studies all use group designs and do not report on whether the different implicit tasks correlate with one another. This question can, therefore, not be looked at using meta-analytic techniques. However, Gebauer and Mackintosh (2007) demonstrated no relationship at all between performance on a serial reaction time task, an artificial grammar learning task and a complex systems process control task, using an individual differences approach. Lending some support to this finding within the group design studies, many find impaired learning in language-disordered participants is confined to a single task. Table 4.12 indicates whether performance of the language-disordered group was impaired compared to controls or not. Only Vicari et al. (2005) found impaired learning on all tasks (although there were only two). Hsu & Bishop (2014) found a significant correlation between Hebb learning and deterministic serial reaction time task performance, such that better Hebb learning correlated with faster decrease in RTs to portions of the serial reaction time task exhibiting a predictable sequence. They concluded that this correlation may have related to individual differences in sequence learning. Lukács & Kemény (2014) also noted impaired learning on two sequence-based tasks, but not a task of probabilistic category learning.

Table 4.12 Studies testing participants on more than one procedural task, highlighting the tasks that report significant implicit learning differences between groups.

	DLD / DD	Age	Correlation between implicit tasks	SRT	Hebb	AGL / SL	WPT	CC	Other
Bennett et al., 2008	DD	Adult	Not reported	✓				x	
Gebauer & Mackintosh, 2007*	-	Teen	Non-significant	x		x			x[1]
Henderson & Warmington, 2017	DD	Adult	Non-significant	x	✓				
Howard et al., 2006	DD	Adult	Not reported	✓				x	
Hsu & Bishop, 2014	DLD	Child	SRT and Hebb ($r = .23$, $p = .09$)	✓	✓				x[2]
Jiménez-Fernández et al., 2011	DD	Child	Different children	✓				x	
Laasonen et al., 2014	DD	Adult	Not reported	x		x			
Lee, 2012[6]	DLD	Adult	Not reported	x			✓		
Lee & Tomblin, 2015	DLD	Adult	Non-significant, except WPT and repetition priming ($r = .35$, $p = .01$)	x			✓		✓[2,3]
Lukacs & Kemeny, 2014	DLD	Child	Non-significant	✓		✓	x		
Mayor-Dubois et al., 2014	DLD	Child	Not reported	x			x		
Rüsseler et al., 2006	DD	Adult	Not reported	x		x			
Vakil et al., 2015	DD	Child	Not reported	x					x[4]
Vicari et al., 2005	DD	Child	Not reported	✓					✓[5]

*Correlational study & tasks related to intelligence, including verbal IQ, but not language-ability specifically; ✓ = significant implicit learning-related differences between groups; x = no significant difference between groups; [1] = Process control; [2] = Pursuit Rotor; [3] = Repetition priming; [4] = Tower of Hanoi; [5] = Mirror Drawing; [6] = unpublished thesis.

Finally, the paradigms measure procedural learning in different ways and these methodological differences may have down-stream consequences for their ability to index both a reliable and a valid measure of procedural learning. It should be noted that these methodological differences do not seem to explain the pattern of findings across the literature.

4.16 Overall summary of review and meta-analysis

Even after this series of meta-analyses, the extent to which procedural learning deficits may be considered a risk factor for developmental language disorder and dyslexia is not clear. The serial reaction time and Hebb analyses both reported significant, but small pooled, effect sizes. However, the effect size for the serial reaction time task meta-analysis was not corroborated by the results of the small meta-analysis of serial reaction time studies using a correlational approach with far larger sample sizes. This analysis found no evidence of a relationship between implicit learning on serial reaction time tasks and language disorder. Failure to replicate the effect size in group design studies in studies using a correlational design undermines the claims of the procedural deficit hypothesis, suggesting that procedural learning as evidenced by these tasks is not a reliable correlate of language-related ability in unselected samples.

The moderate pooled effect size yielded by the group design meta-analysis of artificial grammar and statistical learning tasks suggested that there may be a relationship between statistical learning and language disorder. The result from the corresponding meta-analysis of correlational studies was also broadly in alignment. However, the lower overall effect size suggested by the funnel plot for the group design analysis suggests that this moderate effect size for the group design meta-analysis is likely to be too high. This is in line with the recommendation of a previous meta-analysis (Schmalz et al., 2016). Finally, the significant overall mean difference in the weather prediction task meta-analysis was large and suggested that genuine differences on this task exist between children with language learning disorders and age-matched controls. However, the low number of studies in this meta-analysis raise

166

a question mark over the result. Additionally, there is much discussion about the extent of declarative learning in the task, with initial learning thought to represent a purer measure of implicit learning. The meta-analysis included measures that related to learning over the whole task and a question remains over what proportion of the overall significant effect size can be justifiably apportioned to implicit rather than explicit learning. Together, these results raise a number of issues.

4.16.1 Methodological issues with task design

The small effect sizes in both the serial reaction time and the Hebb learning meta-analyses may reflect their reliance on difference scores to represent procedural learning. It may be unwise, therefore, to draw any firm conclusions about the relationship between implicit learning and language ability using these tasks, as a result of the noise in the data. By contrast, the artificial grammar, statistical learning and weather prediction tasks do not use difference scores as the measure of implicit learning. This benefit should be weighed against other methodological limitations of their measures of implicit learning, such as the explicit testing phase structure of artificial grammar and statistical learning tasks and the possible contribution of declarative learning to performance on the weather prediction task, however. This issue refers us back to the earlier discussion of the relative merits of different measures of implicit learning at the end of Chapter 3.

4.16.2 Moderator influence

A number of participant and task related moderators were considered that might explain the pattern of inconsistent results in the literature. However, no moderator analyses were undertaken for the serial reaction time and Hebb meta-analyses, given the non-significant variation in effect sizes. This is ultimately unsatisfactory. For example, variations in the severity of impairment in the language-disordered groups as well as variations in between group differences in NVIQ were found, but could not be examined as moderators.

A range of moderator analyses were undertaken for the artificial grammar and statistical learning and weather prediction meta-analyses, since there was heterogeneity in the effect sizes for different studies in the analysis. However, not one single moderator was found to explain any of this heterogeneity. The lack of influence of some moderator variables, such as participant diagnosis was expected. Although developmental language disorder and dyslexia themselves are distinct disorders, the practice of attaching labels for one or other disorder to individuals for research purposes is less than perfect. The narrow focus on one or other developmental language disorder in investigations of the procedural deficit hypothesis echoes the separate research traditions behind the study of developmental language disorder and dyslexia by speech and language therapists and educational psychologists respectively. McArthur, Hogben, Edwards, Heath, and Mengler (2000) tested 110 children with dyslexia on the Clinical Evaluation of Language Fundamentals – Revised (CELF-R: Semel, Wiig, & Secord, 1987) a battery of tasks tapping oral language skills, usually used to diagnose developmental language disorder. They also tested the reading ability of 102 children with a diagnosis of developmental language disorder using the Neale Analysis of Reading Ability – Revised (NARA-R: Neale, 1988). They found that 55% of the dyslexic sample scored one standard deviation or more below the mean on the tests used to diagnose developmental language disorder, while 51% of the developmental language disorder sample qualified for a reading impairment. Therefore, just over half of the two samples would have qualified for the alternative diagnosis simply dependent on which tests were administered.

Additionally, although developmental language disorder is frequently diagnosed at an earlier age than dyslexia (Snowling, Bishop, & Stothard, 2000), the age of participants in studies of children in these meta-analyses was very similar, regardless of disorder diagnosis (mean age in studies with children: developmental language disorder = 121 months, *SD* = 18 months; Dyslexia = 130 months, *SD* = 18 months).

However, the lack of explanatory value of other moderators is more confusing. For example, the disorders are both dimensional and one could, therefore, expect that any

procedural learning impairment would also be correspondingly dimensional in nature. However, in spite of significant variation in severity of language impairment in the disordered groups, it did not explain any of the heterogeneity in the effect sizes of the artificial grammar and statistical learning meta-analysis. This raises the possibility that, other participant or task-related factors not accounted for in the studies (such as perhaps attention, motivation, or sequence-specific details of the grammars themselves) may be responsible for this variation. This leads to questions about the optimal choice of experimental design to investigate the procedural deficit hypothesis and this will be explored in more detail in the following chapter as part of the rationale for the experimental design used in this thesis.

4.16.3 Issues of task validity

The pattern of results within the series of meta-analyses also raises the question of whether the tasks specifically index procedural learning and procedural learning alone. The larger effect sizes in some of the meta-analyses may indeed demonstrate that tasks are measuring a true underlying difference between groups, but that this difference does not relate to procedural learning ability. Implicit learning tasks are not process pure and the extent to which they index factors such as declarative learning is not clear (Shanks & John, 1994), while other participant cognitive factors such as endogenous attention (de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016) may also play a role. These potential issues with task validity will be revisited in the following experimental chapters.

4.16.4 Publication bias

Funnel plots are the most commonly used device to measure publication bias, but have recently been criticized, particularly when used to determine bias in analyses with a small number of studies. This has led to the introduction of the p-curve method to measure publication bias (Simonsohn et al., 2013; 2014). There is an important difference between the p-curve and funnel plot analyses of publication bias in this series of analyses, however. The p-curve analyses as used here relate to whether there

is likely to be bias across the literature as a whole, while the funnel plots relate to whether the effect size in the meta-analysis itself is likely to be inflated as a result of publishing bias. The results from funnel plots and p-curves were mainly contradictory. Given that they do not necessarily measure exactly the same thing in each case, the decision was taken to report both.

4.16.5 Conclusion

Finally, the meta-analysis for the extreme groups serial reaction time task studies had the largest number of eligible studies, yet ultimately less than half of that number of studies could be included in the final meta-analysis. This is unfortunate, since this is the definitive task used to investigate the procedural hypothesis. The meta-analysis of serial reaction time and artificial grammar learning tasks by Schmalz et al. (2016) concluded that the quality of the data reported in serial reaction time tasks was of insufficient quality to warrant a meta-analysis and I would reluctantly have to agree with this view. The main reason for the lack of quantitative data in eligible studies is the convention of reporting performance on the task in figure format only. The majority of tasks in the serial reaction time task group design meta-analysis did not report means and standard deviations for group performance numerically, either as difference scores or scores per sequence type. This has down-stream consequences for any attempt at meta-analysis. For example, Siegert et al's (2006) meta-analysis of procedural learning in serial reaction time tasks in Parkinson's disease served as the model for previous meta-analyses of procedural learning and language disorders. It contained thirteen studies, nine of which did not report means and / or standard deviations in numerical form. This led the authors to estimate the means and / or standard deviations used to calculate the effect size from the plotted lines for each group in the relevant figure published in the papers. With the growing popularity of meta-analysis and its great utility in exploring the underlying relationships between variables in literatures with inconsistent results, this issue should be addressed in future research.

To conclude, we return to the questions posed at the start of this chapter. Insufficient data in many studies, group deficits of varying sizes across paradigms, as well as a contradictory result from serial reaction time studies using a correlational approach, meant that it was not possible to conclude definitively whether a relationship exists between procedural learning ability and language disorder on the basis of literature to date. Additionally, nonsignificant variation in effect sizes in two of the meta-analyses precluded the investigation of moderators to answer questions about the possible effects of participant variables, such as age or diagnosis, and task variables such as complexity or length. Moderator analyses were only possible in the artificial grammar and statistical learning group design meta-analysis, but in this case not one of the possible moderators explained any of the variability in the studies. The questions about the extent of any procedural learning impairment in language disorder and about the moderating influence of a range of factors across the literature, therefore, remain largely unanswered. We now turn to the next section of the thesis, which takes a different approach to the existing literature in an attempt to shed light on these questions.

Chapter 5   Study 1: Procedural and declarative learning and language-related attainment in children

5.1   Introduction

According to the procedural deficit hypothesis (Nicolson & Fawcett, 2007; 2011; Ullman, 2004; Ullman & Pierpont, 2005) a key risk factor for language learning disorders such as developmental dyslexia and developmental language disorder is impaired procedural learning. However, as has been documented in the preceding chapters, studies evaluating this hypothesis have produced highly inconsistent results. Such inconsistencies may reflect a reliance on measures with low reliability and the use of extreme group designs with small group sizes. The current study takes a different approach to this issue: it assesses the relationships between measures of language and attainment and a wide range of measures of both procedural and declarative learning in a large unselected sample of children. It also takes care to assess the reliabilities of all measures used.

The hypothesis also suggests that it is procedural sequence learning that is a critical cognitive risk factor for developmental language disorder and dyslexia, while declarative learning mechanisms remain relatively intact (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005). However, investigations into procedural learning and language disorder do not often include declarative learning measures. Research on the relationship between language skills and explicit memory skills has frequently used free recall and serial recall tasks. Impaired free recall (Menghini, Carlesimo, Marotta, Finzi, & Vicari, 2010, Vellutino & Scanlon, 1985) and serial recall (Di Betta & Romani, 2006; Perez, Majerus, Mahot, & Poncelet, 2012) have been found in adults and children with language-learning disorders.

The theory also suggests that problems in a procedural learning system should be found in different modalities (Ullman, 2004), affecting both non-verbal and verbal stimuli, yet studies do not often include measures of both verbal and non-verbal implicit learning to test this. It is also relatively rare for investigations into implicit

learning impairments and language to select more than one or two implicit learning paradigms on which to test participants. The extent to which children might perform consistently across a range of tasks, involving both verbal and non-verbal and both implicit and explicit learning, is not clear. One of the few studies examining the same participants on more than one kind of implicit sequence learning task found children with developmental language disorder performed with some consistency across Hebb and serial reaction time tasks (Hsu & Bishop, 2014). However, any consistency in performance across sequence learning paradigms has so far not been found to extend to implicit learning tasks such as contextual cueing, which index other cognitive domains (Jiménez-Fernández et al., 2011; Howard et al., 2006). It is thought possible that separate mechanisms may mediate performance for implicit learning of information with a spatial or temporally sequential structure (Seger, 1994). This study will, therefore, include a range of tasks, both sequence-related and more broadly implicit to test these claims.

A brief re-cap of the implicit learning tasks that will be used to investigate the procedural deficit hypothesis in this study follows, but for more detail, see Chapter 3. The implicit learning task most frequently used to investigate the procedural deficit hypothesis is the non-verbal serial reaction time task (SRT: Nissen & Bullemer, 1987). In this task participants respond as quickly as possible to a visual stimulus appearing in 1 of 4 locations on a screen. Faster responses to trials that follow a covert sequence compared to random trials is taken as evidence of implicit learning (Seger, 1994). The original deterministic serial reaction time task may not fully dissociate implicit and explicit learning (Shanks & Johnstone, 1999), but more complex, probabilistically structured (Schvaneveldt & Gomez, 1998) or alternating versions (Howard & Howard, 1997) are thought to minimize the risk of explicit learning. Language-disordered children and adults have been reported to perform poorly both on deterministic serial reaction time tasks (Lum et al., 2010; Lum, Ullman, & Conti-Ramsden, 2013; Vicari et al., 2005; Jiménez-Fernández, Vaquero, Jiménez & Defior, 2011) and more complex alternating versions of the task (Hedenius, 2013; Howard, Howard, Japikse, & Eden, 2006). However, findings are mixed with null results in some other studies of adults

(e.g., Rüsseler, Gerth, & Munthe, 2006; Kelly, Griffiths, & Frith, 2002), and children (Gabriel, Maillart, Guillaume, Stefaniak, & Meulemans, 2011; Lum & Bleses, 2012).

The contextual cueing task (Chun & Jiang, 1998) is another non-verbal measure of implicit learning (Goujon, Didierjean, & Thorpe, 2015), although this task does not involve the learning of sequences. In this task, participants are instructed to find the location of a target stimulus within matrices of distractor stimuli. The position of the target in some matrices is predictable and faster responses to these compared to random unpredictable matrices is considered evidence of implicit learning. So far, studies have not found impaired performance in dyslexic adults (Howard et al., 2006; Bennett, Romano, Howard, & Howard, 2008) or children (Jiménez-Fernández et al., 2011), although impaired implicit sequence learning was found in these same participants.

The most widely used measure of verbal implicit learning is the Hebb serial order learning task (Hebb, 1961), which asks participants to recall lists of stimuli in the order of presentation. At regular intervals during the task a covert repeating sequence is introduced. Better recall of the repeated, compared to non-repeated, sequences provides evidence of implicit learning. Once again findings from this task are inconsistent. Poor implicit learning has been found in children with developmental language disorder (Hsu & Bishop, 2014) and in dyslexic adults (Szmalec, Loncke, Page, & Duyck, 2011; Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015). However, null results have also been found in both disorders (Staels & Van den Broek, 2015; Majerus et al., 2009).

There are a number of possible reasons for the inconsistent results from studies of the relationship between implicit learning and language learning disorders. The vast majority of studies use extreme group designs. Yet, dyslexia and specific language impairment are dimensional, heterogenous, often co-morbid, neuro-developmental disorders (Bishop & Snowling, 2004; Peterson & Pennington, 2015). Language-disordered groups from different studies may not, therefore, reflect the same behavioural symptoms or underlying cognitive impairments. Extreme group designs

174

also tend to overestimate the size of any linear association between variables (Preacher, 2015; Preacher, Rucker, MacCallum, & Nicewander, 2005) and potentially produce measures that may be lower in reliability (Preacher, 2015). Additionally, given the difficulties inherent in recruitment and testing of language-disordered participants, sample sizes in these studies are typically small, further reducing confidence in results.

Finally, there are reasons to suspect that the implicit memory tasks themselves may not be reliable (Buchner & Wippich, 2000; Salthouse, McGuthry, & Hambrick, 1999; Reber, Walkenfeld, & Hernstadt, 1991) and tasks with poor reliability produce large errors of measurement and are inherently insensitive to individual differences (Nunnally & Bernstein, 1994). However, previous studies have not reported the reliability of the tasks used to measure implicit learning.

In summary, it has been suggested that language learning impairments (developmental language disorder and dyslexia) may reflect a procedural learning deficit. A variety of different tasks, involving both verbal and non-verbal stimuli, have been used to assess implicit learning in groups with language learning impairments with inconsistent results. An important question is whether the different measures of implicit learning used to investigate procedural learning really do measure a common underlying procedural learning system, which is distinct from a declarative memory system. Another important question is whether the tasks currently used to assess implicit learning are reliable.

The current study uses a large sample of children unselected for ability. This has the advantage that it will not overestimate the size of any association between measures of attainment and memory performance, as an extreme groups design might. It also uses multiple measures of implicit memory (the serial reaction time, Hebb serial learning and contextual cueing tasks) and explicit memory (immediate serial recall and free recall tasks), using both verbal and non-verbal stimuli. Using this wide range of tasks in a concurrent correlational design makes it possible to assess the factor structure of the tasks and explore whether there are separable implicit and explicit

memory systems. It will then also be possible to assess the extent to which variations in language and reading skills are correlated with variations in implicit or explicit memory skills, should these be dissociable. The reliability of the different measures can also be determined, which is imperative when investigating individual differences.

5.2   Method

This is a concurrent correlational study investigating the possible associations between language attainment and explicit and implicit memory skills in 7- and 8-year-old children.

5.2.1   Participants

One hundred and one Year 3 children (64 girls, 37 boys) from three London primary schools took part. Children's ages ranged from 7 years 5 months to 8 years 7 months (mean = 8 years and 1 month; SD = 3.82 months). Fifty two of the participating children used English as an additional language but were judged by their class teachers to be fluent in English. Ethical clearance for the study was provided by the UCL Research Ethics committee.

5.2.2   Tasks and testing procedures

All children completed a battery of attainment measures that was administered in a single session to whole classes. Subsequently, children completed three further individual testing sessions. The final session comprised four tasks the children had completed before (verbal and non-verbal versions of declarative and implicit memory tasks) in order to measure memory consolidation. Tasks were administered in a fixed order to all children.

5.2.2.1 Attainment Tasks

Test of Reception of Grammar (TROG-2: Bishop, 2003). Children's receptive grammar skills were assessed using this 80 item test, which was adapted for group

administration. Children were asked to match spoken sentences to pictures, following a four picture multiple choice format.

Wide Range Achievement spelling subtest (WRAT-3: Wilkinson, 1993). Children were asked to spell the first 15 words of the WRAT Tan spelling test (go, cat, boy, run, will, cut, arm, dress, train, shout, watch, grown, kitchen, result, heaven) that were dictated by the experimenter.

Picture Word Matching (PWM: Caravolas et al., 2012). This timed single word reading test consisted of 63 items, each of which showed a picture of an object or scene with 4 printed words (the correct word and 3 distractor words). Children were given 3 minutes to select the correct word for as many items as possible.

Test of word and non-word reading efficiency (TOWRE-2: Torgesen, Wagner, & Rashotte, 1999). These individually administered tests required children to read aloud as many words (or non-words) as they could in 45 seconds.

Test of basic arithmetic and number skills (TOBANS: Hulme, Brigstocke, & Moll, 2016). These timed tests were designed to assess fluency in addition, subtraction, and multiplication, giving a composite arithmetic score. In addition, dot and digit comparison tasks required children to circle the larger of 2 groups of dots or the larger of 2 Arabic numerals respectively. Finally, a test assessed the speed and accuracy of counting random arrays of dots. The TOBANS subtests had no reading requirement, with all instructions read aloud to the children.

WASI (Wechsler, 1999). The 28 item WASI matrix reasoning subtest (Wechsler, 1999) was used to assess non-verbal ability.

5.2.2.2   Declarative memory tasks

Word Lists (Cohen, 1997). This free recall test from the Children's Memory Scale assessed children's ability to learn a list of 10 unrelated words over 4 learning trials. Children were asked to recall as many words as possible in any order from a list of 10

177

unrelated words read out by the experimenter (Trial 1). After the first trial only words that had been omitted were read out to children for each of the following 3 trials (Trials 2 – 4). Children were then asked to recall a distractor list of 10 different words spoken by the examiner. A final trial on the first list (without representation of the list) was then attempted (Trial 5). The score for the first 5 trials formed the child's Learning Score. A measure of delayed recall was taken by asking the child to recall the list once more at the end of the testing session (Trial 6). A final memory consolidation measure was taken during the last testing session several days later, asking children to recall as many words as possible from the 10 item list (Trial 7). Scheduling constraints meant the time lapse between Trial 6 and 7 was not the same for all children, but restricting inclusion to the majority of participants with a 2 day lapse did not significantly alter results.

Dot Locations (Cohen, 1997). The Dot Locations task from the Children's Memory Scale was used as a non-verbal analogue of the Word Lists free recall task. It tested recall of a static dot pattern configuration, giving a measure of declarative, non-verbal spatial memory. Children were shown a 4 x 3 grid with a pattern of 6 red dots for 5 seconds. Children were then asked to recreate it on an empty grid, using red plastic discs (Trial 1). This was repeated twice (Trial 2-3). A distractor pattern of yellow dots was then shown and the children asked to reproduce it. Without re-presenting the first pattern, children were then asked to reproduce it once again (Trial 4). A point was scored for each correct location on each attempt. The mean of the scores for these 4 trials formed the child's learning score. Delayed recall was tested by asking the children to reproduce the initial configuration at the end of the testing session (Trial 5). A memory consolidation measure was taken during the final session (Trial 6), asking the children to reproduce the pattern once more. Again, the time lapse between Trial 5 and 6 was not the same for all children, but all were included in analysis, as differences in the time lapse did not significantly alter the results.

Immediate serial recall (ISR). These tasks were developed to give declarative verbal and non-verbal measures that specifically targeted memory for sequences. They formed the beginning of the implicit memory Hebb sequence learning tasks.

Two versions of the task were created: a verbal task that used nameable pictures as stimuli and a non-verbal task that used abstract symbols. A total of 8 stimuli were used for each version of the task. The non-verbal and verbal stimuli used are shown in Figure 5.1.
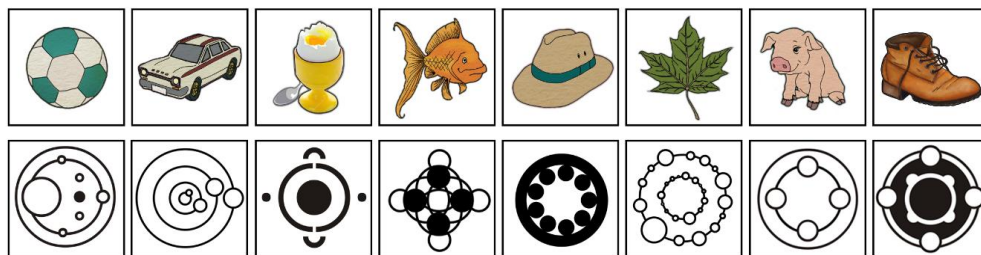
Figure 5.1 Immediate serial recall and Hebb task verbal and non-verbal stimuli.

Eight pictures with dissimilar names were selected that 7 - 8 year-old children would be familiar with (fish, car, egg, shoe, pig, hat, leaf, ball). Symbols for the non-verbal condition were selected that were judged to be difficult to name but were easily discriminable from each other (http://www.dudeman.net/siriusly/cc/phenom.html). Verbal and non-verbal versions were administered as separate tasks during different testing sessions.

On each trial a sequence of stimuli was presented across the top of a computer screen. All 8 possible stimuli then appeared across the middle of the screen in a random order. Children were instructed to use the computer mouse to click on these stimuli to reconstruct the sequence they had just seen. Each item the child clicked on disappeared from the central display, reappearing in the order of selection in the child's reconstructed list at the bottom of the screen. Once an item was selected it could not be changed. All trial sequences were randomly generated.

179

The task began with an 8 trial practice round with each trial presenting a single stimulus. The recorded portion of the task began with 4 trials at sequence length 2. If the child reconstructed one or more of these sequences correctly they proceeded to the next level (3 item sequences). Each subsequent level contained 4 trials, at a sequence length one item longer than the preceding level up to a maximum of 7 items. Trials continued until all 4 trials at a given sequence length were incorrectly reconstructed, at which point testing stopped. At each increase in sequence length the test sequence remained on the screen for an additional one second, starting at 3 seconds for 2 item sequences. The number of trials correctly reconstructed at each sequence length was recorded. This information was used to calculate a span score, consisting of the longest sequence length recalled correctly on all 4 trials, plus .25 for each longer sequence correctly recalled (see Conway et al., 2005; Hulme, Maughan, & Brown, 1991).

5.2.2.3 Implicit memory tasks

All implicit memory tasks were presented on a Dell laptop with a 15 inch screen with resolution set at 1366 x 768 dpi.

Serial Reaction Time Task (SRT). An SRT task (Nissen & Bullemer, 1987) with a probabilistic sequence structure based on Schvaneveldt and Gomez (1998) was used to investigate non-verbal implicit spatial sequence learning. A verbal analogue of the SRT task adapted from Hartman, Knopman, and Nissen (1989) was devised to test verbal implicit sequence learning.

For the non-verbal SRT task (NV-SRT) two 12-item sequences were taken from Shanks, Wilkinson, and Channon (2003): sequence A – 314324213412; sequence B – 431241321423. In both sequences, each location repeated 3 times, each time being preceded by a different location; each sequence contained one reversal (121 or 343) and no repeated locations. They differed only in their second order conditional structure. Each block started with a randomly chosen bigram, e.g.: 3 2. The next location selected was either the location that followed that bigram in sequence A (with a probability of .9, i.e.: 4), or was the location that followed the bigram in Sequence B (with a probability of .1, i.e.: 1). This process then repeated with the new most recent

180

bigram, either 2 4, if the transition had been a probable one, or 2 1 if the transition had been improbable. The task continued in this way until the end of the block.

Children were seated in front of a laptop connected to an Xbox Gamepad controller. For each trial a stimulus of a smiley yellow face appeared on the screen in 1 of 4 locations. The locations formed a diamond pattern that corresponded to the pattern of buttons on the Gamepad controller (see Figure 5.2). The children were told to press the button that corresponded to the position of each stimulus as quickly as possible. There were 500 trials. Ten practice trials began, with equal probabilities of each sequence occurring. There were then 5 blocks of 100 trials that followed the sequence probabilities outlined above.

The program recorded the RT and the button pressed, whether correct or incorrect, but required the child to press the correct button before going on to the next trial. There was a 250 ms interval between trials. A pause between blocks allowed the child a short break if needed, with the experimenter manually starting each new block as soon as the child was ready to continue. The task took approximately 15 minutes to complete. Faster RTs for probable compared to improbable transitions were taken as evidence of implicit learning.

Figure 5.2 Non-verbal serial reaction time task. Children pressed the button on the controller that matched the location of the stimulus.

The verbal SRT task (V-SRT) was adapted from experiments by Hartman, Knopman, & Nissen (1989). The sequences were the same as those used by Schwaneveldt and Gomez (1998: Probable sequence A: 121342314324; Improbable sequence B: 123413214243). The probabilistic structure of the task was otherwise identical to the NV-SRT. This task used 4 nameable pictures as stimuli (bird, hammer, fish, tree) from an online directory of Snodgrass & Vanderwart-like images (Rossion & Pourtois, 2004, see Figure 5.3).

Figure 5.3 Nameable picture stimuli used in the verbal serial reaction time task (V-SRT)

The pictures were approximately 10cm square and were presented one at a time on the left half of the computer screen. Each picture was associated with a particular button on a Gamepad controller. A visual key to this pairing was displayed at all times on the right side of the computer screen, so that the pairings did not need to be memorized. As each picture appeared, the child had to press the button on the GamePad controller that corresponded to the picture as quickly as possible. Although pictures in this task were presented one at a time, requiring the participant to make an additional cognitive step by matching the picture to the spatial location displayed on the on-screen key, in all other ways the task was identical to the NV-SRT.

Hebb serial order learning task (Hebb). Following on seamlessly from the earlier immediate serial recall portion of the task, the implicit Hebb task introduced a covert repeated sequence in order to measure implicit learning of repeated sequences. There were 18 trials. Children were not told that the 6th, 9th, 12th, 15th and 18th trials were repetitions of the 3rd trial sequence. All 18 trials were the same sequence length, with the length of the sequence used for each child determined by their performance

on the immediate serial recall task; the Hebb task sequence length was 1-item longer than the longest sequence the child had correctly recalled 2 or more times in the immediate serial recall task. The stimuli selected and their order of presentation were determined randomly. No stimulus appeared more than once in any sequence. Only items correctly recalled in the correct position were scored as correct (Conway et al., 2005). Points awarded per trial were, therefore, up to a maximum of the length of the list. Proportional scores for the blocks for the repeated and random sequences were calculated by dividing the raw score by the allocated list length. Higher proportional scores for repeated trials compared to random sequence trials were taken as evidence of implicit learning.

Contextual Cueing Task. A dual condition contextual cueing task was used to measure visual search efficiency in both non-verbal and verbal modalities simultaneously. Children were required to search for a target in matrices of distractor stimuli. They then had to indicate the quadrant of the matrix that the target appeared in as fast and accurately as possible, by pressing the key on the laptop keyboard that was associated with that quadrant, (A, Z, K or M; for a similar procedure, see Merrill et al., 2013). Five stimuli were chosen for each condition (verbal and non-verbal): four distractor stimuli and one target stimulus. The verbal condition used line drawings of nameable pictures of familiar animals (frog, cow, rabbit, snail and lion). The non-verbal condition required participants to discriminate between a simplified Chinese symbol and four other simplified Chinese symbols (see Figure 5.4). Both the symbols and the nameable pictures could appear in any of four colours (red, yellow, blue or green). All stimuli were 15 mm square.
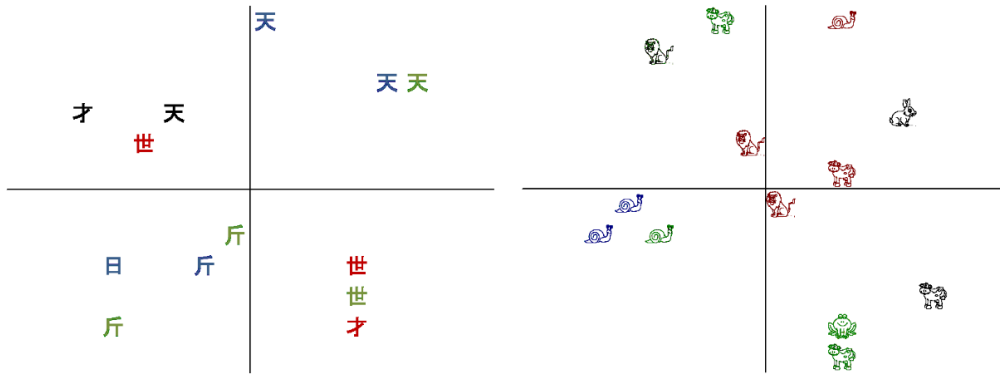
Figure 5.4 Example matrices for the non-verbal and verbal conditions.

All matrices displayed stimuli on invisible 12 x 12 grids divided into 4 easily identifiable quadrants. Three distractor stimuli appeared in each quadrant, such that 12 distractors and the target appeared in every matrix. For each participant the programme randomly selected 8 different locations to contain the target. Half of them were used in the verbal and half in the non-verbal condition. These target locations were sampled from a set of five locations within each quadrant that were all approximately the same distance from the centre of the screen, such that one location was selected in each quadrant per condition. Distractors never appeared in the locations reserved for targets. Each target was used for a different predictable matrix, resulting in 4 different predictable matrices for each condition. Target locations were selected in the same way for unpredictable matrices, but the arrangement of the distractors in each unpredictable matrix was always random and never repeated, so that the positions of distractors in these matrices could not aid visual search.

The experiment was divided into 2 phases. A learning phase of 80 trials included only predictable matrices, with each predictable matrix appearing once in each of 10 blocks. A testing phase of 128 trials subsequently compared speed of response on the "learned" predictable matrices with an equal number of random unrepeated matrices where the position of the target was not predictable. There were 8 blocks in the testing phase, with each block including the 8 predictable matrices plus 8 unpredictable

matrices in random order. Each trial began with a 500 ms fixation cross in the centre of the screen and children were instructed to focus on the cross each time it appeared. There was a 500 ms inter-stimulus interval between trials. To keep accuracy throughout the task high all errors were flagged. A single break was scheduled after 80 trials. The task took most children between 15 and 20 minutes to complete.

5.3 Results

One hundred and one children were tested. Results were available for 86 and 88 children respectively for the non-verbal and verbal ISR and Hebb task, due to computer malfunction. The means, standard deviations and reliabilities for all tasks are shown in Table 5.1.

Table 5.1 Means, standard deviations, reliabilites and sample sizes for the attainment and memory measures.

	N	Mean	*SD*	Reliability
Age in months	101	98.31	3.84	-
Gender (f/m)	101	63/37	-	-
Handedness (right)	90	-	-	-
TROG-2 (Blocks passed)	100	15.25	3.24	.88[s]
TROG-2 (Total correct)	100	71.57	6.53	.88[s]
Literacy composite	101	.0006	.88	
WRAT-3	100	12.11	2.84	.96[s]
PWM	100	37.76	10.89	
TOWRE-2 Words	101	58.83	13.50	.90[r]
TOWRE-2 Nonwords	101	33.93	12.67	.90[r]
Arithmetic composite	100	52.53	23.77	.97[r]
Addition	100	18.08	7.44	.92[r]
Addition plus carry	100	8.42	4.7	.89[r]
Subtraction	100	11.41	5.01	.88[r]
Subtraction plus carry	100	5.2	3.73	.85[r]
Multiplication	100	9.52	6.4	.93[r]
Dot comparison	100	13.14	5.53	.72[r]

Digit comparison	100	21.3	5.73	.80[r]
Dot count	100	11.03	2.78	.79[r]
WASI	100	17.93	5	.94[s]
Dot Locations (DL)				.76[s]/.57[r]
Learning	101	21.02	3.44	
Delay	100	5.29	1.17	
Consolidation	80	4.91	1.23	
Word Lists (WL)				.84[s]
Learning	98	32.81	5.64	
Delay	97	6.18	1.61	
Consolidation	76	5.72	1.73	
ISR (NV)	84	1.66	.397	.49[s]
ISR (V)	87	3.67	.78	.68[s]/.71[r]
NV-SRT1 RT difference	98	58.57	48.49	.75[s]/.21[r]
NV-SRT2 RT difference	90	89.4	48.47	.49[s]/.21[r]
V-SRT1 RT difference	92	40.32	85.58	.17[s]/-.001[r]
V-SRT2 RT difference	86	39.51	87.59	.27[s]/-.001[r]
NV-SRT1 Error difference*	98	2.23	2.73	.18[r]
NV-SRT2 Error difference*	90	4.49	4.62	.18[r]
V-SRT1 Error difference*	92	.69	2.27	-.08[r]
V-SRT2 Error difference*	86	1.03	2.58	-.08[r]
Hebb NV	86	.062	.205	.5[s]
Hebb V	88	.088	.233	.58[s]/.29[r]
Contextual Cueing NV	100	.313	.415	-.03[s]
Contextual Cueing V	100	.248	.483	-.05[s]

[s] = split-half reliability / [r] = test-retest reliability; *Proportional difference in error frequency by sequence

Attainment means were in line with test norms, where applicable. However, performance on the 15 words from the WRAT spelling test approached ceiling, with approximately 20% of the sample getting the maximum score. TROG-2 performance was also high with 75% of the sample scoring over 70 out of 80 items correct. Performance on the non-verbal Dot Locations task was also high. However, there was

186

sufficient variability on all measures to allow correlations to be meaningfully examined.

5.3.1   Learning on the implicit tasks

There was clear evidence of implicit learning on all tasks (see figure 4). Mixed effects models, treating items and subjects as crossed random effects (Rabe-Hesketh & Skrondal, 2012) in Stata (13.0) were chosen to analyse response times (RTs) and recall scores for all implicit tasks in order to take account of item and participant variability. A secondary analysis of error frequencies on the SRT tasks was also undertaken.

5.3.1.1   SRT tasks

Response Times (RTs). For both attempts at both tasks all inaccurate trials and trials over 5,000 ms were removed. Given the unequal number of trials for sequence A and B, a moving criterion, based on sample size (Miller, 1991 as cited in Selst & Jolicoeur, 1994; Cousineau & Chartier, 2010) was used to remove remaining outlying RT observations. This manipulated the parameter applied to the cutoff criterion for each Block by sequence, dependent on sample size, in order to mimic the bias of applying a criterion of 2.5 *SD*s to a sample size of 100. Probable sequence A samples were allocated the default cutoff criterion of 2.5 *SD*s, as they were approaching a sample size of 100 and were, therefore, considered relatively immune to sample bias. Mean RTs and standard deviations for all SRT tasks are in Table 5.2. RTs for the improbable sequence were slower than for the probable sequence for all SRT attempts in every block. However, whereas RTs decreased over time on NV-SRT, they increased over time on the verbal analogue, flagging up possible issues with attention and motivation on this task.

Table 5.2 RT means and standard deviations for all SRT tasks by block

	RT means (SDs) for NV-SRT and V-SRT at time 1 & 2							
Block	NV-SRT1		NV-SRT2		V-SRT1		V-SRT2	
	Prob	Improb	Prob	Improb	Prob	Improb	Prob	Improb
1	603.97	638.81	482.28	540.61	962.46	985.68	859.87	892.25
	(262.24)	(268.42)	(191.65)	(169.67)	(439.87)	(463.14)	(408.27)	(399.51)
2	579.74	629.19	467.38	548.40	994.19	1052.55	895.68	925.35
	(262.44)	(287.76)	(189.41)	(205.95)	(484.06)	(536.54)	(453.37)	(450.87)
3	573.73	623.66	465.55	558.44	1018.34	1049.74	917.37	986.91
	(274.74)	(283.19)	(190.35)	(190.35)	(509.32)	(482.26)	(481.08)	(567.49)
4	547.07	632.39	452.91	562.53	1013.31	1043.0	923.4	981.69
	(258.72)	(276.71)	(188.28)	(201.97)	(527.82)	(591.086)	(522.74)	(532.21)
5	541.62	613.39	445.86	580.91	998.18	1064.49	951.54	984.09
	(252.78)	(233.29)	(207.39)	(302.38)	(498.02)	(585.77)	(542.17)	(508.23)

For all SRT tasks sequence type (Probable or Improbable), block (1 - 5) and the interaction between them were entered as fixed effects and items and participants as crossed random effects. For both non-verbal SRT attempts, RTs for the probable sequence were significantly faster than for the improbable sequence (NV-SRT1: unstandardised regression coefficient = 34.766, z = 4.41, $p < .001$, 95% CI [19.31, 30.22]; NV-SRT2: unstandardised regression coefficient = 54.072, z = 8.69, $p < 0.001$, 95% CI [41.88, 66.27]). On the first task attempt the interaction between block and sequence was significant for the last 2 blocks of the task, providing evidence of implicit learning (Block 4 unstandardised regression coefficient = 48.923, z = 4.34, $p < .001$, 95% CI [26.84, 71.01]; Block 5 unstandardised regression coefficient = 34.751, z = 3.06, $p = .002$, 95% CI [12.52, 56.98]). By the second attempt, this interaction was significant in every block (Block 2 unstandardised regression coefficient = 22.582, z = 2.54, $p = .011$, 95% CI [5.15, 40.01]; Block 3 unstandardised regression coefficient = 33.026, z = 3.69, $p < .001$, 95% CI [15.50, 50.55]; Block 4 unstandardised regression coefficient = 48.910, z = 5.45, $p < .001$, 95% CI [31.32, 66.49]; Block 5 unstandardised regression coefficient = 69.673, z = 7.61, $p < .001$, 95% CI [51.73, 87.61]). For the

verbal SRT task, probable transitions were only significantly faster than improbable transitions on the second attempt (unstandardised regression coefficient = 33.02, z = 20, $p < 0.046$, 95% CI [.61, 65.43]). The interaction between sequence and block was not significant at any point in either verbal task. See Figure 5.5 for performance on all tasks.
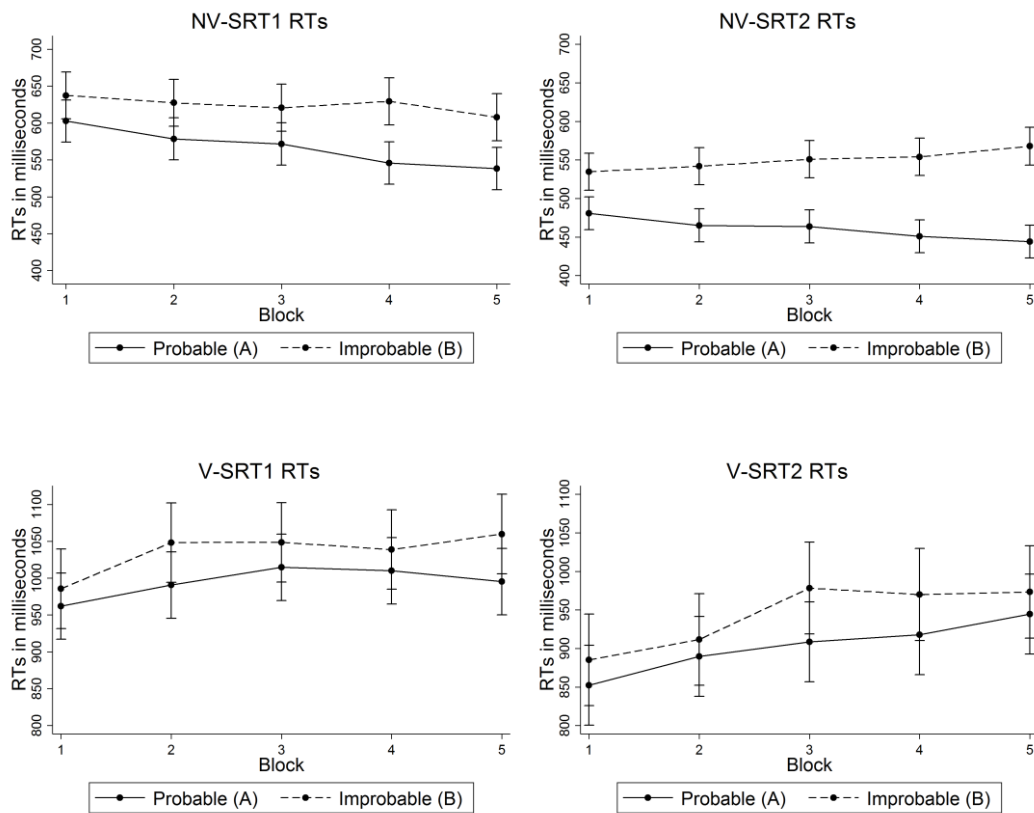
Figure 5.5 RTs per sequence and block for all SRT tasks. Error bars are 95% confidence intervals.

Error frequencies. Error frequencies on probabilistic SRT tasks are considered to be a meaningful index of how much participants are anticipating the sequence (Schwaneveldt & Gomez, 1998). This is in contrast to error rates on deterministic SRT tasks, which are generally low and covary with overall RTs. On probabilistic SRT tasks implicit learning can be inferred from more frequent errors on the improbable compared to the probable sequence. Similarly, the type of error on the improbable

189

sequence can also reflect implicit learning, with more errors that anticipate the probable sequence than random errors indicative of implicit learning. Two analyses were conducted to assess implicit learning from the pattern of errors on all four serial reaction time tasks. Means, standard deviations and reliabilities for relevant error frequencies by sequence and and error type, as well as for the derived difference score measures are shown in Table 5.3.

Table 5.3 Means, standard deviations and test-retest reliabilities for error frequencies for component measures and difference scores.

	Non-verbal SRT tasks			Verbal SRT tasks		
	NV-SRT1	NV-SRT2	Reliability (r)	V-SRT1	V-SRT2	Reliability (r)
Sequence errors analysis						
Probable Errors*	27.44 (20.97)	24.07 (23.30)	.64	43.26 (32.69)	45.92 (44.06)	.81
Improbable Errors	5.28 (4.26)	7.17 (6.53)	.49	5.49 (3.80)	6.13 (5.63)	.89
Proportional sequence errors difference score[1]	2.23 (2.73)	4.49 (4.62)	.18	.69 (2.27)	1.03 (2.58)	-.08
Error type analysis						
Anticipatory errors on improbable sequence	4.12 (3.53)	6.27 (5.84)	.49	3.97 (2.93)	4.66 (4.52)	.41
Random errors on improbable sequence*	1.15 (1.36)	0.90 (1.10)	.29	1.53 (1.42)	1.47 (1.73)	.30
Proportional error type difference score	3.55 (3.32)	5.82 (5.55)	.47	3.20 (2.68)	3.92 (4.12)	.34

*Before proportional adjustment; [1]this measure is included main table of study means.

The first set of analyses investigated whether participants made proportionally more errors on the improbable compared to the probable sequence, with an increasing divergence over the course of the task, as participants progressively anticipated probable locations on improbable sequence trials. In order to compare the frequency of errors, the measure for probable errors was adjusted proportionally, by dividing by 9, as the ratio for probable to improbable trials was 9:1.

For these analyses, sequence, block and the interaction between them were entered into mixed effects models as fixed effects and items and participants as crossed random effects. For the non-verbal tasks Figure 5.6 shows that proportionally there are more errors on the improbable sequence than on the probable sequence (NV-SRT1: unstandardized regression coefficient = .349, $z = 20.08$, $p < .001$, 95% CI [.32, .38]; NV-SRT2: unstandardized regression coefficient = .462, $z = 25.98$, $p < .001$, 95% CI [.43, .50]). On both attempts there was an interaction such that errors increased at a greater rate on the improbable sequence than on the probable sequence over the course of the task (NV-SRT1: unstandardized regression coefficient = .383, $z = 15.58$, $p < .001$, 95% CI [.34, .43]; NV-SRT2: unstandardized regression coefficient = 1.017, $z = 40.47$, $p < .001$, 95% CI [.97, 1.07]). These results are consistent with implicit learning, which is particularly evident on the second attempt on the task.

Equivalent analyses on the verbal analogue tasks did not yield such clear results (also see Figure 5.6), although the data was still suggestive of implicit learning. There were proportionally more errors on the improbable than on the probable sequence (V-SRT1: unstandardized regression coefficient = .220, $z = 12.29$, $p < .001$, 95% CI [.18, .25]; V-SRT2: unstandardized regression coefficient = .087, $z = 4.37$, $p < .001$, 95% CI [.05, .13]). On both attempts there was an interaction. However, the difference between the sequences decreased on the first attempt at the task, which cannot be related to implicit learning, only increasing on the second attempt at the task (Interaction Block 1 – 5 V-SRT1: unstandardized regression coefficient = -.066, $z = -2.63$, $p = .009$, 95% CI [-.12, -.02]; V-SRT2: unstandardized regression coefficient = .185, $z = 6.59$, $p < .001$, 95% CI [.13, .24]).
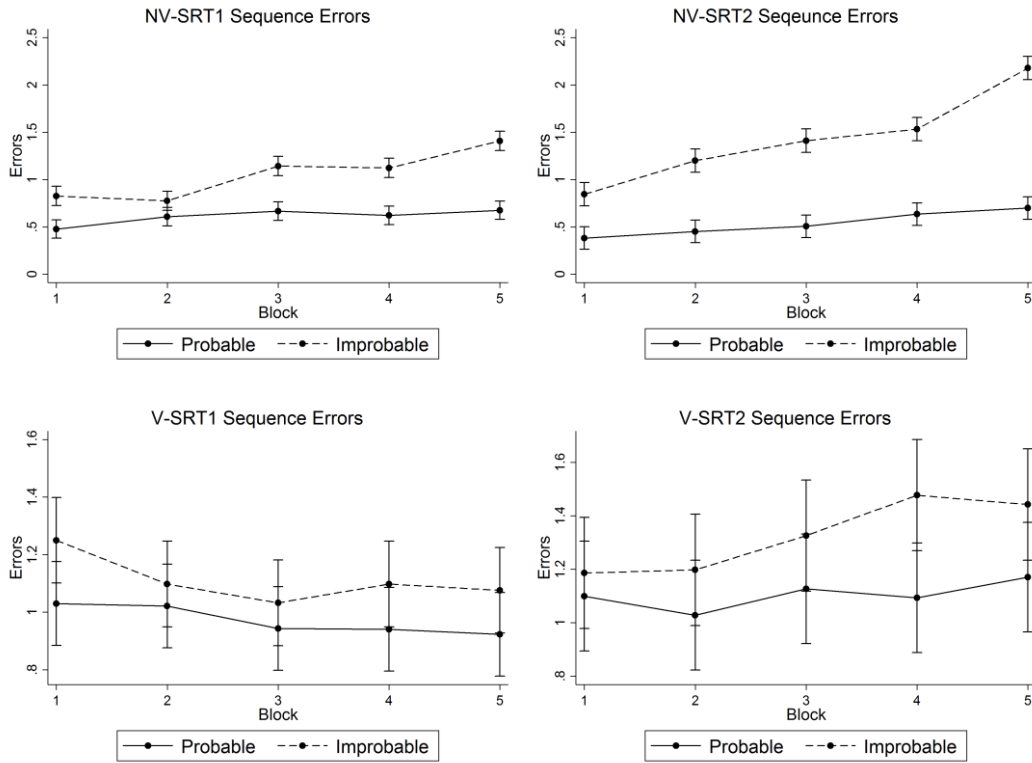
Figure 5.6 Error frequencies for the non-verbal and verbal SRT tasks per sequence. Error bars are 95% confidence intervals

The second set of analyses explored the extent to which the errors that participants made on the improbable sequence were anticipatory of the probable sequence. More anticipatory errors compared to random ones, as well as an increase in the difference between these error types over the course of the task would reflect implicit learning on the task.

Once again, error frequencies were proportionalized, in order to be directly comparable. This was because on each location only one button was associated with an anticipatory response, while two buttons were associated with random responses. The random measure was, therefore, divided by two (NB: since one of the two possible random responses inevitably necessitated a repeated button press, it should be noted

that the pattern of significance detailed below was equivalent for analyses where the random frequency was not proportionately adjusted).

For these analyses, error type (anticipatory & random), block and the interaction between them were entered into mixed effects models as fixed effects and items and participants as crossed random effects (see Figure 5.7). The non-verbal tasks showed that the frequency of anticipatory and random errors were equivalent on both task attempts (NV-SRT1: unstandardized regression coefficient = -.172, z = -1.71, p = .08, 95% CI [-.37, .03]; NV-SRT2: unstandardized regression coefficient = -.348, z = -.180, p = .07, 95% CI [-.73, .03]). None of the interactions were significant, such that anticipatory and random errors did not increase at different rates across the task.

There were proportionally more anticipatory than random errors on the improbable sequence on the first attempt on the verbal SRT task (V-SRT1: unstandardized regression coefficient = -.195, z = -3.98, p < .001, 95% CI [-.29, -.10], but this effect was clearly driven by the large difference on the first epoch and the frequencies of the two types of error were equivalent on the second attempt. Random and anticipatory errors did not behave in a significantly different manner from each other across the blocks of the task.
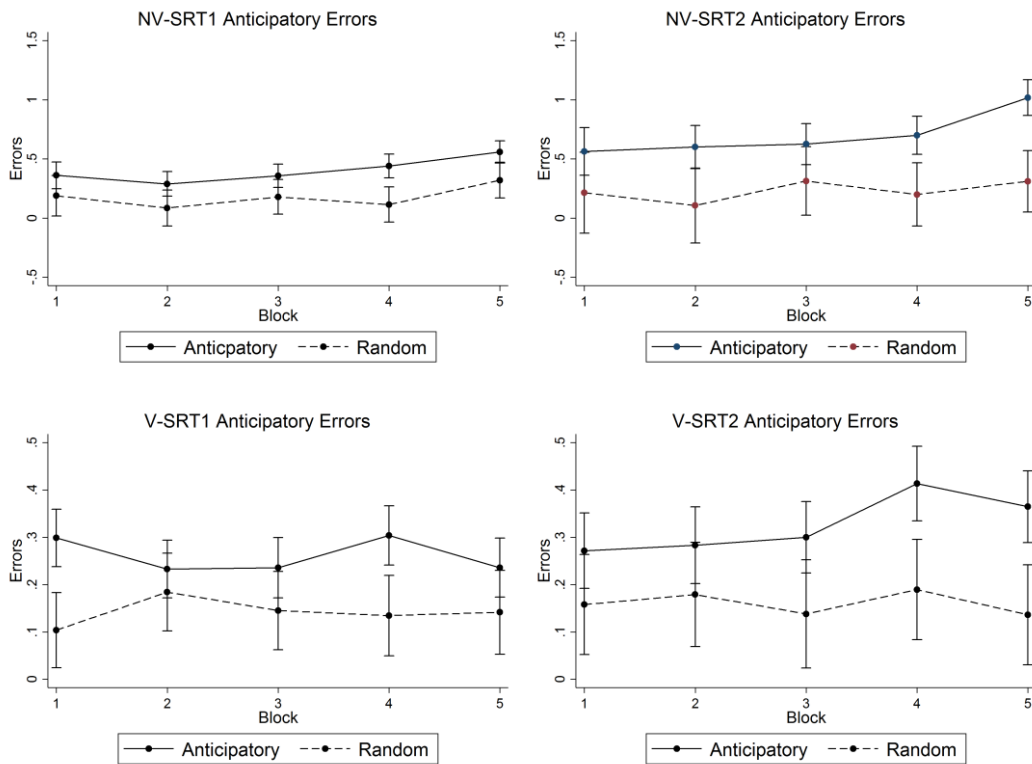
Figure 5.7 Error frequencies for anticipatory and random errors on the improbable sequence of the non-verbal and verbal SRT tasks. Error bars are 95% confidence intervals.

To summarize, the results of the error analyses provide evidence of implicit learning on the SRT tasks. The sequence error measures in particular echoed the pattern of learning seen in the RT measures. Just as RTs on the improbable trials slowed down over the course of the task, as participants anticipated probable sequence locations, the frequency of errors on the improbable sequence also increased. However, as with the RT analyses, results were less conclusive for the verbal tasks.

5.3.1.2 Hebb tasks

For both Hebb tasks, sequence type (repeating or non-repeating), block $(1-6)$ and the interaction between them were entered as fixed effects and items and participants as crossed random effects. Proportional recall scores by sequence and block for the non-verbal and verbal tasks are graphed in Figure 5.8. The random score for each block

was the random trial mean. Mean recall for the repeating Hebb sequence was greater than for random sequences in both the non-verbal and verbal versions on all blocks. The non-verbal Hebb task did not show significant evidence of implicit learning, suggesting the task demands with unnameable stimuli were too high. However, on the verbal task repeated Hebb sequences were recalled significantly better than random sequences (unstandardised regression coefficient = .115, z = 3.21, p = 0.001, 95% CI [.045, .18]). The interactions were not significant, such that recall for the two types of sequence did not develop at different rates across the task.
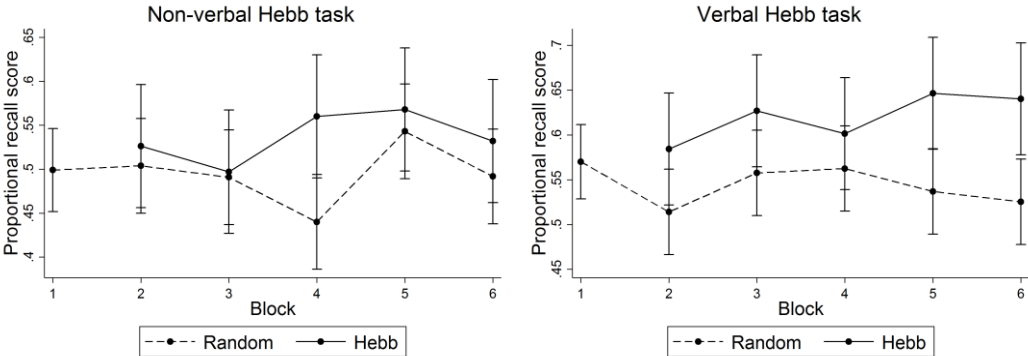
Figure 5.8 Proportional recall scores per sequence and block for Hebb tasks. Error bars are 95% confidence intervals.

5.3.1.3 *Contextual cueing task*

Verbal and non-verbal conditions of the contextual cueing task were analysed separately. Matrix type (Predictable and Unpredictable) and epoch $(6 - 9)$ and the interaction between them were entered as fixed effects and items and participants as crossed random effects. Only RTs in the testing phase were analysed (128 trials, divided into 4 epochs). All inaccurate responses, responses over 10,000 ms and RTs three test phase standard deviations above or below each participant's epoch mean were excluded from analysis, removing 647 trials from analysis across all participants. RTs were positively skewed for both conditions, so analysis was conducted on log transformed RTs. The verbal and non-verbal conditions of the task were analysed separately. Children reported that the verbal condition was more difficult and this was

reflected in slower RTs for the verbal compared to the non-verbal condition. Figure 5.9 graphs RTs across both the learning and testing phase of the experiment for both conditions by matrix type and epoch. Targets were identified significantly faster in predictable matrices than in random ones for both non-verbal and verbal conditions, suggestive of implicit learning (Non-verbal: unstandardized regression coefficient = -.072, z = -3.69, $p < .001$, 95% CI [-.110, -.034]; Verbal: unstandardized regression coefficient = -.0067, z = -3.00, p = .003, 95% CI [-.11, -.023]). No other effects were significant.
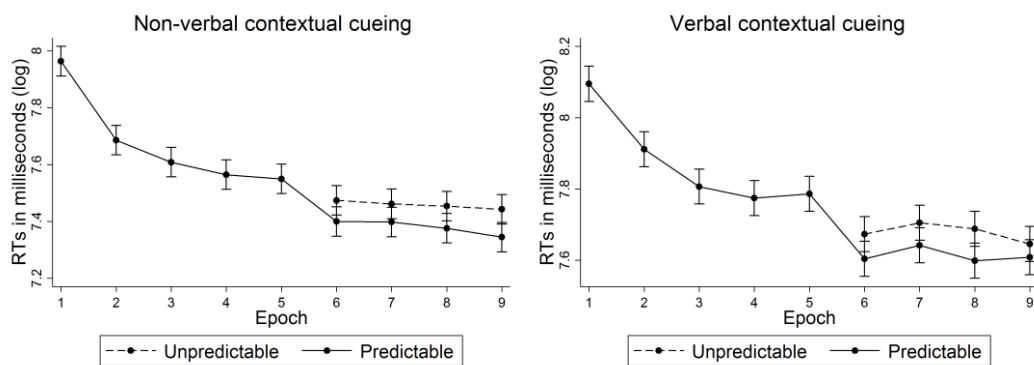
Figure 5.9 RTs for contextual cueing tasks by matrix type (log transformed). Error bars are 95% confidence intervals.

5.3.2 Task measures and reliabilities

Reliabilities for all tasks are shown in Table 5.1. The scores for all declarative tasks were based on the number of items correct. Reliabilities for the declarative tasks were generally good. Reliability estimates for CMS Dot Locations and Word Lists subtests for 7 year olds, as supplied by the CMS manual, depended on the length of delay. The correlation between Delay and Consolidation measures can be regarded as lower bound estimate of the test-retest reliabilities in the current sample. A split-half reliability analysis of 40 of the children in the study (10 children from each class) showed verbal ISR to be more reliable than non-verbal ISR. Children reported that they found the non-verbal ISR difficult and were typically timed out after only 2 or 3

levels compared to 5 or 6 levels for the verbal task and this may be behind its lower reliability. Task difficulty was also reflected in the higher mean span score for the verbal rather than non-verbal ISR. There was no way to establish test-retest reliability for the ISR tasks in the main study, so a separate sample of 27 Year 3 children were tested on the verbal ISR/Hebb task on two occasions 3 days apart, yielding a test-retest reliability for the verbal ISR of ($r = .71$).

Implicit learning tasks required the calculation of derived measures for each participant. As a result of the additive nature of measurement error, the difference between two cognitive test scores is less reliable than either of the scores it is derived from (Lord, 1958; Overall & Woodward 1975). This is particularly the case when their parent scores are unreliable and / or when they are positively correlated (Tisak & Smith, 1994; Johns, 1981; Edwards, 1994). Some researchers have rejected difference scores entirely in favour of regression-based residualised measures (Cronbach & Furby, 1970); others have questioned these too (Griffin, Murray, & Gonzalez, 1999); and yet others have accepted difference scores, as long as they are systematically approached and individual circumstances considered (Rogosa, Brandt & Zimowski, 1982; Rogosa & Willet, 1983). It was, therefore, important in the current study to consider the merits of different scoring methods for implicit learning on a case-by-case basis.

5.3.2.1 Serial reaction time RT measures

The reliability of both difference scores and regression-based residual RT measures were investigated for the serial reaction time tasks. Difference scores were calculated by subtracting the task mean for probable transitions from the task mean for improbable transitions to give a measure that took account of the ratio between probable and improbable trials. Residual measures reflected the degree of deviation from the regression slope for probable transitions that occurred on improbable transitions. For the serial reaction time tasks, the proportional mean difference in RT between sequence types across all trials was preferred, as there was relatively little evidence of an interaction between the sequences on the majority of the serial reaction

time tasks. Split-half reliability was then calculated by sequentially numbering the trials for each sequence and calculating a proportional difference score measure for odd and evenly numbered trials separately. Test-retest reliability was established by correlating children's difference scores on the first and second time of taking the task. Although the test-retest reliability for RTs by each individual sequence was good (see table showing correlations for component scores for all implicit learning tasks in Appendix B), the reliability of the difference scores was poor. It should be noted that the residual measures demonstrated similar levels of unreliability and an equally nonsignificant relationship with language-related attainment.

Additionally, as test-retest reliability for the difference scores was poor, an alternative, coarser-grained, binary measure of difference was also considered. This method was first recommended by Lord (1958) and has been used for serial reaction time tasks with reported success in studies of individual differences in implicit learning (Kaufman et al., 2010; Pretz, Totz, & Kaufman, 2010). The method first calculated the effect size for the difference between the probable and improbable RT means for the sample (Cohen's d for NV-SRT1 = .22; NV-SRT2 = .48; V-SRT1 = .08; V-SRT2 = .09) and allocated a point for each block a participant's learning on the probable trials was as high or higher than this sample effect size. A maximum score of 5 was achievable, which would indicate implicit learning on every block in the task. Test-retest reliability was once more calculated, this time using the binary scores. Although reliability had improved, it was still unacceptably low (NV-SRT r = .28; V-SRT r = .20). Nevertheless, relationships with both language-related and declarative measures were examined. There were no relationships with Language (TROG-2). The non-verbal NV-SRT2 task correlated with Literacy (r = .28), Arithmetic (r = .26), digit comparison (r = .24) and non-verbal long-term memory (DL Consolidation r = .33). The verbal V-SRT1 task correlated with the verbal long-term memory measure (WL Consolidation r = .28). However, all these correlations were low and none of the relationships remained significant once a Bonferroni adjustment has been made. It is interesting to note, however, that the strongest relationship was between the second attempt at the serial reaction time task, which would have taken account of any

consolidation of implicit learning from the first attempt, and a measure of declarative non-verbal memory that also related to long-term memory.

5.3.2.2 SRT error measures

Although the test-retest reliability for the frequency of errors for each individual sequence was generally good, the reliability for the derived measures that represented the proportional difference in the total amount of errors on the probable compared to the improbable sequence was very poor (test retest reliability: NV-SRT r = .16; V-SRT r = -.08). The test retest reliability for the proportional difference between anticipatory and random types of error on the improbable sequence, was a little better (test-retest reliability: NV-SRT r = .42; VSRT r = .34). However, floor effects accounted for the better reliability of this measure. To take the NV-SRT tasks as an example, approximately 37 participants on NV-SRT1 and 40 participants on NV-SRT2 made no random errors on the improbable sequence at all and derived measures for these participants were consequently simply raw score measures of anticipatory errors on the sequence. Similarly, eight participants made no anticipatory errors on the improbable sequence and their derived measures were a raw score measure of random errors. As a result, the derived procedural learning scores on the NV-SRT tasks were effectively not difference scores in approximately half of the participants. This pattern of low or no errors of one or other type on the improbable sequence was similar in the V-SRT tasks too. Importantly, the error measures showed only low and non-significant correlations with the derived RT measures, as well as no significant correlation with any implicit, declarative or attainment variables in the study at all. As a result of the extremely low reliability of the sequence errors measure and the low frequencies on the error type measure, they were not considered further.

5.3.2.3 Hebb serial order implicit learning measures

Recall scores were divided by list length, in order to control for variability in participant recall ability. Although the gradient of improvement on repeated trials compared to random trials has frequently been used to give an index of Hebb sequence learning (Guérard, Saint-Aubin, Boucher, & Tremblay, 2011; Hitch, Flude, & Burgess,

2009; Hsu & Bishop, 2014; Page, Cumming, Norris, McNeil, & Hitch, 2013; Szmalec et al., 2011), this is an effective way of capturing implicit learning only if participants show stable recall on random trials and improved recall for the repeated sequence over time (Hebb, 1961). This is usually the case with adult participants, but children have been shown to exhibit a different pattern of Hebb learning to adults (Archibald & Joannise, 2008; Mosse & Jarrold, 2008), with inconsistent recall on random trials, combined with more consistent, rather than improved, recall of the repeated sequence. This pattern was evident in the current study and a more suitable difference score measure was selected, which summed the difference in proportional recall across blocks 4 to 6, by which time any Hebb learning should have been established.

Split-half reliability could only be calculated for the random trials, correlating the first and second random trial per block, as the Hebb trials were not independent. The random trials form the supposedly stable baseline against which improvement on the Hebb sequence is measured (Hebb, 1961). Split half reliability was moderate for both tasks (Hebb NV $r = .5$; Hebb V $r = .58$), indicating considerable variability in recall of random sequences. Children in this study took each Hebb task once, so test-retest reliability was established in a separate sample of 27 children.

5.3.2.4 Test-retest reliability of verbal Hebb serial order learning task

In order to examine the test-retest reliability of the Hebb learning tasks a separate sample of twenty seven children (13 girls and 14 boys) from two Year 3 classes were tested on the verbal ISR and Hebb learning task on two occasions 3 days apart. Mean age was 7 years and 7 months ($SD = 3.97$ months). The data from one child was not included in the analysis, as he was unable to maintain attention during the retest session. As stimuli selection was randomised, children were presented with different Hebb, as well as random, sequences on each occasion. Administration of the tasks was identical on both occasions.

Both test and retest verbal Hebb tasks showed evidence of significant implicit learning. Means and standard deviations are in Table 5.4. Mixed effects models with block and sequence as fixed effects and items and participants as crossed random

effects found recall for the Hebb trials to be significantly better than recall for random trials both times (Time 1: unstandardized regression coefficient = .143, z = 2.20, p = .028, 95% CIs [.016, .271]; Time 2: unstandardized regression coefficient = .225, z = 3.48, p = .001, 95% CIs [.098, .352]). The implicit learning score for each child's test and retest attempt was the proportional difference across the last 3 blocks of the task. However, the test retest reliability of this score was poor (r = .29), suggesting that implicit learning as shown on Hebb learning tasks is not a stable characteristic in children. By comparison, test retest reliability for the preliminary ISR portion of the task was estimated at r = .71 (ISR mean recall score at time 1 was 3.79 (SD = .68) and time 2 it was 3.98 (SD = .67)).

Table 5.4 Means and standard deviations for Hebb and random trials per block and implicit learning difference score.

Block	Time 1				Time 2			
	Hebb		Random		Hebb		Random	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Block 1	-	-	.66	.31	-	-	.58	.30
Block 2	.67	.31	.64	.28	.57	.33	.57	.31
Block 3	.69	.33	.63	.31	.71	.32	.61	.29
Block 4	.69	.30	.61	.33	.61	.33	.60	.32
Block 5	.82	.24	.61	.30	.71	.25	.58	.30
Block 6	.75	.30	.60	.28	.73	.29	.51	.34
Difference Score	.75	.22	.61	.17	.69	.22	.56	.17

5.3.2.5 Contextual Cueing implicit learning measures

RT variability was controlled in the same way as in the serial reaction time tasks. A single overall facilitation measure was created for each condition (NV & V) that was the mean difference between predictable and unpredictable matrices across the entire testing phase. Similar difference scores have been used in published research (Dixon, Zelazo, & De Rosa, 2010; Brown, Aczel, Jiménez, Kaufman, & Grant, 2010). A measure that attempted to remove noise from the data by dividing each participant's

mean RT by their testing phase standard deviation, was also investigated. This method of reducing participant level variability is frequently used in analysis of implicit association tasks (Greenwald, Nosek, & Banaji, 2013), although it is not commonly used within the contextual cueing paradigm. However, it did not improve reliability and the simpler difference score was, therefore, preferred.

Split-half reliability for raw RTs (on accurate odd and evenly generated trials in the second half of the learning phase) was good for both conditions. However, split-half reliability for the testing phase difference scores, estimated by correlating odd and even positioned difference scores for the 4 epochs of the testing phase for each condition, was negligible (non-verbal $r = -.03$; verbal $r = -.05$). This suggested that the contextual cueing difference scores did not reliably capture the degree of implicit learning shown by participants on this task.

5.3.3 Correlations

Correlations between all literacy measures were high (WRAT spelling, PWM reading test and TOWRE word and non-word reading r's from .62 to .81). Z-scores for these measures were summed to create a composite literacy measure. Correlations between all measures are shown in Table 5.5. Measures of literacy, language, counting and NVIQ showed moderate to strong correlations with each other as expected. Measures of declarative memory showed moderate correlations with literacy, and somewhat lower correlations with language (TROG-2) and arithmetic. The declarative memory tasks correlated with each other broadly as expected, with the dot location memory measures correlating strongly with each other, as did the word list learning measures. Measures of immediate serial recall (both verbal and nonverbal) showed moderate correlations with most of the other memory tasks, and with each other. Finally, the measures of procedural memory correlated weakly and non-significantly with measures of attainment (language, literacy, and arithmetic) and poorly with each other, reflecting the poor reliability of these measures.

Table 5.5 Correlations between all attainment and memory measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Age (months)																					
2. TROG-2	.14																				
3. Literacy composite	.26**	.49**																			
4. Arithmetic composite	.15	.38**	.50**																		
5. Non-verbal IQ	.04	.36**	.25*	.12																	
6. DL Learning	.11	.36**	.18	.13	.34**																
7. DL Delay	.09	.32**	.13	.17	.34**	.77**															
8. DL Consolidation	.04	.37**	.21	.2	.38**	.68**	.62**														
9. WL Learning	.23*	.48**	.26**	.33**	.20	.29**	.22*	.25*													
10. WL Delay	.15	.30**	.24*	.25*	-.00	.10	.08	.16	.60**												
11. WL Consolidation	.22	.25*	.19	.24*	.14	.17	.17	.17	.59**	.79**											
12. ISR (NV)	-.18	.33*	.15	.14	.46**	.40**	.28**	.40**	.23*	.05	-.03										
13. ISR (V)	.18	.52*	.28*	.31**	.33**	.33**	.24*	.40**	.49**	.35**	.27*	.36**									
14. Contextual Cueing NV	.12	.11	.10	.20	-.02	-.07	-.11	-.02	.06	.02	.04	-.02	.02								
15. Contextual Cueing V	.09	-.02	.05	.11	-.01	.06	.08	-.02	-.04	.02	.22	.11	-.07	-.10							
16. NV-SRT1 RT	-.03	-.03	-.20*	-.06	-.05	-.07	-.09	-.18	.10	-.03	-.04	-.25*	.08	-.03	-.12						
17. NV-SRT2 RT	.03	.03	.16	.01	.13	.07	.10	-.01	.15	.19	.26*	-.01	-.01	-.18	.15	.21					
18. V-SRT1 RT	-.01	.01	.06	-.02	.15	-.09	-.05	-.17	.14	.28**	.38**	-.15	-.01	.06	.12	-.08	.24*				
19. V-SRT2 RT	.04	.01	-.04	-.03	.09	.06	.11	.10	-.10	-.12	-.09	.02	.07	-.10	-.12	.11	.20	-.00			
20. Hebb NV	.06	.05	-.01	.12	.03	.07	.03	-.04	.14	.04	.10	.10	-.04	.12	-.05	-.02	.03	-.12	-.14		
21. Hebb V	-.08	.13	.04	.10	.14	.07	.10	.14	.09	-.03	-.02	.23*	.14	-.08	.01	.09	.03	.14	-.00	-.15	

*p < .05; **p < .01

5.3.4 Effects of children's language background

It was important to check that the pattern of results obtained was not influenced by differences between monolingual children and those with English as an additional language (EAL). As shown in Table 5.6 there were no statistically significant differences in language attainment between the EAL and monolingual children after Bonferroni corrections for multiple comparisons were made; and the EAL children actually performed slightly but non-significantly better than the monolingual children on tests of word reading. Effect sizes for the TROG-2 show that the level of grammatical proficiency demonstrated by the EAL children was lower than their English mother-tongue counterparts. Twenty of the monolingual children scored over 75 out of 80 on the TROG-2 task, compared to eleven of EAL children, who showed a greater range of scores.

Table 5.6 Language attainment means and standard deviations by monolingual and EAL subgroups and *t*-test comparisons.

Attainment test	Mean (SD)		*t*(df = 98)	*p*	*Cohen's d*
	Monolingual (*n* = 49)	EAL (*n* = 52)			
TROG-2 (Blocks passed)	15.85 (3.21)	14.69 (3.18)	1.81	.07	.36
Trog-2 (Total correct)	73.02 (6.18)	70.23 (6.62)	2.17	.03	.43
WRAT-3	12.12 (3.21)	12.10 (2.47)	.05	.96	.01
PWM	36.77 (11.73)	38.67 (10.07)	-.87	.38	-.17
TOWRE-2 Words	57.65 (16.45)	60.04 (9.85)	-.81	.38	-.18

Crucially, correlations that included only the monolingual children showed the same pattern as those for the overall sample (see Table 5.7). The relationship between outcome measures and predictors showed no evidence of meaningful group differences between the EAL fluent English speakers and monolingual (EMT) children, other than the monolingual group exhibited even stronger relationships between verbal declarative memory and language (TROG-2) than the sample as a whole (verbal free recall score: monolingual $r = .52$; overall $r = .48$; verbal ISR

205

learning score: monolingual $r = .71$; overall $r = .52$), while the correlations with non-verbal declarative free recall scores ceased to be significant. The implicit learning measures for the monolingual group still failed to correlate significantly with language or with each other. This is consistent with the study's findings that it is specifically verbal declarative memory that relates to language, demonstrating that the pattern of results in the main study is not an artefact of having EAL fluent English speakers in the sample.

Table 5.7 Correlations between language (TROG-2) and other measures by subgroup and overall sample.

Measures	TROG-2		
	Complete Sample (*n* = 101)	EMT (*n* = 49)	EAL (*n* = 52)
Literacy composite	.49**	.58**	.46
Arithmetic composite	.38**	.38	.44
WASI	.36**	.53*	.26
Dot Locations (DL)			
Learning	.36**	.33	.44
Delay	.32**	.23	.39
Consolidation	.37**	.41	.39
Word Lists (WL)			
Learning	.48**	.52*	.51*
Delay	.30**	.39	.24
Consolidation	.25*	.41	.13
ISR (NV)	.33*	.41	.31
ISR (V)	.52*	.71**	.36
NV-SRT1 RT difference	-.03	-.29	.14
NV-SRT2 RT difference	.03	-.05	.12
V-SRT1 RT difference	.01	.04	-.06
V-SRT2 RT difference	.01	-.11	-05
Hebb NV	.05	-.12	.20
Hebb V	.13	-.13	.28
Contextual Cueing NV	.11	.21	.06
Contextual Cueing V	-.02	.12	-.07

*p < .05; **p < .01; Bonferroni corrections applied; EMT = English mother tongue; EAL = English as an additional language

5.3.5   Confirmatory Factor Analysis

Given the low reliabilities of the measures of implicit learning, and the low correlations between these measures, they were not considered further.   A confirmatory factor analysis model for the eight declarative memory measures and

measures of attainment for the complete sample of children was estimated in Mplus 7.4 (Muthén & Muthén, 1998-2016) with missing values being handled with Full Information Maximum Likelihood estimation. The model used is shown in Figure 5.10 and provides an excellent fit to the data (χ^2 (38) = 40.60, p = .356; RMSEA = .026 [90% CI .000 -.076]; CFI = .99; TLI = .99). In this model the verbal and non-verbal declarative memory measures defined two separable factors which correlated moderately with each other (r = .29). The verbal factor correlated moderately with measures of attainment (language (TROG-2) r = .54; literacy r = .28; arithmetic r = .34). The non-verbal factor did not correlate significantly with literacy or arithmetic, but did correlate with language as measured by TROG-2 (r = .32).
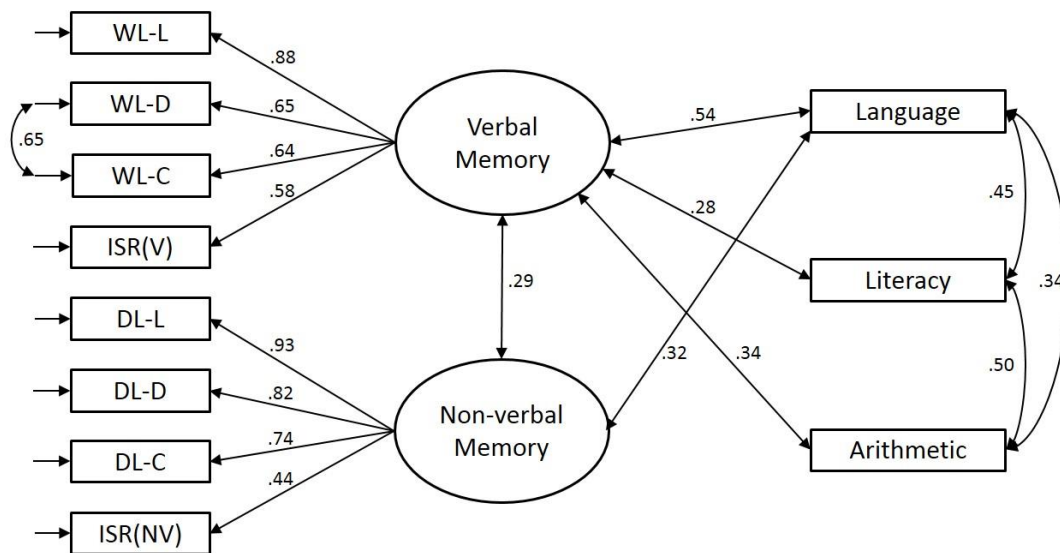
Figure 5.10 Confirmatory factor analysis showing relationship of memory and attainment tasks to latent variables of verbal and non-verbal memory. WL-L = Word Lists learning score; WL-D = Word Lists delay score; WL-C = Word Lists consolidation score; DL-L = Dot Locations learning score; DL-D = Dot Locations delay score; DL-C = Dot Locations consolidation score: ISR(V) = verbal immediate serial recall; ISR(NV) = non-verbal immediate serial recall; Language = TROG-2 total score; Literacy = Literacy composite of WRAT spelling, TOWRE word and non-word reading and Picture Word Matching; Arithmetic = composite of TOBANS addition, subtraction and multiplication subtests.

5.3.6   Investigating attainment using an ability groups design

The relationship between implicit memory and language in children has been most frequently investigated using extreme groups designs that enter performance on

different types of sequence in the serial reaction time task, matrices in contextual cueing or repeated and random sequences in Hebb learning directly into mixed ANOVAs or ANCOVAs, obviating the need to calculate difference scores first (e.g., Majerus et al., 2009; Archibald & Joanisse, 2013; Lum et al., 2010; Hedenius et al., 2013; Stoodley, Harrison, & Stein, 2006; Jiménez-Fernández et al., 2011). While the use of a group design in this study was rejected on methodological grounds, mixed ANOVAs that allocated children to groups based on a median split of the language factor were nevertheless run for all tasks to exclude the possibility of a relationship between implicit memory and language existing when data was analysed in this fashion. This analysis was done simply to facilitate comparison with previous studies. The between subjects factor of language attainment was not significant for all but two of the implicit learning tasks (see Table 5.8). It was a significant factor for the non-verbal contextual cueing condition and the non-verbal Hebb task, but only in so far as children scoring below the median for language ability were slower for both matrix types on the cueing task and recalled fewer stimuli for both repeated and random sequences on the Hebb task than children above the median. Neither of these differences related to implicit learning. Given the range of methodological problems with analyzing data in this fashion (MacCallum, Zhang, Preacher, & Rucker, 2002; Irwin & MacClelland, 2003), these results were not considered further.

Table 5.8 Mixed ANOVA results for implicit tasks.

Task	F	p	η^2_p
NV-SRT1	$F(1,96) = .48$	.492	
NV-SRT2	$F(1,87) = 1.11$	.294	
V-SRT1	$F(1,90) = .215$	.644	
V-SRT2	$F(1,84) = 1.33$	.252	
Contextual Cueing NV	$F(1,98) = 8.28*$	.005	.078
Contextual Cueing V	$F(1,98) = 2.89$	.093	
Hebb NV	$F(1,84) = 5.00*$	.028	.056
Hebb V	$F(1,86) = 1.98$	.163	

5.4 Discussion

This study assessed the claims of the procedural deficit hypothesis that impairments in a procedural learning system are a causal risk factor for language learning deficits in children with developmental language disorder and dyslexia (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005). The study examined the relationship between both implicit and explicit memory skills and language-related attainment in a large sample of children aged 7 – 8, using verbal and non-verbal versions of all tasks. Results showed that verbal declarative memory skills related to language attainment, supporting much previous research in this area. The implicit memory tasks, on the other hand, did not correlate with language at all. Additionally, although aggregated mean response times for the implicit memory tasks showed considerable evidence of implicit learning, none of the implicit tasks proved to be reliable indicators at an individual level. This has clear implications for research seeking to establish a causal relationship between procedural learning and language.

As documented in Chapters 3 and 4 many studies have reported deficits on a range of implicit learning measures in children with language disorder (Hedenius, 2013; Hsu & Bishop, 2014; Lum et al., 2010) or dyslexia (Vicari et al., 2005; Howard et al., 2006). However, the findings from previous studies are distinctly mixed, with many null results (Gabriel et al., 2011; Lum & Bleses, 2012; Staels & Van den Broeck, 2015; Majerus et al., 2009). Methodologically, most studies in this area share a number of undesirable characteristics: 1. The studies use extreme group designs. 2. Sample sizes are small, giving low statistical power. 3. Only a single measure, or a limited range of measures of learning and memory are used in any one study. 4. The studies do not report reliability estimates for the measures of learning.

Extreme group designs will tend to overestimate the extent of any true linear relationship between two variables in the population as a whole. Decreased reliability, inflated effect size estimates, misclassification of participants into groups and problems related to regression to the mean for extreme scores are all potential issues for this design (Preacher et al., 2005; DeCoster, Iselin, & Gallucci, 2009).

Additionally, studies with low statistical power suffer from both type 1 and type 2 errors to a far greater extent than high-powered studies (Button et al., 2013) and are likely to yield many false positive results as a result. They are also more likely to suffer from vibration effects, whereby the significance or not of the results depends as much on the methods of analysis used, as on any underlying true effects (Ioannidis, 2008). This is an important point given the variety of different methods that are recruited for screening, confound control and analysis across the range of procedural learning and language studies.

The solution to these problems taken by the current study was to administer a wide range of measures of procedural and declarative learning and language and attainment to a large and representative sample of children. Clear evidence of learning was found in the procedural memory tasks, but such measures proved to have extremely low reliabilities, consistent with some previous evidence (Buchner & Wippich, 2000; Salthouse, McGuthry, & Hambrick, 1999; Reber et al., 1991.

One potentially important determinant of the reliability of any task is the number of trials used (Nunnally & Bernstein, 1994). The length of implicit learning tasks used in this study was similar to the length of tasks used by many others in the field. Serial reaction time tasks have occasionally used over 1000 trials (Rüsseler et al., 2011; Kelly et al., 2002), but they have often been much shorter, with some including as few as around 300 trials (Lum & Bleses, 2012; Vicari et al., 2005; Stoodley, Harrison, & Stein, 2006; Menghini et al., 2006). The length of contextual cueing tasks varies across studies, but evidence of cueing in children has been shown in tasks containing as few as 80 trials in total (Dixon et al., 2010). The number of Hebb repetitions used here was the same as in Hsu and Bishop (2014). The reliability of the implicit learning tasks in this study is, therefore, likely to be broadly comparable to the reliabilities of measures used in previous studies in this area. Further research will investigate whether increases in the number of trials used in procedural learning tasks such as those used here will result in estimates of learning with adequate reliability.

In addition, although children over the age of six are able to cope with the demands of cognitive testing across multiple tasks, they are more prone to boredom and fatigue than adults (Luciana & Nelson, 2002), with resultant down-stream effects on the quality of data they produce. For example, it has been demonstrated that children can be inconsistent performers on tasks such as Hebb learning compared to adults (Mosse & Jarrold, 2008; Archibald & Joanisse, 2013), which may explain the unreliable results on this task in particular.

Evidence from the current study seriously questions the viability of the procedural deficit hypothesis. It is clear, however, that in order to adequately test such a hypothesis more work will be required to develop measures of procedural learning with adequate reliabilities. If reliable measures can be developed, only then can the procedural learning hypothesis be adequately assessed. The mixed evidence to date for this hypothesis likely reflects the low statistical power (and unreliable measures) of studies in this area.

In contrast to the findings for procedural learning, measures of declarative memory showed reasonable reliabilities and moderate correlations with measures of language skills and academic attainment. The correlation found here between the measure of verbal serial recall and measures of attainment are in line with many earlier findings. For example, Melby-Lervåg, Lyster, and Hulme (2012) reported a robust correlation between measures of immediate verbal memory span and reading ability (pooled effect size estimate ($r = .34$)). Similarly, verbal free recall performance is typically poor in children with dyslexia or developmental language disorder (Kramer, Knee, & Delis, 2000; Baird, Dworzynski, Slonims, & Simonoff, 2010). Finally, the correlation between arithmetic performance and verbal declarative memory, as well as with both language and literacy measures, is consistent with research that language, in particular the manipulation of verbal codes, is integral to arithmetical fluency (Durand, Hulme, & Larkin, 2005; Simmons & Singleton, 2007).

The above correlations may or may not reflect causal effects of declarative memory on the development of reading and language skills, since some have argued that

212

phonological processing deficits and verbal memory impairments in dyslexia are two expressions of the same underlying problem (Tijms, 2004) and that verbal short-term memory skills may be a by-product of the mechanisms that subserve language itself (Hulme & Snowling, 2009; Acheson, Hamidi, Binder, & Postle, 2011; Allen & Hulme, 2006).

The results of the current study also showed that non-verbal declarative memory shared a relationship with language attainment to a degree, as well as being moderately correlated with verbal declarative memory. Some previous research has found non-verbal memory performance to relate to language impairments in dyslexia and developmental language disorder (Menghini et al., 2010; Bavin, Wilson, Maruff, & Sleeman, 2005), but others have found no link with language disorder (Vellutino, 1979; Alloway & Archibald, 2008; Nation, Adams, Bowyer-Crane, & Snowling, 1999. However, it is possible that the non-verbal tasks were not entirely modality-pure, with children using verbal cues to aid recall. Equally, the relationship between non-verbal declarative memory skills and the language attainment measure of receptive grammar, in particular, may have been bolstered by shared domain-general memory demands of the tasks. The lack of a relationship between non-verbal declarative memory and the less memory-intensive literacy measures is consistent with this explanation.

In summary, this study has shown that verbal declarative measures correlate with language attainment, yet in spite of considerable evidence of implicit learning on most implicit tasks, no relationship between implicit learning and language attainment was found. Crucially, the derived measures representing implicit learning displayed very low reliability. The development of implicit tasks with demonstrable reliability is needed, before any questions about the relationship between implicit procedural learning and language can be answered definitively.

Chapter 6   Study 2: Test-retest reliability of an extended serial reaction time task in adults

6.1   Introduction

The serial reaction time task administered to children in the study reported in Chapter 5 contained 500 trials, which is a fairly typical within the field. However, in spite of acceptable levels of internal consistency, this task proved to have low test-retest reliability. Task length in cognitive testing has been found to be a determinant of reliability (Nunnally & Bernstein, 1994; Charter, 2003). For this reason a longer task was trialed in healthy adults to investigate whether increasing the length of the task would serve to improve its reliability. Measures of reading fluency were also taken, in order to examine whether any relationship with procedural learning on the serial reaction time task existed in adults.

6.2   Method

This was a correlational study assessing the test-retest reliability of a serial reaction time task with 1500 trials in a sample of healthy adults.

Ethical clearance for the study was given by the UCL Language and Cognition Department's ethics committee. An opportunity sample of forty six healthy adult participants were recruited to take part. Ages ranged from 18 to 61 (mean = 25 years and 4 months; $SD = 10$ years and 4 months). Thirty two were female and fourteen were male. Forty four of the participants were right handed. Sixteen spoke English as an additional language, but only five of these rated themselves as not fluent in English.

6.2.1.1   Tasks and testing procedures

Two extended probabilistic SRT tasks were developed with an identical probabilistic sequence structure to the SRT tasks in the first study reported in Chapter 5. The length of the tasks was increased from 500 trials to 1,500 trials. The first task used the sequences from NV-SRT, taken from Shanks et al. (2003). The second task used the sequences from the verbal analogue task in the first study, taken from

Schvaneveldt & Gomez (1998). Just as in the first study, five breaks were scheduled during the tasks, occurring every 300 trials (compared to every 100 trials in the first study). Participants completed both tasks on different occasions, two or three days apart.

Reading fluency of words and non-words was also assessed using the TOWRE-2 word and non-word reading tests (Torgesen, Wagner & Rashotte, 1999). TOWRE test scores for the five participants who were not fluent in English were not included in analysis. Tasks were administered in fixed order for all participants. Both short TOWRE reading tests were administered at the beginning of the first testing session.

6.3 Results

All but 7 participants took the serial reaction time tasks 2 or 3 days apart. However, scheduling constraints meant that three participants took the tasks on consecutive days and four had a 4 day gap. Data was screened in an identical fashion to Study 1, with all inaccurate trials and trials over 5000 ms removed and a subsequent moving criterion based on sample size (Selst & Jolcoeur, 1994) applied to remove the remaining outlying observations. RTs for the first two trials in each block were excluded, because their locations were not predictable.

6.3.1 Analysis of Response Times (RTs)

For analysis, blocks were consolidated into epochs of 300 trials. Means and standard deviations for RTs per sequence and epoch for both task attempts are shown in Table 6.1. RTs for the probable trials remained fairly constant throughout the first task attempt, but increased in the last epochs of the second attempt, suggestive of fatigue. RTs for the improbable sequences increased in every epoch on both task attempts.

Table 6.1 Means and standard deviations for SRT task RTs in milliseconds by sequence.

	SRT Time 1 sequence means (SDs)		SRT Time 2 sequence means (SDs)	
	Probable	Improbable	Probable	Improbable
Epoch 1	388.91 (129.37)	470.54 (130.01)	374.78 (114.08)	457.22 (112.61)
Epoch 2	379.97 (137.26)	484.25 (118.03)	378.06 (123.63)	481.09 (118.65)
Epoch 3	380.24 (142.37)	494.89 (133.06)	382.62 (123.54)	507.65 (156.05)
Epoch 4	377.81 (136.16)	500.15 (125.05)	398.72 (128.09)	519.34 (140.32)
Epoch 5	383.10 (137.79)	518.49 (124.69)	405.04 (131.38)	529.97 (126.38)

Mixed effect models (Rabe-Hesketh & Skrodal, 2012) in Stata (13.0) were chosen to analyse response times for both attempts at the extended serial reaction time tasks (see Figure 6.1). The tasks were analysed separately. Sequence, epoch and the interaction between them were entered as fixed effects and items and participants as crossed random effects.

Sequence was a significant predictor of RT at both Time 1 (unstandardized regression coefficient = 68.642, $z = 21.33$, $p < .0001$, 95% CIs [62.33, 74.95]) and at Time 2 (unstandardized regression coefficient = 72.71, $z = 22.39$, $p < .0001$, 95% CIs [66.34, 79.07]). The interaction between the sequence and epoch was also significant for both tasks (Time 1 Epoch $1 - 5$ unstandardized regression coefficient = 45.47, $z = 9.47$, $p < .0001$, 95% CIs [36.06, 54.89]; Time 2 Epoch $1 - 5$ unstandardized regression coefficient = 36.79, $z = 7.73$, $p < .0001$, 95% CIs [27.46, 46.11]), with the difference between sequences at Time 1 increasing throughout the task from 81.63 ms in Epoch 1 to 135.39 ms in Epoch 5. This was due to the ever-increasing RTs in the improbable sequence. The increase in difference between the sequences followed a similar pattern at Time 2 rising from 82.44 ms in Epoch 1 to 124.93 ms in Epoch 5. Notably, RTs for both sequence types slowed over the course of the task at Time 2, with the improbable sequence RTs slowing at a greater rate than the probable sequence RTs.

The effect of epoch was also significant on both occasions, but in opposite directions. At Time 1 RTs decreased over the course of the task (Epoch $1 - 5$

216

unstandardized regression coefficient = -6.58, z = -4.80, p < .0001, 95% CIs [-9.27, -3.89]), while Time 2 RTs increased in every epoch (Epoch 1 − 5 unstandardized regression coefficient = 30.00, z = 22.21, p < .0001, 95% CIs [27.35, 32.65]), suggesting fatigue and / or lack of motivation played an increasing role during Time 2.
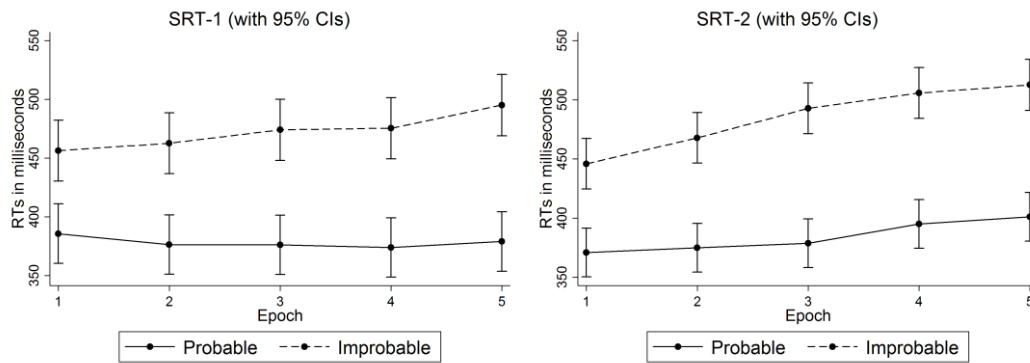
Figure 6.1 Graph showing RTs per sequence and epoch for both task attempts, with 95% confidence intervals.

6.3.2   Analysis of error frequencies

The same error analyses used in the first study were again undertaken to examine implicit learning on the adult pilot tasks. Means, standard deviations and reliabilities for error measures (both component error frequencies and difference score measures) are in Table 6.2.

Table 6.2 Means (SDs) and test-retest reliabilities for component error measures and difference scores

	SRT1	SRT2	Reliability (r)
Sequence errors analysis			
Probable errors*	43.93 (35.82)	61.46 (44.45)	.78
Improbable errors	33.83 (32.58)	38.26 (30.07)	.82
Proportional sequence errors difference score	28.94 (30.07)	31.43 (26.43)	.81
Error type analysis			
Anticipatory errors on improbable sequence	32.02(32.20)	36.20 (30.42)	.82
Random errors on improbable sequence*	1.80 (1.78)	2.07 (1.74)	.14
Proportional error type difference score	31.12 (32.05)	35.16 (30.63)	.82

*Before proportional adjustment.

The sequence errors analysis showed that there were proportionately more errors on the probable sequence than on the improbable sequence on both tasks (SRT1: unstandardized regression coefficient = 3.331, z = 57.71, p < .001, 95% CI [3.22, 3.44]; SRT2: unstandardized regression coefficient = 4.437, z = 86.86, p < .001, 95% CI [4.34, 4.54]) and that errors increased at a greater rate on the improbable sequence than on the probable sequence over the course of the task (SRT1: unstandardized regression coefficient = 4.193, z = 51.37, p < .001, 95% CI [4.03, 4.35]; SRT2: : unstandardized regression coefficient = 2.701, z = 37.37, p < .001, 95% CI [2.56, 2.84]). These results are highly suggestive of implicit learning of the probable sequence that continues to develop over the course of the tasks (see Figure 6.2).
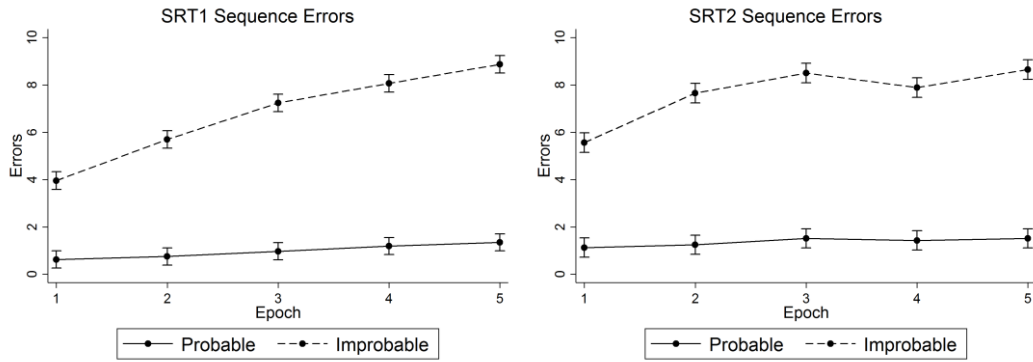
Figure 6.2 Error frequencies for the extended 1500 trial SRT tasks per sequence. Error bars are 95% confidence intervals.

The second analysis investigated whether there were more anticipatory than random errors on the improbable sequence compared to the probable one. Figure 6.3 shows that there were proportionately significantly more anticipatory than random errors on both task attempts (SRT1: unstandardized regression coefficient = -4.201, z = -4.79, $p < .001$, 95% CI [-5.92, -2.48]; SRT2: unstandardized regression coefficient = -2.933, $z = -4.41$, $p < .001$, 95% CI [-4.23, -1.63]). Indeed, this was the case regardless of any proportional adjustment of random errors. However, the interactions between error type and epoch were not significant, as the ratio of anticipatory to random errors stayed at similar levels across the task.
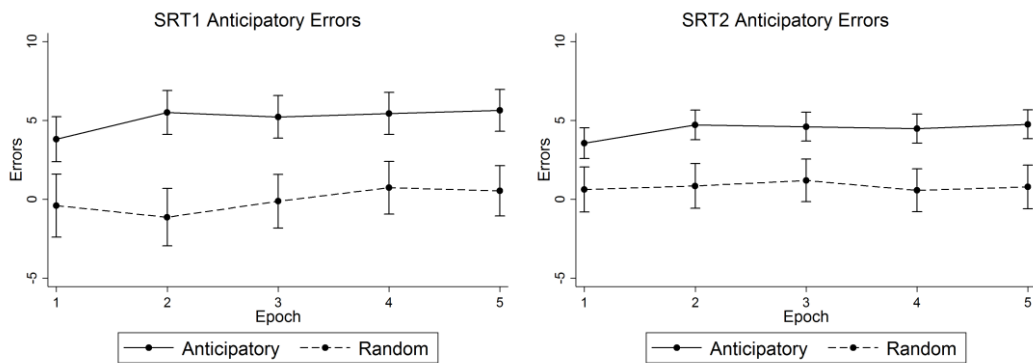
Figure 6.3 Error frequencies for anticipatory and random errors on the improbable sequence of the verbal SRT tasks. Error bars are 95% confidence intervals.

219

6.3.3 Task measures and reliabilities

RT measures. Split half and test-retest reliability was established for response times. An overall difference score per task for implicit learning was calculated by subtracting the participant's mean probable RT across the whole task from their mean improbable RT. Subsequently, similar measures were calculated for the 1st 500 trials and the 1st 1000 trials of the task, in order to examine reliability at different lengths of the task. Means, standard deviations and reliabilities for these difference scores are in Table 6.3.

Table 6.3 Response Times difference score means, standard deviations and reliabilities

Task	Difference Score	N	Mean	SD	Reliability
SRT1	Total	46	107.04	(70.05)	.95[s] / .68[r]
	1st 500 trials	46	83.82	(67.48)	.95[s] / .53[r]
	1st 1000 trials	46	96.06	(69.73)	.92[s] / .66[r]
SRT2	Total	46	108.50	(54.49)	.88[s] / .68[r]
	1st 500 trials	46	87.69	(54.15)	.71[s] / .52[r]
	1st 1000 trials	46	103.33	(55.02)	.84[s] / .66[r]

[s] = split-half reliability; [r] = test-retest reliability

Split-half reliability for the overall mean difference in RTs was calculated by correlating odd and even numbered trials per sequence and was good for both Time 1 ($r = .95$) and Time 2 ($r = .88$). Internal consistency for the first 500 trials was also calculated to give a measure that could be more directly compared to the internal consistency of the 500 trial SRT task in Study 1. It was also good at Time 1 ($r = .95$), but dropped at Time 2 ($r = .71$). Split-half reliability for the first 1000 trials at Time 1 was ($r = .92$) and at Time 2 was ($r = .84$), which approached the level of correlations for the overall tasks containing 1,500 trials.

Test-retest reliability for the tasks using the overall mean difference between response times for the sequences for each task was moderate to good ($r = .68$). There was evidence that the length of the task improved test-retest reliability, as the correlation between the mean differences per block increased over the course of the task from a low starting point for the first 100 trials ($r = .28$). Test-retest reliability for the first 500 trials (the length of the SRT tasks in Study 1) was ($r = .53$), increasing to ($r = .66$) for the first 1000 trials. Test-retest reliability improved only slightly (by .02) when the difference scores from the final 500 trials were added to the analysis.

Additionally, when variability in participants' baseline speed of response was controlled by dividing the difference between sequences by participants' overall speed of response (improbable sequence − probable sequence) / ((improbable sequence + probable sequence)/2) test retest reliability for the task improved further (1st 1000 trials r = .71; 1500 trials r = .70).

This suggests that cutting the task to 1000 trials would not have a large effect on the internal consistency or test-retest reliability of the task. Importantly, fatigue effects in children may be less of an issue with a 1000 trial task than a 1500 task.

Error measures. These were calculated in the same way as in Chapter 5 (see Table 6.2). Test-retest reliability for the difference between the frequency of errors on the two sequences and for the difference between anticipatory errors on the improbable sequence were good (Sequence errors test-retest $r = .81$; Error type test-retest $r = .82$), but once again the frequency of random errors on the improbable sequence were low, with 11 of the 46 participants making no random errors on SRT1 and the same number making no random errors on SRT2. More importantly, once the baseline frequency of errors made by participants on the task was controlled [sequence errors: (Improbable errors − Proportional probable errors) / (Improbable errors + Proportional probable errors); Error type on the improbable sequence: (anticipatory improbable errors − proportional random improbable errors) / (anticipatory improbable errors − proportional random improbable errors)] the reliability of the measures was much reduced (Sequence errors test retest $r = .14$; Error type test retest $r = .44$), suggesting

that much of the reliability of the sequence errors measure was a result of the variability in the overall level of errors by participants across both sequences.

These reliabilities suggest that implicit learning measures derived from error frequencies are either not reliable once variability in baseline frequency of errors is controlled (Sequence errors) or are subject to floor effects (Error type on the improbable sequence) in adults just as they are in children.

6.3.4　The relationship with reading fluency

There was no relationship at all between TOWRE word or TOWRE non-word reading and any of the RT or error-based implicit learning measures on either SRT task (r's from -.06 to -.11). TOWRE word and non-word reading correlated with one another at (r = .65)

6.3.5　The effect of motivation and boredom on task performance

At the end of the second testing session participants were asked to rate their enjoyment of the serial reaction time task on a scale of 1 to 10, with 1 equal to "not enjoyable at all" and 10 equal to "very enjoyable". The mean rating was 3.74, with 48% of the sample rating the tasks at 3 or lower, suggesting the experience of completing two serial reaction time tasks of this length was not a pleasant one.

Counter-intuitively, self-rated task enjoyment was negatively correlated to the implicit learning response time measure, such that enjoying the tasks less correlated with better implicit learning (Time 1 r = -.32; Time 2 r = -.40). When task enjoyment was correlated to the 1st, 2nd and last thirds of the tasks, the negative relationship was most pronounced during the last third of the first task (r = -.44). Correlations between task enjoyment and error rate measures of implicit learning were not significant.

6.4　Discussion

Implicit learning and test-retest reliability of an extended version of the probabilistic serial reaction time task was examined in an opportunistic sample of 46

healthy adults. Identical tasks with differently numbered sequences, were administered twice on separate occasions. Measures of reading fluency were also taken. Results showed robust levels of implicit learning on both occasions and moderate to good test-retest reliability for the majority of RT-based implicit learning measures. No relationship between implicit learning and reading fluency was found. These results show that the length of a serial reaction time task influences its reliability as expected from classical test theory (Nunnally & Bernstein, 1994).

These results are in line with recent findings by Siegelman & Frost (2015), who established test retest reliability of a probabilistic serial reaction time task with 960 trials with a sample of 76 adults. The probabilities associated with the probable and improbable sequences for their task were .85 and .15 respectively (as opposed to .9 and .1 in the current task) and it was taken on two occasions approximately 3 months apart, using the same sequences on both occasions. The corresponding test-retest reliability was $r = .47$. The reliability of the current task is better than this, but the testing lag was only one week.

The results of the current study also suggest that controlling for overall speed of response delivers measureable improvement in reliability for RT implicit learning measures. The reduction in reliability for the error frequency measures as a result of controlling for baseline frequency could be interpreted as further evidence that these measures are not reliable indicators of implicit learning.

However, split-half and test-retest reliability of the first 500 trials of the extended tasks in adults in the current study were already considerably higher than in the 500 trial task with children in Chapter 5, suggesting that participant age may also be a factor in the reliability of the serial reaction time tasks and that improvements in reliability in the current study should not be attributed to task length alone.

There was an obvious rationale for piloting the extended serial reaction time task in adults: that of practicality. However, if implicit learning is age-independent, as has been claimed (eg, Kirkham et al., 2007), then the reliability of an implicit learning

223

measure in adults should be no different in children. That is, as long as the measure of implicit learning is accurately indexing implicit learning and is not confounded by variables that are age-dependent, such as for example, executive function or attention. It is, therefore, possible that adults may perform more reliably on the SRT tasks because they are both more motivated and are better able to sustain attention across the task than are children.

The finding that levels of implicit learning in adults were negatively correlated with self-rated enjoyment could be tentatively linked to this interpretation. Sustaining attention across a lengthy cognitive task requires effort, which may correspond to decreased enjoyment. Children may find it harder than adults to sustain their attention throughout both attempts at such a demanding task (Betts, Mckay, Maruff, & Anderson, 2006). The increase in RTs towards the end of the second attempt at the task in this study suggests that fatigue may become a confounding factor in lengthy tasks, even in adult participants. It is therefore, important to identify the optimum task length to balance reliability with participant motivation.

Finally, no relationship between implicit learning and reading fluency was found, but this result should not be given undue weight. The sample was relatively small; included English as an additional language speakers whose fluency was self-rated; and since the majority of participants were university students, there may not have been sufficient variation in reading fluency scores to reveal a relationship with implicit learning.

To conclude, the results of this extended serial reaction time task in adults suggests that task length is a determinant of reliability. Investigation of the reliability and relationship to attainment of an extended SRT task in children therefore seems an obvious next step. However, participant age may also influence test-retest reliability to an, as yet unknown, degree. For this reason it is recommended that sustained attention and task engagement is monitored to explore whether this is a factor in the reliability of implicit learning in this younger age group.

Chapter 7   Study 3: An extended serial reaction time task in children

Following on from the results of the studies in Chapters 5 and 6, this study investigated the claims of the procedural deficit hypothesis by examining the relationships between measures of procedural and declarative memory and measures of attainment in a large sample of 7 and 8 year-old children unselected for ability.

The most widely used measure of procedural learning in studies of the procedural deficit hypothesis is the serial reaction time task (SRT: Nissen & Bullemer, 1987). As shown in Chapters 3 and 4, results from studies using the serial reaction time task with children are mixed, with some studies finding evidence of implicit learning on the task in children with dyslexia and developmental language disorder and others finding no evidence of such impairments (for reviews see: Lum, Ullman, & Conti-Ramsden, 2013; Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Schmalz, Altoè, & Mulatti, 2016).

There are a number of possible reasons why the results from studies investigating the relationship between the serial reaction time task and measures of language and reading skills are so mixed.

1. The majority of studies use extreme groups designs, comparing children with dyslexia or language impairment with an age-matched control group with normal language skills. Extreme groups designs are subject to a number of methodological limitations (Preacher, Rucker, MacCallum, & Nicewander, 2005), such as regression to the mean for extreme scores, misclassification into groups, inflated effect sizes and decreased reliability. Sample sizes are also typically small, further compounding these issues.

2. Reliability for the serial reaction time task may be poor. Difference scores, such as those used to measure procedural learning on the serial reaction time task, are frequently unreliable (Lord, 1958; Overall & Woodward, 1975). Previous studies have

not reported the reliability of serial reaction time tasks used, but the reliability of the task used in the Chapter 5 study was extremely low (test-retest reliability $r = .21$).

3. The putative poor reliability of the serial reaction time tasks typically used in studies of language learning disorders in children may, in part, be due to their length. Procedural learning tasks in children are typically kept short in an effort to maintain motivation and avoid fatigue (e.g., Perlant & Largy, 2011). However, task length is usually an important determinant of reliability (Nunally & Bernstein, 1994) and an extended serial reaction time task with 1500 trials piloted in adults in Chapter 6 demonstrated moderate to good reliability. The small number of trials used in many serial reaction time tasks with children is, therefore, a cause for concern.

4. Finally, self-regulatory mechanisms develop with age (Putzke, Williams, Adams, & Boll, 1998), so unreliable procedural learning scores may be a particular issue in children, where levels of sustained attention vary and may influence learning. This may be a particular complication for children with language learning disorders, since a variety of developmental cognitive disorders, including language disorders, are frequently comorbid with attentional difficulties (Finneran, Francis, & Leonard, 2009; Rabiner, Coie, & The Conduct Problems Prevention Research Group, 2000).

In summary, the current study uses a concurrent correlational design with a large sample of children unselected for ability. Procedural learning is assessed using a serial reaction time task with a large number of trials in an attempt to ensure that the task is reliable. By testing all participants on two separate occasions, test-retest reliability can be assessed. In addition, a novel procedure of rating the child's level of attentional engagement with the serial reaction time task on both occasions is used. Measures of attainment (reading, language [grammar] and arithmetic) are related to performance on the serial reaction time task as well as to measures of verbal declarative memory.

7.1 Method

A concurrent correlational design was used to explore the relationships between procedural and declarative memory skills, attention and attainment in 7 and 8 year-old children. Ethical approval for the study was granted by the UCL Research Ethics committee.

7.1.1 Participants

One hundred and twelve primary school children (mean age = 92. 11 months; *SD* = 5.71; range 83 – 105 months) unselected for ability took part in the study (53 girls and 59 boys). Children in participating schools were enrolled on an opt-out basis. Children registered as speaking English as an additional language were excluded, unless they were born in the UK; had attended a UK primary school since the beginning of Reception (age 4-5) or earlier; confirmed that they spoke English to at least one parent at home; and were judged to be completely fluent English speakers by their class teachers. Seventeen children fitted this description and were included in the study.

7.1.2 Tasks and testing procedures

Testing took place across three sessions. Children first completed a battery of attainment measures during a whole-class testing session. Two individual sessions for each child were then scheduled a week apart to complete the memory measures. Tasks were administered in the same fixed order to all children.

7.1.2.1 Attainment Tasks

A battery of the same language-related attainment tasks used in the initial study with children in Chapter 5 were used:

- Test of Reception of Grammar (TROG-2: Bishop, 2003).
- Test of word and non-word reading efficiency (TOWRE-2: Torgesen, Wagner, & Rashotte, 1999).

- Test of basic arithmetic and number skills (TOBANS: Brigstocke, Moll, & Hulme, 2016).

- Wechsler Abbreviated Scale of Intelligence (WASI: Wechsler, 1999).

7.1.2.2 Verbal Declarative Memory tasks

Word Lists (Cohen, 1997). This measure of verbal free recall from the Children's Memory Scale was also used in the initial study in Chapter 5. It required children to recall as many words as possible in any order from a list of 10 spoken words (Trial 1). After the first presentation only words omitted by the children were re-presented (Trials 2 – 4). After a distractor list children were asked to recall the initial list again without re-presentation (Trial 5). These trials were summed to form the child's Learning Score. Recall was requested once more at the end of the testing session (Trial 6: Delayed Recall) and during the second individual testing session a week later (Trial 7: Consolidation). The testing lapse for the consolidation measure was not the same for all children owing to scheduling constraints (93% of children were assessed with a lag of between 6 and 8 days and 7% of children with a lag of 2 or 5 days), but there was no significant difference in consolidation scores for those with a smaller or greater testing interval.

7.1.2.3 Procedural Learning task

Serial Reaction Time Task (SRT: Nissen & Bullemer, 1987*).* The two versions of the probabilistic serial reaction time task used in the pilot study in Chapter 6 were adjusted to a length of 1000 trials, since test-retest reliability for the first 1000 trials of the pilot task was moderate to high ($r = .66$, increasing to .71 once baseline RT was controlled) and showed little improvement when computed over all 1500 trials. The task began with 10 practice trials, with equal probabilities of each sequence occurring. There were then 10 blocks of 100 trials divided into 5 epochs. The task took 20-25 minutes to complete.

7.1.2.4 Attention Measure

In order to quantify the attention paid by each child to the serial reaction time tasks a 9 point rating scale was devised.

228

1 = Did not complete Epoch

2 = Poor attention throughout; continual stopping, talking, difficulty with restarting

3 = Distracted multiple times throughout with talking and short pauses (7 or more)

4 = Some talking and short pauses throughout, getting worse towards end (4 − 6)

5 = Occasional chatting and short pauses (3 or less)

6 = Good attention, fading slightly towards the end (2 or fewer short pauses towards the end)

7 = Good attention throughout with occasional comments and no real pauses

8 = Good attention throughout

9 = Fast, accurate, highly focussed on the task

An attention score based on this scale was allocated to each epoch of both tasks by the experimenter during testing and averaged for each task.

7.2   Results

7.2.1   Learning on the SRT task

RT analyses. Results for two children were missing for SRT1 due to computer malfunction and one for SRT2 due to absence. Inaccurate trials, RTs for the first bigram in each block, and those over 5000 ms were removed and a moving criterion based on sample size (Selst & Jolicoeur, 1994) was used to remove outlying RTs. RTs for the first two trials in each block were excluded, since they could not be predicted. Remaining response times for each task were analysed using mixed effects models (Rabe-Hesketh & Skrondal, 2012) in Stata 13.0 to take account of variability across participants and trials. Sequence type (probable or improbable), epoch and the interaction between the two were entered as fixed effects and participants and trials as crossed random effects. Mean RTs for each sequence type (probable vs. improbable) and epoch are shown in Figure 7.1. RTs for improbable trials were slower than for probable trials in all epochs of both tasks.
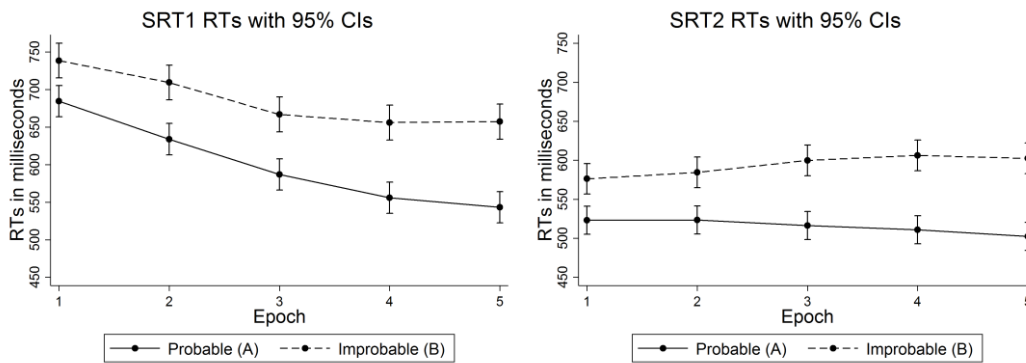
Figure 7.1 Graphs showing RTs per sequence and epoch for both SRT tasks, with 95% confidence intervals.

For both tasks RTs were slower on improbable trials (SRT1: marginal mean difference = 54.105 [95% CI 42.85, 65.36], $z = 9.42$, $p < .001$, SRT2: marginal mean difference = 53.170 [95% CI 44.46, 61.88], $z = 11.96$, $p < .001$). RTs became faster across epochs (SRT1 Epoch 1 - 5: unstandardized slope = -141.203 [95% CI -146.26, -136.15], $z = -54.76$, $p < .001$; SRT2 Epoch 1 - 5: unstandardized slope = -20.578 [95% CI -24.44, -16.71], $z = -10.43$, $p < .001$). Crucially, the interaction between sequence type and epoch was also significant on all blocks of SRT1 and the last three blocks of SRT2. These interactions confirm what can be seen in Figure 7.1: the difference in reaction time between the probable and improbable trials tends to increase from earlier to later epochs of both tasks. This difference in RT provides evidence that the children are learning the task structure of the probable trials.

Error Analyses. Once again the same error analyses that were used in the first study and the piloted extended task in adults were used to examine implicit learning in children on the extended SRT task. Figure 7.2 shows that children made proportionally more errors on the improbable sequence than on the probable sequence on both tasks (SRT1: unstandardized regression coefficient = .267, $z = 16.17$, $p < .001$, 95% CI [.23, .30]; SRT2: unstandardized regression coefficient = .701, $z = 43.60$, $p < .001$, 95% CI [.67, .22]) and errors increased at a greater rate on the improbable sequence than on the probable sequence over the course of the task (SRT1: unstandardized regression coefficient = 1.256, $z = 53.83$, $p < .001$, 95% CI [1.21, 1.30]; SRT2: unstandardized

regression coefficient = .850, $z = 37.39$, $p < .001$, 95% CI [.81, .89]). These results are highly suggestive of implicit learning of the probable sequence that develops over the course of the tasks.
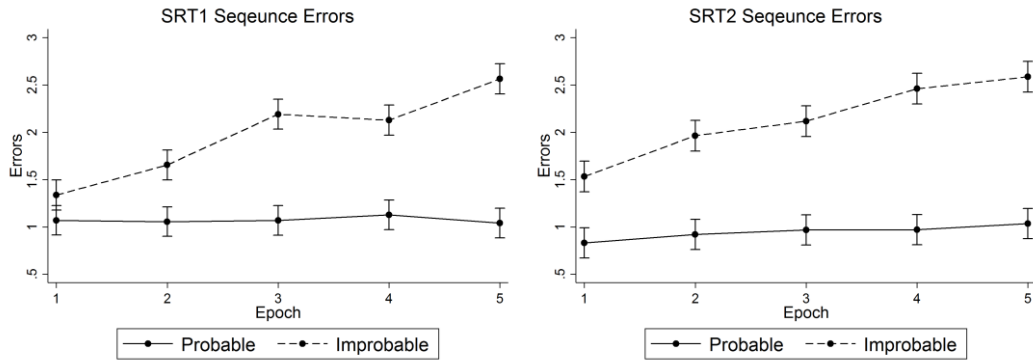
Figure 7.2 Error frequencies for the probabilistic non-verbal SRT tasks per sequence. Error bars are 95% confidence intervals.

Figure 7.3 shows that the frequency of anticipatory and random errors were not significantly different on both task attempts (SRT1: unstandardized regression coefficient = -.128, $z = -1.43$, $p = .15$, 95% CI [-.31, .05]; SRT2: unstandardized regression coefficient = -.260, $z = -1.76$, $p < .08$, 95% CI [-.55, .03]). However, the increase in anticipatory errors in the last epoch of SRT1 and over the last two blocks of SRT2 meant that the interaction between error type and epoch was significant (SRT1: unstandardized regression coefficient = -.35, $z = -2.68$, $p = .01$, 95% CI [-.61, -.09]; SRT2: unstandardized regression coefficient = -.43, $z = -2.23$, $p = .03$, 95% CI [-.82, -.05]. These results show that the pattern of errors on the SRT tasks reflected implicit learning of the probable sequence.
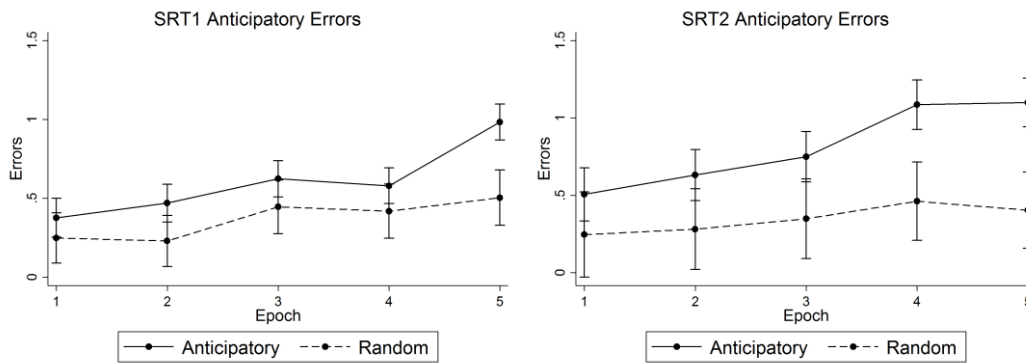
Figure 7.3 Error frequencies for anticipatory and random errors on the improbable sequence of the non-verbal SRT tasks. Error bars are 95% confidence intervals.

7.2.2   Task measures, reliability and correlations

Implicit learning scores for each child were derived from performance on the two serial reaction time tasks. The principal measure chosen to represent this learning was the overall proportional mean difference in RT between the probable and improbable sequences. Children's baseline speed of response was further controlled for by dividing this learning score by their overall response speed (improbable sequence RT − proportional probable sequence RT) / ((improbable sequence RT + proportional probable sequence RT)/2)). A positive score, therefore, reflected implicit learning (108 children had positive scores on the first task and 110 on the second task). Split-half reliability for the procedural learning RT score was moderate (SRT1 r = .51; SRT2 r = .62) though test-retest reliability was much lower (r = .26).

Implicit learning scores derived from error frequencies were also calculated. Test-retest for these was a little better than in the first study using the 500 trial SRT task (Sequence Errors test-retest reliability r = .38; Error Type test-retest reliability r = .62). However, once again the apparent reliability of the Error Type measure on the improbable sequence was deceptive. A substantial number of participants did not make any random errors on the improbable sequence (SRT1 n = 26; SRT2 n = 41) and their resulting derived measure was simply a measure of anticipatory errors as a result. Additionally, once baseline error frequency was controlled, reliability was very poor

for both measures (Sequence Errors test-retest reliability $r = -.07$; Error Type test-retest reliability $r = .12$). As a result neither measure was considered further.

Means, standard deviations and reliabilities for all measures are given in Table 7.1 and correlations between all measures in Table 7.2.

Table 7.1 Performance on attainment and memory measures.

	N	Mean	SD	Reliability
Age in months	112	92.11	5.71	-
TROG-2 (Total correct)	112	71.58	6.89	.88 [s]
TOWRE-2 Words	111	61.12	11.44	.90 [r]
TOWRE-2 Nonwords	110	36.75	10.71	.90 [r]
Arithmetic composite	111	53.72	25.62	.97 [r]
Addition	111	17.62	8.14	.92 [r]
Addition plus carry	112	8.62	5.31	.89 [r]
Subtraction	112	12.14	5.76	.88 [r]
Subtraction plus carry	112	6.02	3.88	.85 [r]
Multiplication	112	9.43	5.82	.93 [r]
WASI	112	18.12	5.38	.94 [s]
Word Lists (WL)				.84 [s]
Learning	112	32.44	6.03	
Delay	111	6.09	1.94	
Consolidation	111	4.29	1.78	
SRT1 RT Difference Score	110	.13	07	.51 [s] /.26 [r]
SRT2 RT Difference Score	111	.14	.08	.62 [s] /.26 [r]
SRT1 Sequence error score	110	.17	.17	-.07 [r]
SRT2 Sequence error score	111	.16	.30	-.07 [r]
SRT1 Error type score	110	.73	.32	.12 [r]
SRT2 Error type score	111	.81	.32	.12 [r]
SRT1 Attention Rating	112	7.25	1.29	.75 [r]
SRT2 Attention Rating	111	7.23	1.45	.75 [r]

[s] = split-half reliability / [r] = test-retest reliability; All procedural learning measures control for baseline performance.

Table 7.2 Correlation matrix for all measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Age (months)																	
2. TROG-2	.31**																
3. Towre Words	.18	.45*															
4. Towre Nonwords	.19*	.35*	.83**														
5. Arithmetic composite	.44**	.42**	.50**	.55**													
6. Addition	.42**	.40**	.50**	.52**	.92**												
7. Addition with carry	.33**	.32**	.38**	.43**	.89**	.78**											
8. Subtraction	.36**	.38**	.42**	.48**	.93**	.78**	.81**										
9. Subtraction with carry	.25**	.36**	.50**	.52**	.85**	.69**	.73**	.78**									
10. Multiplication	.50**	.37**	.40**	.46**	.83**	.68**	.62**	.72**	.63**								
11. Non-verbal IQ	.35**	.58**	.36**	.36**	.43**	.37**	.29**	.38**	.29**	.52**							
12. WL Learning	.07	.47**	.37**	.32**	.26**	.19*	.19*	.23*	.29**	.29**	.17						
13. WL Delay	-.07	.34**	.24*	.19	.15	.12	.06	.13	.17	.22*	.10	.76**					
14. WL Consolidation	.03	.30**	.24**	.21*	.19*	.14	.10	.17	.22*	.25**	.14	.63**	.68**				
15. Attention SRT1	.17	.39**	.37**	.33**	.29**	.26**	.31**	.23*	.32**	.19*	.32**	.18	.10	.04			
16. Attention SRT2	.17	.37**	.44**	.37**	.29**	.27**	.33**	.28**	.31**	.10	.28**	.16	.05	.12	.75**		
17. SRT1 RT	.09	.05	.07	.05	-.01	.04	-.03	-.05	.01	-.05	.06	-.05	.02	-.03	.18	.26**	
18. SRT2 RT	.15	.32**	.31**	.22*	.25**	.21	.23*	.27**	.28**	.15	.28**	.20*	.13	.10	.19*	.36**	.26**

*p < .05; **p < .01

235

7.2.3 Predicting individual differences in attainment from measures of procedural and declarative memory

The principal objective of the study was to assess the relationships between measures of basic cognitive skills (procedural learning [SRT RT]), attention during the SRT task, and declarative memory (free recall) on the one hand and measures of attainment (Reading [TOWRE word reading, TOWRE nonword reading]) Arithmetic, and Grammatical skills (Test of Reception of Grammar) on the other. For this purpose a latent variable path model was estimated and is shown in Figure 7.4. Modelling was conducted in Mplus 7.31 (Muthén & Muthén, 1998-2016) with the very small amount of missing data being handled with Full Information Maximum Likelihood estimation.

The starting point was a model in which each of the predictors (procedural learning in the SRT task, attention in the SRT task, and declarative memory [verbal free recall]) predicted each of the measures of attainment (reading, grammar, and arithmetic). Nonsignificant relationships between the predictors and measures of attainment were dropped iteratively and changes in χ^2 were used to check that there was no significant loss of model fit as a result of omitting each of the nonsignificant paths. The final simplified model is shown in Figure 7.4; the parameter estimates are the standardized regression coefficients and correlations. The model yields a good fit to the data, $\chi^2(79)$ = 94.526, p = .112; RMSEA = . 042 [90% CI . 0.000 0.071]; CFI = .98; TLI = .98.
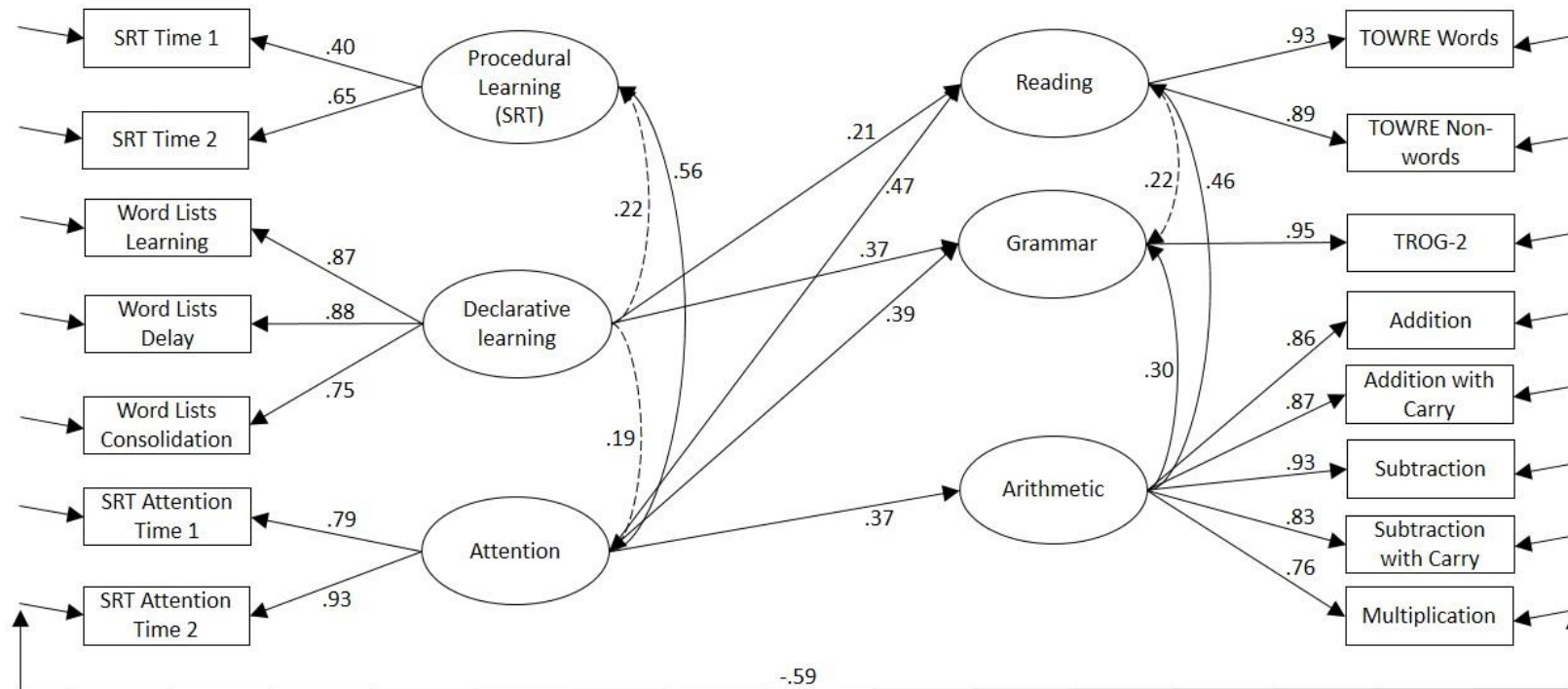
Figure 7.4 Path model of relationships between constructs. Single headed arrows represent regression paths or factor loadings. Twin headed arrows represent correlations between variables. Standardized regression coefficients and correlations are shown.

$(\chi^2 (79) = 94.526, p = . 112; RMSEA = . 042[90\% CI . 0.000\ 0.071]; CFI = .98; TLI = .98)$

237

One notable feature of this model is that the level of procedural learning on the serial reaction time task is quite strongly correlated ($r = .56$) with ratings of children's attention while performing the task. The other critical features of this model are the patterns of prediction from the serial reaction time task, attention in this task and declarative learning to measures of attainment (reading, grammatical skills and arithmetic). Procedural learning as measured by the serial reaction time task is not a significant predictor of any of these measures. In contrast, attention during the serial reaction time task is a predictor of all these outcomes, and in addition declarative memory is a predictor of both reading and grammatical skills.

The significant relationship between rated attention during the SRT task and performance on this task (and on all measures of attainment) is a novel and unexpected finding. Attention ratings were allocated during testing, prior to any derivation of procedural learning scores from the tasks, so ratings could not have been biased by any knowledge of the extent of each child's degree of procedural learning. The pattern of results here suggests that procedural learning in the SRT task is not itself a good predictor of variations in reading, grammatical, or arithmetic skills. However, the SRT task is highly attentionally demanding, and the extent to which children can maintain attention during the task is a relatively powerful predictor of variations in reading, grammatical, and arithmetic skills. This pattern is clarified by inspecting the pattern of correlations between the latent variables in the model (see Table 7.3).

The most powerful correlate of all three measures of attainment is rated attention during the SRT task. In comparison the learning score on the SRT task is a weaker correlate of all outcomes. Finally, declarative memory is a moderate correlate of grammatical skills and a weaker correlate of reading.

Table 7.3 Correlations between the latent variables in model in Figure 7.4

	1	2	3	4	5
1. Procedural Learning (SRT)					
2. Declarative Learning (CMS - WL)	.22				
3. Attention	.56	.19			
4. Reading (TOWRE)	.31	.29	.51		
5. Grammar (TROG-2)	.30	.44	.46	.45	
6. Arithmetic (TOBANS)	.21	.07	.37	.54	.39

7.3   Discussion

This study used latent variable path analysis to explore the extent to which procedural learning (as assessed by a serial reaction time task), declarative learning and sustained attention during the serial reaction time task predicted variations in attainment on measures of reading, grammar and arithmetic. Latent variable models allow us to assess the relationship between constructs without the confounding influence of measurement error (McCoach, Black, & O'Connell, 2007).

The model showed that declarative learning was a moderate predictor of attainment and was not significantly associated with variations in attention on the serial reaction time task. This finding is in line with many other studies that have found a robust relationship between measures of verbal declarative memory and reading ability (Melby-Lervåg, Lyster, & Hulme, 2012).

The findings for procedural learning are more surprising. Procedural learning (as assessed by a serial reaction time task) was strongly related to ratings of children's attention while performing the task ($r = .56$), and measured attention on this task was highly predictive of both procedural learning and variations in attainment. Conversely, once such attentional effects were controlled, there was no relationship between procedural learning on the serial reaction time task and measures of attainment. The model shows that it is a child's ability to attend to a lengthy and repetitive

computerized task that is an important predictor of attainment, not their procedural learning ability.

Few studies have examined the impact of attentional variability on procedural learning in language disordered children, although some have included measures of attention in their testing battery, in order to control for potential group confounds. Sengottuvel and Rao (2013b; 2014) measured attention in children with language impairment and typically developing children performing an adapted serial reaction time task, having found a correlation between attention and procedural learning in children in a previous study (Sengottuvel & Rao, 2013a). However, as they found similar levels of attention in both groups, they did not go on to examine whether a relationship between attentional scores, procedural learning and language ability existed.

A study by Staels and Van den Broeck (2017) used a deterministic serial reaction time task in groups of dyslexic and typically developing children. Although the principal measures of implicit learning on this deterministic task (increased RTs on introduction of a block of random trials and the rebound effect on returning to the sequence) were equivalent across groups, they found that a slight deficit in the initial learning of the sequence in the dyslexic group disappeared after controlling for attentional functioning, using parent and teacher ratings on an ADHD scale. This finding is particularly interesting as it is the first extreme group design to suggest that SRT implicit learning performance in children with dyslexia may be differentially impaired by attentional functioning.

Only one experiment has related attentional functioning to procedural learning and language ability across the sample. Waber et al. (2003) examined serial reaction time performance in 422 children with a wide range of reading ability. Parental ratings of attention predicted error rates on the serial reaction time task, but not response times. These studies suggest that measures of attention are useful additions to testing batteries examining procedural learning.

The current results might be seen as lending support to these findings; though the within-task measure of attention in this study is very different from measures of global parent or teacher rated attention used in previous studies. In addition the current study demonstrated associations between attention on the task and the principal measure of implicit learning, as indexed by differences in reaction time to probable vs. improbable sequences. Furthermore, the study shows that attention during the serial reaction time task accounts entirely for the relationship between procedural learning performance and measures of attainment (reading, grammar and arithmetic). A possible weakness of the current study is that it did not also include a standardised ADHD scale, such as those used above. It could be speculated that ratings for ADHD and the ratings of attentional functioning on the serial reaction time task used here would be highly correlated, but future research could explore whether including both measures sheds additional light on the relationship between procedural learning, attention and attainment.

An interesting question is whether attention would still account for the relationship between procedural learning and attainment in a procedural sequence learning task better able to maintain the focussed attention of children throughout the task? The 1000 trial SRT task used in the current study is highly demanding of sustained attention. Some previous studies have highlighted the importance of keeping motivation high and avoiding fatigue in children during monotonous or cognitively demanding procedural learning tasks. Most have responded by introducing abbreviated procedural learning tasks (e.g., Perlant & Largy, 2011; Stoodley, Ray, Jack, & Stein, 2008). However, generally brief tasks are less likely to yield reliable measures than longer tasks (Nunally & Bernstein, 1994), and the shorter 500 trial serial reaction time task in Chapter 5 in particular was found to be unreliable in children.

One potential consequence of extending a procedural learning task to increase reliability is that participants may start to learn it declaratively, rendering the task no longer a valid or pure measure of procedural learning. However this possibility is inconsistent with the pattern of results obtained here, in which serial reaction time learning scores and measures of declarative learning had different relationships to

241

attainment (the former mediated by variations in attention, the latter not). Moreover, by using probabilistic rather than deterministic regularities between target locations, the serial reaction time task employed here was designed specifically to render the structure hard to learn declaratively (Schvaneveldt & Gomez, 1998; Song, Howard, & Howard, 2007).

In summary, this study found that verbal declarative memory skill predicted reading, grammar and arithmetic attainment. However, the relationship between procedural learning (as assessed by a serial reaction time task) and attainment was entirely explained by rated attention during the task. Previous research has documented the frequent comorbidity between language disorder and attention deficits (Carroll, Maughan, Goodman, & Meltzer, 2005; Finneran, Francis, & Leonard, 2009; Laasonen et al., 2014). On the basis of the present findings, it can be conjectured that the poorer performance of such groups on procedural learning tasks may well reflect poorer attentional functioning during testing, rather than any separable procedural learning impairment. Furthermore, poor attention evident during the task is likely not to be an isolated effect and ongoing difficulties with sustaining attention will likely have knock-on effects on attainment. We recommend that future studies of procedural learning and language disorders include measures of attention and use procedural learning tasks specifically designed both to have good reliability and to address the inevitable attentional variability of participants.

Chapter 8   Study 4: Probabilistic categorization in children

8.1   Introduction

The weather prediction task (WPT: Knowlton, Squire, & Gluck, 1994) examines implicit probabilistic perceptual category learning, such as may be evident in early language learning as children detect statistical regularities in a continuous speech stream and subsequently map those sounds to meanings (Evans et al., 2009; Saffran et al., 2003). The learning of such probabilistic relationships is also crucially involved in the mastery of the rules of grammar (Ullman et al., 1997) and similar processes are at work when learning to read, where the co-occurrence of letters can serve as probabilistic cues for those that follow (Arciuli & Simpson, 2012).

As documented in Chapters 3 and 4, the weather prediction task has been used in a number of studies to investigate the claims of the procedural deficit hypothesis (Gabay et al., 2015; Kemény & Lukács, 2010; Lee et al., 2016; Lee & Tomblin, 2015; Lukacs & Kemeny, 2014; Mayor-Dubois et al., 2014). Results of previous research are mixed, but the bulk of evidence so far, along with the results of the meta-analysis in Chapter 4, suggests that performance on the task is impaired in both children and adults with developmental language disorder and dyslexia.

The weather prediction task has some advantages over other implicit learning tasks. Unlike in the artificial grammar or statistical learning tasks, the existence of hidden conditional probabilities in this task are not revealed to the participants, which means the task can be used twice to gauge test-retest reliability and it likewise follows that it does not rely on a testing phase that may draw on explicit memory processes to make legality judgements about rules or associations learned earlier.

The weather prediction task also holds an advantage over the serial reaction time and Hebb serial order learning tasks by using a unitary measure of above chance performance on the task as a reflection of implicit learning, rather than relying on a difference score. Given the problems associated with the use of difference scores that

has been demonstrated in previous chapters, the use of this measure may serve to improve the reliability of the task.

However, one important drawback of the weather prediction task as an index of implicit learning is the uncertainty over the extent to which implicit and explicit learning processes are involved in the task. It has been argued that participants are supposedly unaware that they are applying probabilistic reasoning strategies as they select outcomes in the task. In support of this, Gluck et al (2002) found self-reports of even simple strategies corresponded poorly with actual performance, suggesting that rules to guide performance were acquired in an unconscious, non-verbalisable way.

However, Knowlton et al's (1994) original experiments with amnesiac patients and controls suggested that while early performance is supported by implicit associational learning, declarative memory processes are increasingly brought to bear to aid performance as the task develops. Certainly, experimental manipulations of the weather prediction task aimed at reducing the contribution of explicit processing, such as the introduction of a secondary tone counting task (Foerde, Poldrack, & Knowlton, 2007) or reducing the amount of time available to process feedback (Price, 2009), result in lower accuracy on the task.

The current study investigates whether performance on the task is predictive of language-related attainment in a large unselected sample of children. It also explores the extent to which performance on the task may involve declarative learning. Consistent with the overall aims of the thesis, it also takes care to examine the reliability of the task.

8.2 Method

8.2.1 Participants

The same participants recruited into the study in Chapter 7 were tested on a weather prediction task in small groups on two occasions a week apart. Means and standard

deviations for attainment and declarative memory measures were, therefore, the same as those in Chapter 7 (see Table 7.1).

8.2.2 Task and testing procedures

In the weather prediction task participants are shown arrangements of stimuli (cue arrangements) and are asked to classify these arrangements into one of two possible outcomes. Cue arrangements include 1 to 3 stimuli out of a total possible set of 4. The task has a probabilistic structure, with each stimulus given a fixed probability of a certain outcome. The probability of an outcome on any given trial is, therefore, based on the combined probability of the cue arrangement that is displayed. Overall the two outcomes occur with equal frequency. Participants are not aware of the probabilistic nature of the task. After each trial feedback based on the cue-outcome probabilities informs them whether their choice was correct or incorrect. A trial is scored correct if it accords with the conditional probabilities of the cue arrangement shown, regardless of "correct / incorrect" feedback to the participant. This means the percentage score allocated to each participant reflects how well they have learned the probabilistic associations between cue arrangements and outcomes.

The cue-outcome probabilities associated with outcome 1 in the original weather prediction task were .75, .57, .43, and .25 (Knowlton et al., 1994). However, a study with children with developmental language disorder and typically developing controls found no learning on a task using these weights (Kemény & Lukács, 2010). A weather prediction task was, therefore, designed that used the identical probabilistic structure to the task in Gluck et al. (2002, Experiment 2). The cue-outcome probabilities associated with outcome 1 in this version of the task (0.2, 0.4, 0.6, and 0.8) were such that responding with the most likely category for each pattern would result in the correct prediction in 83% of trials, as opposed to the 76% in Knowlton et al's (1994) original task. This version of the task was selected in order to maximize the number of participants who perform above chance. See Table 8.1 for the outcome frequencies of each arrangement.

Table 8.1 Cue arrangements and outcome frequencies of the weather prediction task

Pattern	Cues present	Park	Pool	Total
A	0001	17	2	19
B	0010	7	2	9
C	0011	24	2	26
D	0100	2	7	9
E	0101	10	2	12
F	0110	3	3	6
G	0111	17	2	19
H	1000	2	17	19
I	1001	3	3	6
J	1010	2	10	12
K	1011	5	4	9
L	1100	2	24	26
M	1101	4	5	9
N	1110	2	17	19
TOTAL		100	100	200

8.2.2.1 Stimuli

The task used verbal nameable pictures to investigate the link between verbal probabilistic category learning and language-related ability and was given a theme to appeal to children. The task introduced the children to four minion friends who each attended a different class at Minion school (see Figure 8.1). Each minion held a placard emblazoned with their class emblem. These placards were the verbal nameable pictures associated with the probabilistic outcomes. Participants were informed that each break time the minions decided whether to visit the park or the pool after school. On each trial the children were asked to predict which of these two alternatives the minions would pick, depending on which minions appeared in the playground at break time.

Two versions of the task were developed that were identical in all but the four nameable pictures used as stimuli (WPT1: Hat, Leaf, Book, Cup; WPT2: Ball, Car,

Fork, Shoe). The tasks each contained 200 trials and the trial order was pseudo-random, such that no repeats of cue arrangements on consecutive trials. The left to right order of the individual stimuli within each prescribed cue arrangement was randomized, so that individual stimulus locations within each arrangement could not serve as an additional cue. The testing lag between task attempts was one week.

Figure 8.1 Two screenshots from the weather prediction task.

8.2.2.2 *Group administration Procedure*

One to one testing on implicit learning tasks is a lengthy process. For this reason a process was trialled that administered tasks to children in small groups. The task was hosted on an external server, which enabled groups of 8 to 12 children at a time to dial in to the task from school Information and Communication Technology (ICT) suites, using school computers. Group administration of the weather prediction task, made it possible to test all participating children in any given class within a single 90 minute session.

At the start of every group session each participating child was allocated to a computer and was asked to click a link on a web-page pre-loaded by the experimenter in order to enter the experiment. The children were then asked to enter a simple unique code as an identifier. The experimenter then explained the task to the children as a group, with children asked to follow the written version of these instructions on the screens in front of them, clicking the "Next" button in time with the rest of the group,

as prompted by the experimenter until the instructions were completed. The children were then told to click the "Start button" on the final instruction screen to begin the experiment individually. The experiment itself was self-paced and children returned to class on completion, in order not to disturb those still working on the task. The sessions were closely monitored by the experimenter to keep communication between the children to a minimum during the task.

8.2.2.3 Test of task knowledge

At the end of the second task children were asked four questions to gauge the extent of their explicit knowledge of single cue-outcome associations. For each of the four cues, they were asked to judge how likely the Minion was to want to go to the park or the pool. To respond the children had to click on 1 of 5 options displayed horizontally from left to right on the screen. Each option was associated with a score on a Lickert scale (1 to 5)

Does [Minion Cue] love the park or the pool?

Press this button if you think that [Minion cue]…

…always wants to go to the park (1)

…prefers the park, but not all the time (2)

…likes them both the same (3)

…prefers the pool, but not all the time (4)

…always wants to go to the pool (5)

8.3   Results

The 200 trials on the weather prediction task were split into four blocks of 50 trials. The percentage of optimal responses were measured for each block. Procedural learning was defined as above chance performance that significantly improved across the 4 blocks of the tasks. Two of the arrangements (F and I) had an average predictive value of 50% and were scored as correct, regardless of the answer given by the participant. This meant that performance above chance level would be reflected by an

overall percentage score of over 56% rather than 50%, since these two cue arrangements appeared on 6% of the trials (12 times across the 200 trials).

Means and standard deviations by block and task total for both tasks are in Table 8.2. The percentages of optimal responses were higher on the second compared to the first task. Performance on the first block of WPT1 was at chance, but was slightly greater than chance on the remaining three blocks. The mean percentage of optimal responses was considerably higher on the first block of WPT2, suggesting that a strategic approach to the task may have developed by the second attempt, since although the cues themselves were different on the second task the probability weights associated with the cues were identical to the first task. Performance continued to improve over the course of the second task.

Table 8.2 Means and standard deviations by block and task total for both task attempts.

	WPT1 ($n = 107$)		WPT2 ($n = 103$)	
	Percentage of Optimal Responses	SD	Percentage of Optimal Responses	SD
Block 1	56.6	5.98	60.37	9.31
Block 2	58.21	8.83	62.56	11.29
Block 3	58.22	8.10	62.37	11.49
Block 4	58.36	8.48	63.34	11.48
Total	57.85	6.26	62.16	9.03

Once again, mixed effects models (Rabe-Hesketh & Skrondal, 2012) in Stata (13.0) were used to analyse the percentage of optimal responses for each task separately. Block was entered as a fixed effect and item and participants as crossed random effects to take account of participant variability.

For the first task Block was a significant predictor of percentage of optimum responses (Block 1 – 4 unstandardised regression coefficient = 1.757, $z = 2.30$, $p = .02$, 95% CI [.26, 3.25]. The result was similar on the second task (Block 1 – 4

unstandardised regression coefficient = 2.97, $z = 3.01$, $p < .001$, 95% CI [1.04, 4.91], such that both tasks showed significant improvement in performance over the course of the four blocks of trials (see Figure 8.2).
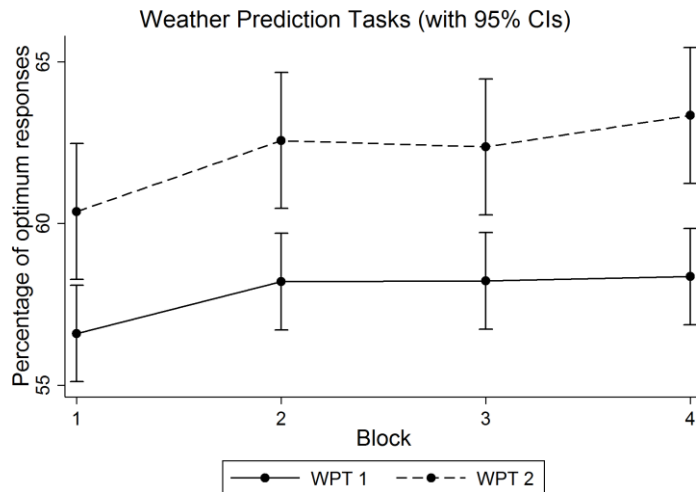
Figure 8.2 Percentage of optimum responses on both attempts at the weather prediction task with 95% confidence intervals.

8.3.1   Task measures and reliability

The implicit learning measure used for each child was the mean percentage of optimal responses across all four blocks of each task. Test-retest reliability of this measure was moderate ($r = .46$). Test-retest reliability of the last block of the task was slightly higher ($r = .50$).

8.3.2   Correlations

Correlations between the weather prediction tasks and the attainment and declarative memory variables from Chapter 7 for the whole sample were uniformly low (see Table 8.3). Low but significant correlations were noted between WPT2 and NVIQ ($r = .21$), the TOWRE word reading test ($r = .26$) and long-term verbal declarative memory (WL-C) on the CMS word lists free recall task ($r = .21$), but these were no longer significant once Bonferroni corrections had been made. There were no

correlations between implicit learning on the serial reaction time task and performance on either of the weather prediction tasks.

However, it was noted during the task that a number of children appeared to be randomly pressing buttons throughout the tasks. Observations were, therefore, restricted to the 38 participants who displayed learning on both tasks (ie: those participants whose percentages of optimal responses were above 56 % on both tasks), in case the chance-level performance of children who did not engage with the tasks was masking any underlying relationships between variables (see second two columns of Table 8.3). There were no significant correlations for the first task, but the correlation between WPT2 and TOWRE word reading increased ($r = .36$), as did the correlation between WPT2 and long-term verbal declarative memory (WL-C) on the CMS word lists free recall task ($r = .53$). This relationship was even evident on the first block of the task (WL-C and WPT2 Block 1 $r = .36$). The correlation between WPT2 and the verbal short-term memory (WL-L) was also significant ($r = .32$). These results suggest that children who performed the weather prediction tasks well may have better than average verbal declarative memory and word reading skills, although once again the correlations were not significant after applying Bonferroni corrections for multiple comparisons.

Table 8.3 Correlations between weather prediction tasks and attainment and memory measures (from Chapter 7) by overall sample and for the subgroup of participants who achieved above chance performance on both attempts at the weather prediction tasks.

	Overall Sample		Subgroup performing above chance (*n*= 38)	
	WPT1	WPT2	WPT1	WPT2
Age in months	-.06	.01	-.14	-.08
TROG-2 (Total correct)	.11	.15	.09	.16
TOWRE-2 Words	.09	.26*	.15	.36*
TOWRE-2 Nonwords	-.03	.17	.00	.24
Arithmetic composite	.05	.19	.04	.05
WASI	.17	.21*	.11	.21
Word Lists (WL)				
Learning	.05	.14	.19	.32*
Delay	-.02	.12	.15	.28
Consolidation	-.05	.21*	.12	.53*
SRT1 RT Difference Score	-.04	-.13	-.23	.02
SRT2 RT Difference Score	-.00	.08	-.09	.10

*$p < .05$

8.3.3 Task performance and declarative learning

To explore the question of whether improved performance towards the end of the task reflected declarative learning, rather than implicit learning, the children's judgements on the 5-choice explicit tests on WPT2 were examined. Cue 1 was strongly associated with "The Pool", Cue 4 was strongly associated with "The Park. Cues 2 and 3 had weak associations with the outcomes. This meant that the appropriate response for Cue 1 was a high score on the Lickert scale (4 or 5) and the appropriate response for Cue 4 was a low score (1 or 2). Cue 2 should also tend towards higher scores and Cue 3 lower ones.

The explicit judgement scores for each cue were entered into a multiple regression with the percentage of optimal responses on the last block of the WPT2 task as the dependent variable, since by then any declarative learning should be well established.

A significant model emerged, $F(4,98) = 5.68$, $p < .001$, which explained 15.51% of the variance in performance on the last block of the task (Adjusted $R^2 = .155$). Table 8.4 gives the regression coefficients and other information for the four predictor variables. As hypothesized, Cue 1 and Cue 4, which were both associated with outcomes with a high probability, were significant predictors of performance. Cue 2 and 3 were not significant (although their coefficients were in the right direction). This suggests that children used explicit knowledge of the two predictors with strong associations with outcome to guide their performance on the task, such that seeing Cue 1 within a cue arrangement would lead them to select "The Pool" and seeing Cue 4 within a cue arrangement would lead them to select "The Park".

Table 8.4 Regression coefficients for each cue for the model predicting performance on last block of WPT2.

Predictor variable	B	SE	β	p	95% CI
Cue 1	3.29	1.01	.32	<.01	[1.28, 5.30]
Cue 2	.73	.86	.08	.40	[-.97, 2.43]
Cue 3	-.03	.84	-.00	.97	[-1.70, 1.64]
Cue 4	-2.00	.78	-.24	.01	[-3.55, -.46]

Coefficients for Cues 3 and 4 are negative, as they were associated with lower values on the Lickert scale.

Similar regression models for the other blocks of the task showed that explicit knowledge of the cue-outcome associations did not predict performance on the first 50 trials of the task. On the second and third blocks such knowledge was already a significant predictor of performance, however, explaining 12.69% and 13.23% respectively of the variance in performance (2nd Block: $F(4.98) = 4.71$, $p = .002$; 3rd Block: $F(4.98) = 4.89$, $p = .001$), although only explicit knowledge of Cue 1 was significant in both cases (Cue 1 for Block 2 and Block 3: beta = .30, $p = .003$). The regression model for the task as a whole was almost identical to the one for the last block (WPT2 Total: $F(4.98) = 5.51$, $p < .001$).

8.4   Discussion

Performance on the first attempt at the weather prediction task was only just above chance, but this improved on the second attempt, where performance with new stimuli was already above chance in the first block of 50 trials and continued to improve across the course of the task. At first glance these results could be taken to indicate the presence of implicit learning on the second task. However, explicit judgements at the end of the second task showed that knowledge of single cue-outcome associations guided performance. Correlations between overall task performance and other variables were low, but relationships between performance on the second task and both verbal declarative memory and word reading fluency were evident in the subset of participants who displayed learning on the tasks, such that better prediction of probabilistic associations between cue and outcome was related to more proficient verbal memory and reading performance. The tasks were also shown to be moderately reliable.

8.4.1   The reliability of the task

The reliability of the weather prediction task in this study was not high, but it was considerably better than the other implicit learning tasks used in this thesis. Why might this be the case? Firstly, the implicit learning score for the weather prediction task avoided the use of difference scores, relying on the number of optimal responses in a dichotomous forced choice format instead. Difference scores are inherently unreliable (Lord, 1958; Overall & Woodward, 1975), as has been discussed in Study 1 (see Chapter 7) and this may have played a role in the better reliability of the task.

The task also required participants to make judgements that were associated with varying levels of probability. Siegelman, Bogaerts, & Frost (2017) report that tasks that include trials of varying levels of difficulty are likely to be more reliable when measuring complex, multi-faceted constructs, such as implicit learning, than tasks that constrain trials to a single type and level of difficulty. The weather prediction task performs well against this criterion.

However, Siegelman et al. (2017) also cautioned that task reliability will be adversely affected when the proportion of participants that perform better than chance is low, since chance level performance adds nothing but noise to the data. In this study less than 40% of the participants performed above chance on both occasions of taking the task. This can be compared unfavourably to the 80% of healthy adult participants that performed above a criterion of 65% on a task with identical cue outcome probabilities in Gluck et al. (2002). One obvious difference between Gluck et al's study and the current study is the age of participants. Perhaps the reliability of the current weather prediction task is no better than moderate on account of the young age of the children taking the task. This conjecture is consistent with the reasonable level of reliability found in the pilot serial reaction time task in adults and the poorer reliability in the subsequent study in children reported in Chapter 7.

This study also trialled group administration of the weather prediction task in an effort to reduce testing time. This has clear benefits for researchers, but also benefits schools, as testing of large numbers of children can be accomplished in a short time with minimum disruption to teaching. For budgetary reasons, children in this study accessed the task using the schools' own ICT equipment. This had the advantage that children were familiar with the computerized equipment. However, ICT provision differed from school to school. Some provided desktops and mice with a dedicated ICT suite. Others provided laptops with integrated trackpads and no dedicated space for testing. While the weather prediction task does not rely on response times which would have been confounded by such differences, this is still less than ideal. In spite of these differences, group administration of the tasks worked well and the children were minimally distracted by the presence of others. However, it is possible that this mode of administration may have had a further negative impact on task reliability. Not least, giving the task to children in small groups, rather than individually, may have influenced how assiduously some of the children applied themselves to the task and this may have been a factor in the chance level performance in some cases.

8.4.2 The extent of explicit learning on the task

This study was also concerned with the extent to which performance on the weather prediction task reflected declarative rather than implicit memory processes. If successful task performance is shown to utilize declarative learning, then ascribing the impaired performance on the task of language-disordered groups to deficits in procedural learning may be premature. Certainly, Gabay et al. (2015) found evidence that impaired performance on the weather prediction task by dyslexic participants was due to impaired explicit learning of cue-outcome relationships, rather than deficits in procedural learning. This study used a simplified version of a method devised by Newell, Lagnado, & Shanks (2007) to probe the extent of participants' explicit task knowledge about single cue-outcome probabilities. Newell et al's method asked participants to estimate the probability that each cue predicts a particular outcome. The difference between this figure and the actual probability was then taken as an indication of task knowledge. This explicit testing method was considered too mathematically sophisticated for 7 year-old children, so a simpler version with a 5 point scale was designed. Regression analyses showed that the explicit judgements for the two probable cues predicted performance, while the judgements for the two weaker cues did not. These results indicate that explicit learning of single cue-outcome associations on the weather prediction task in children is limited to the cues with a strong association with a particular outcome, ie: the children focus attention on the highly predictive cues and respond based on their presence or absence in the stimuli arrangements. This conjecture is in line with similar findings for the artificial grammar learning task by Perruchet and Pacteau (1990) who demonstrated how implicit learning performance in such tasks could be explained by the explicit learning of simple bigrams, rather than by abstraction of complex rules. However, overall the number of children using this simple strategy successfully in the current study was small and generally performance on the task was not very good.

The study also found performance on the second weather prediction task related to measures of verbal declarative memory on the CMS Word Lists free recall task, such that children with superior declarative memory skills were better at the task. This

256

finding is consistent with the hypothesis that explicit memory processes aid performance on the task. This was particularly the case for the long-term verbal declarative memory measure (WL-C) and a significant correlation was even found with the first block of the second task. This may indicate that, in spite of the different stimuli set used on each task, the improved performance from the very beginning of the second task compared to the first one was linked to recall of statistical or deductive reasoning strategies that were honed on the first occasion of taking the task.

However, there are two specific characteristics of the weather prediction tasks used in this study that may make it more susceptible to declarative learning than the original version of the task. Firstly, the tasks in this study used verbal stimuli (nameable pictures) as cues, compared to the ostensibly non-verbal tarot cards or geometric shapes more typically used, and it is possible that this contributed to the development of declarative knowledge of cue-outcome relationships. Being able to apply verbal labels to the cues makes it possible to formulate verbalisable rule-based strategies, which in turn facilitates explicit learning processes, thereby improving categorization accuracy (Fotiadis & Protopapas, 2013; Price, 2009).

A second characteristic that can be linked to declarative learning is the probabilities assigned to the cues in the tasks. Gluck et al. (2002) tested adult participants on two versions of the weather prediction task. The first version used Knowlton's et al's (1994) original cue-outcome weights and resulted in 30% of participants reaching a performance criterion of 65% correct judgements. The second version used the probabilities in the current task and, as already noted, 80% of participants achieved the same criterion. Fotiadis and Protopapas (2013) linked this improvement in performance to declarative learning, with the stimuli in the latter version made more explicitly discriminable by virtue of their adjusted probabilities. The version of the task used in this study was selected with the aim of maximizing the number of children performing above chance, but it may be that improved performance on the task as a result of more discriminable probabilities necessarily involved declarative learning.

To conclude, the results of the current study suggest that the weather prediction task may not be a good task to investigate implicit learning processes, while also highlighting some particular difficulties in using this task with young participants. Although the reliability of the task was moderate, only a relatively small proportion of children performed above chance. In this subset of participants, declarative knowledge of single cue-outcome probabilities for the two more heavily weighted cues significantly predicted task performance in all but the first block of 50 trials, becoming a progressively stronger predictor as the task progressed. Finally, regardless of declarative or implicit labels, this study ultimately found little evidence that performance on the task was related to language skills. The relationship with word reading fluency was weak. Even in the subset of participants who performed above chance, correlations were not significant once adjustments for multiple comparisons were made.

Chapter 9   Study 5: Procedural and declarative learning in dyslexia

9.1   Introduction

This study investigated whether impaired procedural learning is a potential cause of reading difficulties in dyslexia. The sample comprised a group of children with dyslexia and a younger typically developing group matched for reading ability. Measures of procedural learning, declarative learning and attention were taken for both groups. This design addresses a possible criticism of the correlational design used in earlier chapters. The studies in the previous chapters have recruited children unselected for ability in order to investigate the procedural deficit hypothesis without the methodological limitations of group designs. However, if procedural memory deficits are linked to language pathology only and not to language ability more broadly, it is possible that correlational designs do not include enough children with severe reading and/or language disorder in order to show evidence of any significant relationship. In support of this view, Stoodley et al. (2008) found impaired procedural learning performance in a group of children with dyslexia, but unimpaired performance of both typically developing controls, as well as a group of "garden variety poor readers", whose reading impairment was linked to general underachievement.

An additional rationale for this study is that the majority of studies investigating the procedural deficit hypothesis have used an age-matched control group. An advantage of using a reading ability-matched design instead is that it allows an inference to be made about the direction of causation, if any deficit in procedural learning is found. This is because any difference in procedural learning cannot be the result of differences in reading ability. This study, therefore, investigates whether impaired procedural learning is a potential cause of reading difficulties in dyslexia, using a sample of children with dyslexia and a younger typically developing group matched for reading ability.

In this study, the dyslexic sample was drawn from a group of children with a formal diagnosis of dyslexia attending specialist schools, while a comparison group that was twice as large was drawn from younger children in mainstream classrooms that had

been recruited for the study in Chapter 7 study. Maximising the number of typically developing controls should minimize the effect of individual differences in this group (Lukacs & Kemeny, 2014).

9.2 Method

A comparison of children with dyslexia and reading age-matched controls was used to evaluate differences in procedural and declarative learning. Ethical approval for the study was granted by the UCL Research Ethics committee.

9.2.1 Participants

Seventy two children aged between 7 and 11 years took part in the study. There were 24 children with dyslexic difficulties (mean age = 117.67 months; *SD* = 13.81) and 48 typically developing reading-ability matched controls (mean age = 91.67 months; *SD* = 5.76). The dyslexic group was recruited from specialist schools for children with dyslexia in both Surrey and North London. Data from 48 children who were part of the sample in Study 4 were used to form the reading age control group. The control group children were matched with the dyslexic group on reading fluency on the TOWRE sight word efficiency test.

9.2.2 Measures and procedure

The dyslexic children were tested individually on the following measures that have been used in previous chapters. Children were tested in two 25-minute sessions one week apart. Tasks were completed in a fixed order.

Reading ability was examined using the test of word and non-word reading efficiency (TOWRE-2: Torgesen, Wagner, & Rashotte, 1999); verbal declarative learning was assessed using the Word Lists subtest from the Children's Memory Scale (Cohen, 1997), including learning, delayed and consolidation scores; procedural learning was assessed using the 1000 trial version of the probabilistic serial reaction time task used in Study 4; and the 9-item SRT attention rating scale was used to assess

the level of attention the children paid to the serial reaction time task. An additional measure of attention (GoNoGo Task) was also included and is described below.

GoNoGo task. Children completed a computer task to provide a measure of behavioural inhibition. The first 30 trials of this task presented a cartoon bug which children were instructed to 'splat' as quickly as they could in order to establish their pre-potent response. There were variable lengths of inter-stimulus interval (300 ms, 600 ms, 900 ms) to ensure children waited until the target was presented before responding. The bug stimulus was presented for 800 ms, during which time children were able to make their response by pressing the spacebar on the computer keyboard. If children responded in less than 800 ms, "Splat!" appeared on the screen for 500 ms. If the button was not pressed in this time "Too slow!" appeared for 500 ms. Following these trials, a test phase presented 80 GoNoGo trials. Children were instructed to press the button to splat bugs, but to inhibit their response when a ladybird was presented. There were 60 presentations of the bug (Go trials) and 20 presentations of the ladybird (NoGo trials), with trials presented in a random order. The task lasted approximately 5 minutes. Stimuli were presented and responses recorded using E-Prime Software (version 2.0). The number of NoGo trials successfully inhibited was used as a measure of behavioural inhibition.

9.3   Results

Twenty four dyslexic children were tested and compared to a control group of 48 children matched for reading fluency on the TOWRE-2 word reading test that were taken from the sample of 112 children tested in Study 4 (Chapter 7). The means, standard deviations and reliabilities for all tasks are shown in Table 9.1 by group. As expected, in spite of being matched for word reading fluency (TOWRE-2 Word Reading Raw Score) the word and non-word reading standard scores of the dyslexic group were considerably lower than the typically developing group. Although the mean scores for the verbal declarative memory measures and for SRT rated attention were higher for the dyslexic group than the control group, further analyses showed these differences were not significant (ANOVA with group (2 levels) and declarative

memory (3 levels): $F(1, 70) = 2.03$, $p = .159$, $\eta_p^2 = .03$; t-test for SRT rated attention: $t(70) = -1.14$, $p = .26$, $d = -.28$). However, results showed that the two groups were not equated for performance on the GoNoGo task, with the dyslexic group demonstrating significantly lower levels of response inhibition, $t(69) = 3.13$, $p = .003$, $d = .78$.

Table 9.1 Means, standard deviations and task reliabilities by group, with group differences and effect sizes.

	Typically developing Group (*n* = 48)				Dyslexic Group (*n* = 24)				*t*	*d*	Reliability
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max			
Age in months	91.67	5.76	83	105	117.67	13.81	93	138	-11.28	-2.82	-
Gender (f/m)	29 / 19	-	-	-	17 / 7	-	-	-			-
Handedness	47 / 1	-	-	-	19 / 3	-	-	-			-
TOWRE-2											.90 [r]
Words (raw)	51.96	11.60	11	62	50.58	17.60	20	88	.40	.10	
Words (standard)	108.56	9.47	82	124	90.13	11.38	65	120	7.28	1.82	
Non-words (raw)	29.58	9.92	0	50	22.67	10.68	5	48	2.72	.68	
Non-words (standard)	112.67	10.76	79	135	92.00	8.98	73	113	8.10	2.02	
Word Lists											.84 [s]
Learning	30.96	6.20	16	45	33.38	5.75	21	40	-1.60	-.40	
Delay	5.60	1.93	1	9	6.17	1.79	3	9	-1.19	-.30	
Consolidation	4.08	1.92	0	8	4.21	1.67	1	7	-.27	-.07	
NoGo Accuracy	13.79	2.93	5	20	11.42	3.19	4	17	3.13	.78	-
SRT1 RT Difference	0.13	0.05	0.02	0.23	0.14	0.06			-.37	-.09	.26 [r]
SRT Attention	6.88	1.51	3	9	7.31	1.51	3	9	-1.14	-.28	-

9.3.1 Procedural Learning

The principal measure of procedural learning for serial reaction time tasks is the difference between the response times for trials that follow a predictable sequence and those that do not. For the RT analyses in this task, only correct responses were included and data was screened and prepared using the same procedures followed in previous chapters. See Table 9.2 for means and standard deviations for each group.

Table 9.2 Serial reaction time task RT means and standard deviations for dyslexic and control groups by sequence and epoch

Epoch	Dyslexic group ($n = 24$)		Control group ($n = 48$)	
	Probable	Improbable	Probable	Improbable
1	583.21 (150.13)	629.83 (152.42)	725.25 (177.89)	778.11 (211.96)
2	550.82 (117.32)	624.97 (139.72)	650.20 (134.96)	726.68 (163.21)
3	520.24 (95.07)	617.30 (148.69)	608.98 (128.17)	676.65 (130.23)
4	508.46 (101.85)	583.52 (113.40)	573.02 (107.04)	685.78 (171.87)
5	492.31 (97.89)	585.92 (126.94)	559.06 (109.98)	691.66 (174.47)

There was evidence of procedural learning on the serial reaction time task in both groups. As shown in Figure 9.1, RTs for the probable sequence were faster than for the improbable sequence across all blocks in both groups.
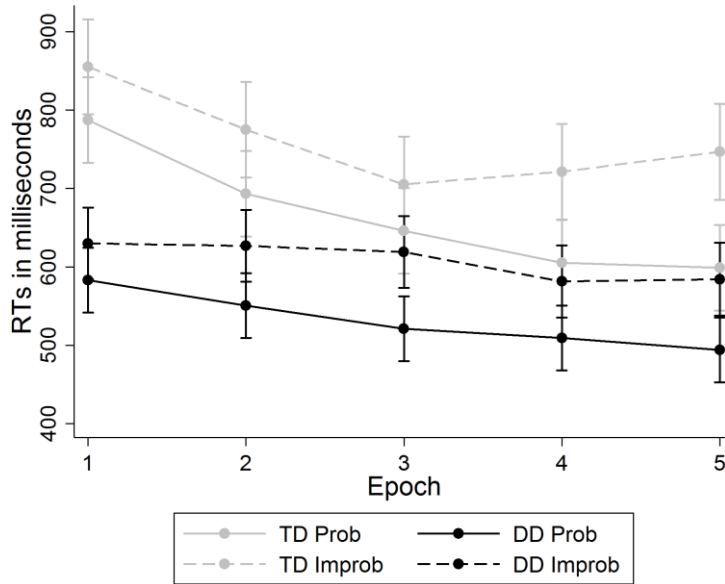
Figure 9.1 SRT task RTs for the probable and improbable sequences by dyslexic (DD) and typically developing control groups (TD), with 95% confidence intervals.

To investigate whether there was a difference between groups in the level of procedural learning on the serial reaction time task a mixed factorial Analysis of Variance (ANOVA) was performed with group (2 levels: control group vs. DD) as a between participants variable, and sequence (2 levels: probable vs. improbable) and epoch (5 levels: epochs 1-5) as within participant variables. This method of analysis was selected to facilitate comparison with the results of previous group design studies of procedural learning and dyslexia. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of epoch ($\chi^2(9)= 67.39$, $p < .001$) and for the interaction between sequence and epoch ($\chi^2(9)= 27.42$, $p = .001$). Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

Results showed that RTs on probable sequences were significantly faster than on improbable sequences across groups ($F (1, 70) = 280.58$, $p < .001$, partial eta squared $= .80$) (Probable sequence mean $= 592$ ms, $SD = 124.11$ ms; improbable sequence mean $= 677.29$ ms, $SD = 140.76$ ms). There was a main effect of epoch ($F (2.50, 174.69) = 20.39$, $p < .001$, partial eta squared $= .23$), indicating a significant decrease in RT across the epochs of the task. There was also a main effect of group ($F (1, 70)$

= 10.04, p = .002, partial eta squared = .13) such that children with dyslexia were significantly faster on the task compared to the control group, regardless of sequence (Dyslexic group mean = 569.66 ms, SD = 108.61 ms; Control group mean = 667.54 ms, SD = 130.28 ms). Critically, however, despite overall differences in RT the two groups did not differ in the level of procedural learning as indicated by the non-significant interaction between group and sequence (F (1, 70) = 1.27, p = .26, partial eta squared = .02) and group, sequence and epoch (F (3.30, 230.79) = 1.72, p = .16, partial eta squared = .02). Therefore, procedural learning appears to be equivalent in both groups.

9.3.2   Correlations by group

Sample sizes were small, but correlations for each group were examined to see if the pattern of relationships between the measures in the study differed by group. Correlations between the measures for each group were generally low (apart from the expected relationships for reading measures and for verbal declarative memory measures) and similar to one another (see Table 9.3 and Table 9.4 for correlations by group).

Table 9.3 Correlations between measures for the dyslexic group. Pairwise correlations are above the diagonal and partial correlations controlling for age are below the diagonal (*p < .05).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Age in months		.76*	.67*	.35	.29	.07	-.02	-.14	.49*
2. TOWRE-2Words	.76*		.82*	.11	.11	-.15	-.02	-.06	.48*
3. TOWRE-2 Non-words	.67*	.63*		.28	.14	.13	-.08	-.16	.48*
4. WL-L	.35	-.26	.06		.68*	.41*	.19	-.30	.14
5. WL-D	.29	-.18	-.08	.65*		.32	.09	-.16	-.17
6. WL-C	.07	-.31	.11	.41	.32		-.34	-.07	-.09
7. NoGo Acc	-.02	-.00	-.09	.21	.11	-.34		.01	.27
8. SRT RT Difference	-.14	.07	-.09	-.28	-.13	-.06	.01		.04
9. SRT Rated Attention	.49*	.19	.24	-.04	-.37	-.15	.33	.12	

Table 9.4 Correlations for the control group. Pairwise correlations are above the diagonal and partial correlations controlling for age are below the diagonal.

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Age in months		.19	.29*	.09	-.09	.08	.23	.07	.11
2. TOWRE-2Words	.19		.77*	.32*	.07	.23	.42*	.16	.36*
3. TOWRE-2 Non-words	.29*	.76*		.34*	.06	.22	.24	.12	.30*
4. WL-L	.09	.32*	.33*		.72*	.60*	.06	-.01	.13
5. WL-D	-.09	.09	.09	.73*		.67*	-.18	-.10	-.05
6. WL-C	.08	.22	.21	.60*	.69*		-.10	.03	-.10
7. NoGo Acc	.23	.40*	.19	.05	-.16	-.11		.02	.43*
8. SRT RT Difference	.07	.15	.10	-.02	-.09	.03	.01		.17
9. SRT Rated Attention	.11	.35*	.29	.12	-.05	-.11	.43*	.16	

The control group showed a low but significant correlation between short-term declarative verbal memory (WL-L) and word and non-word reading (r's = .32 and .34 respectively), which was absent in the dyslexic group. Both groups also displayed significant correlations between rated attention on the SRT task and word and non-word reading and these were larger for the dyslexic group (r's .48 and .48) = than the typically developing control group (r's = .36 and .30 respectively).

However, the mean age of the dyslexic group was approximately 2 years older than the typically developing group, and the age range far greater. Age in the dyslexic group was significantly related to word ($r = .76$) and non-word reading ($r = .67$), as well as to the level of attention paid to the SRT task ($r = .49$), so partial correlations were also estimated for both groups that controlled for the effect of age (see Table 9.3 and Table 9.4, which also list partial correlations for all measures for both groups). Once this was done the relationship between SRT rated attention and word and non-word reading in the dyslexic group was no longer significant. The difference between the correlation for word reading and short-term verbal declarative memory for the two groups increased (see Figure 9.2), only remaining positive in the control group (Dyslexic group $r = -.26$; Control group $r = .32^*$).
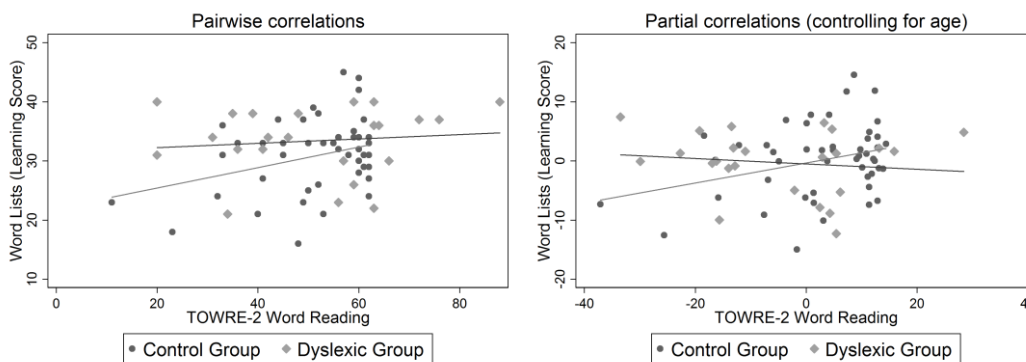

Figure 9.2 Left scatterplot showing relationship between word reading and short-term declarative memory for each group, with linear fit lines. Right plot shows same relationship, but controlling for age.

A moderate correlation between the two measures of attention (SRT rated attention and GoNoGo accuracy) in the control group ($r = .43$), suggested a degree of construct

validity for the SRT attention rating scale, as both tasks are broadly indexing aspects of attention, although this correlation was absent in the dyslexic group. Crucially, there was no relationship between procedural learning on the serial reaction time task and any other variables for either group, regardless of whether the correlations controlled for age or not.

9.4 Discussion

This study compared the performance of a group of children with dyslexia and a typically developing control group matched for reading ability on a range of tasks investigating procedural and declarative learning and attention. This design enabled comparison of children who differ in their ability to learn to read but who have developed equal levels of reading skill. Reading level designs like this can shed light on the causal direction of relationships between memory skills and reading that an age-matched comparison cannot do, since differences in task performance cannot be attributed to difference in reading ability. However, a drawback of this design is that while significant group differences might suggest that procedural learning plays a causal role in reading development, the absence of a significant difference between groups does not necessarily imply there is not a causal relationship (Goswami & Bryant, 1989).

9.4.1 Procedural learning

This study showed that children with dyslexia show comparable levels of procedural learning to that shown by the reading ability-matched control group of typically developing children. Furthermore, individual differences in procedural learning did not correlate with word or non-word reading in either group.

Examining performance of the two groups on the serial reaction time task in more detail does highlight some interesting differences, however. First, the dyslexic group were faster overall on the task than the control group. This can be attributed to the fact that they were older, but a second difference between the groups is more interesting. Although, the interaction between group and epoch was non-significant, Figure 9.1

269

shows that the slopes for both sequenced and random trials for dyslexic children in the first half of the task is flatter than for typically developing children. This might indicate poorer initial motor learning in the dyslexic group. Furthermore, while children with dyslexia show evidence of implicit learning, RTs to improbable trials continue to decrease in speed even at the end of the task. Given that children with dyslexia are showing clear evidence of implicit learning we should expect RTs on improbable trials to increase as they do in the control group or plateau. Decreasing RTs to the end of the task on the random sequence can only indicate continued motor learning right through to the late stages of the task. This pattern may indicate that extraneous factors, such as impaired coordination or motor learning (Ramus, Pidgeon, & Frith, 2003) influence the performance of children with dyslexia on serial reaction time tasks.

9.4.2 Declarative learning

More surprisingly, there was no significant difference between the two groups in terms of declarative learning. However, it is possible that the increased age range in the dyslexic group is an explanation for this finding, Regardless of this, the dyslexic group did not show the same relationship between declarative learning and reading as the control group. There was a clear positive relationship between declarative learning and both word and non-word reading in the typically developing group that was absent in the dyslexic group. The procedural deficit hypothesis claims that declarative learning plays a compensatory role in language disorder (Ullman & Pullman, 2015). If this is correct, then we should expect to find a relationship between verbal declarative learning and reading in the dyslexic group and yet we do not. In spite of overall equivalent scores between groups on the verbal declarative memory task, the lack of such a relationship in the dyslexic group supports the view that verbal memory is indeed impaired in dyslexia (Howes, Bigler, Burlingame, & Lawson, 2003; Griffiths & Snowling, 2002; Kramer, Knee, & Delis, 2000; Mcdougall, Hulme, Ellis, & Monk, 1994; Ovler, Obrzut, Asbiornsen, 2012; Tijms, 2004), at least in so far as it relates to reading ability.

9.4.3 Attention and response inhibition

The two groups also showed similar levels of attention during the serial reaction time task, in spite of the older age of the dyslexic group. However, one interesting finding is that the dyslexic group was significantly worse at inhibiting responses on the GoNoGo task compared to the typically developing group. Poorer performance on executive function tasks such as GoNoGo in children with dyslexia has been linked to ADHD comorbidity (Gooch, Snowling, & Hulme, 2011; Wilcutt et al., 2001). However, this study found no link between response inhibition and attention rating on the task in this group, which could have been expected if comorbid ADHD symptoms were responsible for impaired serial reaction time task performance. This suggests that the level of attention paid to the serial reaction time task reflects a somewhat different aspect of attention from that which is impaired in ADHD.

In summary, we found that both typically developing and dyslexic children matched for reading ability demonstrated equivalent levels of procedural learning on a serial reaction time task, but that this procedural learning did not relate to measures of reading fluency in either group. Declarative learning in the two groups was also comparable. However, here some differences between the groups were apparent. Declarative learning was related to reading fluency in the typically developing group, while no such relationship was found in the dyslexic group. The results of this study call into question the claims of the procedural deficit hypothesis, casting doubt on the assumption that impaired procedural learning is behind the pattern of deficits seen in dyslexia or that declarative learning plays a compensatory role over time.

Chapter 10 General discussion

10.1 Summary

This thesis set out to investigate the claims of the procedural deficit hypothesis of language learning disorders. The hypothesis incorporates a multiple systems view of memory, whereby a domain-general procedural memory system is responsible for the extraction of the statistical regularities that govern language. This system is separate from a declarative memory system that supports word-specific knowledge. It has further been suggested that it is particularly the procedural learning of sequentially, or probabilistically structured, information that is involved in language learning (Christiansen, Dale, Ellefson, & Conway, 2002). According to this hypothesis deficits in the procedural memory system are one major causal risk factor for language learning disorders, such as dyslexia and developmental language disorder (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005). Research to date has used a number of implicit learning tasks to test the theory, but has so far delivered inconsistent results.

A review of the literature in Chapter 3 revealed that there are a number of possible reasons why this might be the case. In the main, research has used extreme group designs, frequently with small sample sizes, giving rise to potential issues with reliability, as well as with classification into groups, since developmental language disorder and dyslexia are both heterogenous and dimensional disorders. Another issue with existing research is that typically only one or very few procedural learning measures are taken, often without declarative measures for comparison. Similarly, learning in verbal and non-verbal modalities is rarely examined in the same participants. However, the multiple systems view of memory is central to the claims of the hypothesis. Without multiple measures of different types of memory (procedural, declarative, verbal and non-verbal) across several cognitive domains it is not possible to assess the claim that any procedural deficit in language learning disorders is the product of a unitary, domain-general procedural memory system or that declarative memory is unimpaired.

10.1.1 The meta-analysis

The starting point for this thesis was a series of meta-analyses (Chapter 4) assessing the existing evidence for a procedural learning deficit in language learning disorders. It included separate analyses for the main implicit learning tasks (serial reaction time task, Hebb serial order learning task, artificial grammar learning and statistical learning tasks and the weather prediction probabilistic category learning task).

The meta-analyses yielded a significant, but small, pooled effect size reflecting poorer procedural learning in language-disordered compared to control groups in studies using serial reaction time tasks (pooled effect size g = -.28, 95% CI [-0.37, 0.03]) or the Hebb learning task (pooled effect size g = -.31, 95% CI [-0.21, 0.23]). By far the largest number of studies investigating the procedural deficit hypothesis used the serial reaction time task. This meta-analysis could only include approximately half of these studies, as most reported insufficient data to calculate effect sizes and many were unable to supply additional data on request. However, the number of studies included was considerably greater than in any previous relevant meta-analyses. However, the small effect size in the analysis of group design studies was not corroborated by a similar result for the analysis of correlational studies using the serial reaction time task, which did not find a significant relationship between serial reaction time performance and language ability. If performance on the serial reaction time task taps a cause of language or reading difficulties, one should expect to find a correlation between learning on the task and language or reading skills in unselected groups. The failure to replicate group deficits in procedural learning on the serial reaction time task in correlational designs casts doubt on the view that the serial reaction time task is a genuine correlate of reading or language skills.

The effect size for group design studies using the serial reaction time task found by the current meta-analysis is much lower than those of several previous smaller meta-analyses (Lum et al., 2013; Lum et al., 2014; Obeid et al., 2016). This is likely to be because the analyses in this thesis took account of the between participants

variability for both components of the derived implicit learning measures, something previous meta-analyses had not done. The small effect sizes for the studies in the analysis likely result from the large amount of variance they encorporate, which in turn suggests that the tasks used in previous studies may not have been reliable and this is a theme we will return to later in this chapter. It should be noted that none of the studies in the series of meta-analyses in Chapter 4 reported the reliability of the implicit learning tasks they used.

A meta-analysis of probabilistic category learning using the weather prediction task returned the largest significant overall effect size for the difference between language disordered and control groups in the series of analyses (g = -0.63, 95% CI [-1.07, -0.19]), but there were too few studies to examine the potential effect of moderators such as IQ and declarative learning. The group design meta-analysis of artificial grammar learning and statistical learning concluded that there was a significant overall difference between groups (g = -0.53, 95% CI [-0.79, -0.28]), although the accompanying funnel plot indicated that the true effect size, although significant, may be quite small. Surprisingly, no moderators were found to explain any of the variance between studies in these analyses. Since the nonsignificant variation in effect sizes (or size) of the other analyses precluded investigation of potential moderators, it is clear that there are still many unanswered questions surrounding the claims of the procedural deficit hypothesis. Overall, however, the results of the series of meta-analyses in this thesis were in line with a recent review of group design studies using serial reaction time tasks and / or artificial grammar tasks with dyslexia and control participants (Schmalz et al., 2016). They concluded that the effect size for studies using the artificial grammar learning task was likely to be small. They further judged that evidence for a deficit in procedural learning in groups with dyslexia in studies using the serial reaction time task was "underwhelming" (p. 10), but that these studies reported too little of the necessary information to conduct a meta-analysis at all.

A final notable finding was the scarcity of existing evidence to support the claim that any procedural deficit on the tasks was domain-general, with only four studies

included in the series of meta-analyses reporting significant deficits in language-disordered groups across more than one type of task (Hsu & Bishop, 2014; Lee & Tomblin, 2015; Lukacs & Kemeny, 2014; Vicari et al., 2005).

10.1.2 The experimental studies

The experimental studies in this thesis were formulated in order to avoid the methodological limitations of the existing research detailed at the beginning of this chapter. The decision was taken to use a correlational design to examine the range of language and literacy ability in a large unselected sample of children and to investigate the relationship with both procedural and declarative memory skills on several tasks and to do so in both verbal and non-verbal modalities. Crucially, the reliability of all the implicit learning tasks was carefully assessed. None of the resulting studies found any evidence that procedural learning is related to language ability.

In the first study (Chapter 5), 101 children completed a battery of language-related attainment tests; serial reaction time, Hebb serial order learning and contextual cueing tasks, as well as measures of free and immediate serial recall. Implicit learning measures were of a comparable length to those in the literature. The measures of declarative memory were reliable and loaded on separable verbal and non-verbal latent factors. Furthermore, in line with previous research into language disorders and declarative memory, variations in verbal declarative memory were stronger correlates of language, literacy and arithmetic skills than variations in non-verbal declarative memory. By contrast, the procedural memory measures were found to have poor reliability and showed no appreciable correlation with each other or with measures of attainment.

In an effort to improve the reliability of the serial reaction time task the next study trialled an extended version of the task with adult participants (Chapter 6). The reliability of the extended task was acceptable in this sample and so the second large-scale study in the thesis examined the relationships between performance on an extended 1000 trial serial reaction time task, verbal declarative learning and attainment

in a representative sample of 112 children (Chapter 7). A measure of attention assessed during the serial reaction time task was also taken. Reliability of the extended task in children was poor, displaying only a slight improvement compared to the first study and the task remained a weak correlate of reading and language. Crucially, a latent variable path analysis showed that the measure of attention taken during the task accounted entirely for the relationship between procedural learning and measures of reading, grammar and arithmetic. Verbal declarative learning, on the other hand, once again showed a significant relationship with attainment.

The next study was motivated by the decision to explore another facet of implicit learning critically involved in both language acquisition and use: probabilistic category learning. The weather prediction task is a category learning task that has been used by several studies to explore the procedural deficit hypothesis (see Chapter 4). The aim of this was to further explore the claim that deficits in procedural learning in language disorder are domain-general. It assessed the relationship between the weather prediction task, declarative learning and attainment (Chapter 8). While the reliability of this task was moderate, further investigation suggested that this may have been a consequence of the extent of declarative learning required to perform the task well. The task did not correlate with attainment, nor did it correlate with the extended serial reaction time task administered to the same children in Chapter 7. The study also highlighted additional limitations of this task, particularly in children, with the majority of participants failing to perform above chance. The subset of children who did perform above chance showed a slightly different pattern of results to the overall sample, but participant numbers in this subset were low ($n = 38$ of 107). This has clear implications for the task from a psychometric point of view, since tasks where a large proportion of participants perform no better than chance are typically insensitive to individual differences (Siegelman, Bogaerts, Christiansen, & Frost, 2017).

The final study investigated procedural learning in dyslexic children using the extended serial reaction time task (Chapter 9). Previous studies investigating the procedural deficit hypothesis have almost all used age-matched group designs (NB: Hsu & Bishop, 2014 also included an ability-matched group). In age-matched group

designs it is impossible to disentangle significantly poorer procedural learning in language-impaired groups from reading ability, since poorer reading may be a cause rather than a consequence of procedural learning impairment. The study in this chapter instead matched a group of dyslexic children with a control group on reading ability. A significant difference between groups with this design would show that any difference between groups was not simply a consequence of differences in reading ability. However, the results of the study showed that the groups were equated for procedural learning performance. In addition, there was some indication that reading ability related to verbal declarative memory skill in the control group that was not found in the dyslexic group. While a null result using this design should not be interpreted as evidence of no relationship between procedural learning and reading ability, the weight of the evidence from the studies conducted for this thesis clearly fails to support the procedural deficit hypothesis.

10.2 Key findings

We will now look at the key implications of these findings in more detail.

10.2.1 Task reliability and different approaches to psychological investigation

The procedural learning tasks developed in Chapter 5 are representative of those used in the literature, yet they were all found to be unreliable. Previous research has not reported task reliability and this may be because historically reliability has meant different things to the experimental and correlational approaches to psychological investigation. A reliable task from a group design perspective is a task that consistently returns the same results (e.g., faster RTs to sequenced compared to random trials or better recall for repeated compared to unrepeated lists). Here, replicability is key. For correlational studies a reliable task is one that consistently ranks individuals in the same order across attempts. Hedge, Powell, and Sumner (2017) point out that tasks that have a proven track record in group design research may not translate well to research using correlational designs. This is because what makes a task reliable from a group design perspective is low variance between participants in a group, so that the averaged response is consistently the same from experiment to experiment. However,

this is the same quality that can make a task unsuited for use in correlational designs, since low between participants variance makes is difficult to detect relationships between variables at an individual differences level. It is, therefore, important to consider whether this is the case for the tasks in this thesis. Are the procedural learning tasks used in this thesis reliable from an experimental perspective, justifying the approach taken in previous research, but just not suited to the correlational approach used here? It is true that a substantial number of group design studies have found poorer implicit learning in groups with developmental language disorder or dyslexia compared to normal or typically developing groups. However, if the independent variable of language disorder (either dyslexia or developmental language disorder) in group design studies is supposed to relate to a deficit in procedural learning, then every well-conducted study should produce a similar result (impaired procedural learning performance of the experimental group) and this is not what is found. The results across the literature are highly inconsistent (see Chapters 3 and 4). While the existence of the occasional null result can be ascribed to experimental confounds, the 50:50 split of significant to non-significant results for the serial reaction time task documented in Chapter 4 is very difficult to reconcile with the procedural deficit hypothesis. Results for the other paradigms are similarly inconsistent. Alongside the large amount of variance forming part of the effect size calculation for each study in the meta-analyses for the serial reaction time and Hebb serial order learning tasks, this suggests that the tasks are no more reliable as tools for experimental research than they are for correlational use. The poor reliability of the implicit learning tasks found in this thesis is, therefore, of critical importance regardless of the choice of experimental design.

10.2.2 Why are implicit learning tasks unreliable?

Why might the reliability of the procedural learning tasks be so low? Ostergaard (1998) noted that the relative contribution of learned information is likely to be far lower in procedural than declarative tasks. In a declarative task like word list recall, there is minimal external stimulus information for the participant to process at recall and hence variation in memory integrity is likely to cause most of the variance in performance. In a procedural task such as contextual cuing, by contrast, each trial

278

evokes a number of perceptual as well as motoric processes that will contribute to variance in performance over and above learned sequence knowledge. If a target is embedded amongst 12 distractors in a contextual cueing display, for example, then variation in basic perceptual processes (scanning across the objects until the target is identified) and response selection and execution will all contribute to measured variance. Any relevant procedural information that can be retrieved from memory about the likely location of the target in a familiar display will make only a small contribution to the RT on a given trial. The same can be said of the other procedural learning tasks, such as the serial reaction time task. Ostergaard formalized this idea in his Information Availability model. According to this model, when the relative contribution of learned information to performance is low, the reliability of the task for measuring that learned information will be low too.

10.2.3 Task reliability from a developmental perspective

Another related finding from the thesis is that procedural learning tasks may be particularly unreliable in children. The age invariance of implicit learning (Reber, 1993; Vinter & Perruchet, 2000; Meulemans, Van der Linden, & Perruchet, 1998) has been cited as a key difference between implicit and declarative learning (e.g., Schacter & Moscovitch, 1984) and this difference used as evidence to support multiple systems views of memory. While many studies have reported intact implicit learning in childhood and even in infancy (see Chapter 2), the developmental invariance of implicit learning is not consistently agreed upon, with a number of studies demonstrating developmental trajectories across a range of tasks (e.g., Arciuli & Simpson, 2011; Thomas et al., 2004; Vaidya, Huger, Howard, & Howard, 2007). However, if implicit learning tasks are particularly unreliable in children, compared to older children or adults (as demonstrated by the poor reliability of the serial reaction time task in Chapters 5 and 7, compared to reasonable reliability of the same task in adults in Chapter 6), then the developmental invariance of implicit learning cannot be reliably assessed, since implicit learning measures in younger participants are likely to be less reliable than those in older participants. The multiple memory systems model evolved primarily from neuropsychological research into dissociations in performance

279

between normal and clinical populations, for the most part once normal development had taken place. The poor reliability of the tasks in young participants thus raises a question over how useful it is to apply such a model in a developmental context.

10.2.4 The role of attention and task engagement in implicit learning

Another key finding in the thesis links the issue of reliability with concerns about the validity of the tasks as measures of implicit learning. Chapter 7 demonstrated that the level of attention children paid to the SRT task was strongly related to procedural learning and, further, that it accounted entirely for the relationship between procedural learning and attainment. This finding is in line with research that has found that selective attention of relevant stimuli is a prerequisite for robust implicit learning (Frensch & Runger, 2003; Jiang & Chun, 2001; Shanks, 2005). Certainly, dividing attention by administering implicit learning tasks in dual task scenarios has been shown to have a detrimental effect on performance (e.g., Nicolson & Fawcett, 1990; Fawcett & Nicolson, 1992; Shanks & Channon, 2002; Turk-Browne et al., 2005; Yap & Leij, 1994), although there is debate over whether the secondary task interferes with the implicit learning itself or just with the expression of that learning.

Consideration of attention as a moderator on implicit learning in studies surrounding the procedural deficit hypothesis has typically been limited to the screening of participants for symptoms or history of ADHD. A few previous studies have questioned whether their results were related to the poorer attentional resources of their experimental groups (e.g., Kelly et al., 2002; Saffran & Robe-Torres, 2009; Waber et al., 2003; Staels & Van den Broek, 2015). However, there is now growing interest in the potentially confounding influence of attention in implicit learning (Staels & Van den Broek, 2017; Sigurdardottir et al, 2017), as well as the role attention plays in language development (de Diego-Balaguer, Alvarez, & Pons, 2016). The results in Chapter 7 lend weight to the suggestion that procedural learning performance is moderated by attention. However, future investigation will benefit from refining the definition of attention, in order to better distinguish between attentional capacity and the allocation of attentional resources (Frensch & Runger, 2003). This can perhaps be

summed up as the difference between considering attention as a trait or as a state. Sustained attention as a trait could be described as the ability to maintain a vigilant state (Posner & Petersen, 1990). It refers to enduring attentional resources and is investigated using measures such as continuous performance tasks that ask participants to monitor a stream of stimuli and respond to pre-specified targets. ADHD scales also index attentional resources from a trait perspective. Attention as a state, on the other hand, also encompasses more fluctuating states, such as level of task engagement and focus, which have been shown to facilitate sustained attention (Matthews et al., 2010) and task performance. These two aspects of attention (trait and state) may be related, but they are not synonymous. The former has implications for the validity of the tasks as measures of procedural learning, but the implications of the latter are more focused on the reliability of the task. Intriguingly, Sigurdardottir et al. (2017) found a continuous performance test a poor predictor of statistical learning in adult dyslexic and control participants, while a history of childhood ADHD was a far better predictor, so there may be important differences between measures of current and historical trait attention too.

10.2.5 The role of declarative memory processes in implicit learning

A further finding from the thesis, that declarative learning related to performance on the weather prediction task (Chapter 8) links specifically to an important ongoing debate about the process purity of implicit learning tasks and hence their validity as measures of procedural learning. It also provides another possible reason for inconsistent findings in the previous literature surrounding the procedural deficit hypothesis. The current results lend some support to the assertion that implicit learning tasks are not process pure (e.g., Reber, 1989; Shanks & John, 1994; Cleeremans, Destrebecqz, & Boyer, 1998) and are likely to involve declarative learning to a variable and an, as yet, unknown degree.

This has been noted in previous research. For example, Guerard et al. (2011) found that aware participants recalled repeated sequences better than non-aware participants on a Hebb serial order learing task. Buchner, Steffens, Erdfelder, and Rothkegel (1998)

demonstrated that declarative learning of sequence fragments contributed to performance on a serial reaction time task and was present from early in the task, while Perruchet & Amorim (1992) found implicit learning on the serial reaction time task was highly correlated to explicit knowledge of the sequence, as tested by a post-task generation test (*r*'s > .75).

Disentangling the relative contributions of declarative and implicit learning to task performance is far from simple, although attempts have been made (e.g., the process dissociation procedure: Jacoby, 1991). However, the use of direct tests of explicit learning were not included in the serial reaction time, Hebb serial order learning or contextual cueing tasks in this thesis for the following reasons. Shanks and John (1994) highlight difficulties in devising measures that actually index the knowledge that is responsible for the change in performance, and that are sensitive to all relevant declarative knowledge, without falsely attributing elements of declarative knowledge as implicit. This is even more of an issue for probabilistic tasks, such as the SRT task in this thesis, as a result of the greater variety of material seen by participants. Additionally, both declarative and procedural meta-cognitive knowledge is thought to develop with age (Schneider, 2010; Schneider & Lockl, 2008) and the children in these studies were young.

Nevertheless, the extent of declarative learning on the weather prediction task was assessed. In so far as the post-task explicit judgement test can be considered an accurate measure of declarative learning on the task[3], results demonstrated that declarative learning contributed to task performance. In addition, task performance related to a separate measure of verbal declarative learning in the subset of participants who performed above chance. Given these results, it is reasonable to conjecture that declarative learning does contribute to performance on probabilistic category learning tasks and that the superior performance of control groups compared to groups with

---

[3] Gluck et al (2002) caution that labelling performance strategies on the task as declarative or implicit may be overly simplistic. Incremental learning may occur in a non-declarative fashion, but the learning strategies that develop using this process may also be verbalisable. Similarly, a simple and memorable performance strategy, such as the single cue strategy on the weather prediction task, does not necessarily have to be declarative.

language learning disorders in previous research may reflect declarative memory skills to a degree.

The extent to which declarative learning may contribute to measures of procedural learning in serial reaction time tasks is less clear. The latent variable path model in Chapter 7 showed that verbal declarative learning correlated only weakly with performance on the extended serial reaction time task ($r = .22$). This suggests that superior verbal declarative memory skills did not greatly aid performance on the serial reaction time task in this thesis. However, task variations, such as using complex sequence structures, purportedly upweight the contribution implicit learning makes to derived procedural learning measures (Cohen, Ivry, & Keele, 1990) and such complex sequence structures were purposely used in this thesis. By contrast, the bulk of studies investigating the procedural deficit hypothesis have used simpler sequences and / or deterministic tasks. It is, therefore, not unreasonable to suppose a greater contribution of declarative learning to serial reaction time task performance in the previous studies using simpler sequence structures.

The impact of declarative learning on implicit learning task performance is too often ignored in the literature surrounding the procedural deficit hypothesis, with significantly poorer task performance in groups with developmental language disorder or dyslexia attributed entirely to a deficit in the procedural memory system. There is a problem inherent in ascribing the poorer performance of experimental groups to a deficit in procedural learning on the basis of their performance on a task that is not a process pure measurement of procedural learning. Additionally, Sun, Slusarz, and Terry (2005) proposed that not only do implicit learning tasks inevitably involve both implicit and declarative learning, but that these processes interact, often synergistically, to aid performance. This makes holding procedural memory responsible for differences in performance on procedural learning tasks especially problematic.

10.2.6 The role of motor learning in implicit learning

The results in this thesis also suggest that motor learning ability may contribute to the group differences on the serial reaction time task found in developmental language disorder and dyslexia in previous research. Children with lower levels of attainment in the first study (Chapter 5) were slower across both probable and improbable sequences on the serial reaction time task. The results in Chapter 9 also showed that the slope reflecting early (motor) learning was flatter for the dyslexic group than for the control group on each of the component sequences of the serial reaction time task. Some previous studies have also noted similar differences in motor learning between groups with dyslexia and developmental language disorder (e.g., Henderson & Warmington, 2017; Laasonen et al., 2014; Mayor-Dubois et al., 2014; Vakil et al., 2015).

Evidence of motor deficits have been found in poor readers using the serial reaction time task (Stoodley, Harrison, & Stein, 2006) and in those with developmental language disorder more generally (Bishop, 2002; Brookman, McDonald, McDonald, & Bishop, 2013; Hill, 2001). Ise & Schulte-Körne (2012) conjectured that group differences in performance on the serial reaction time task in dyslexia may reflect problems with motor learning on the task. They comment that a disproportionate number of studies using simple sequences report a significant difference between groups compared to the overall ratio of significant to null results. They relate this to the fact that simple sequences can be more easily learned as a sequence of finger movements than complex sequences. The motor learning component is, therefore, more evident in results from studies with simpler sequences, but more likely to be overlaid by other extraneous factors in studies using complex sequences.

So it can be seen that a number of cognitive functions may influence procedural learning in the tasks used to assess the procedural deficit hypothesis. This may happen in unsystematic ways, adding noise to the data and contributing to the poor reliability of the tasks, but may also happen in more systematic ways, whereby cognitive processes unrelated to a strict definition of procedural learning may be at least partially

responsible for the impaired procedural learning shown by groups with developmental language disorder and dyslexia in some previous research.

10.2.7 Implications for the multiple memory systems debate

The results of the studies in this thesis also have implications for the procedural deficit hypothesis position that the procedural learning deficit in language learning disorders is domain-general. This issue forms part of the larger debate about the existence of multiple memory systems that govern different types of learning and memory.

On one side of the debate, multiple memory systems proponents furnish evidence from neuro-psychological and neuro-imaging research in support of the existence of separable memory systems (see Chapter 2), which include a domain-general implicit learning system (e.g., Poldrack & Foerde, 2008). However, this evidence is only persuasive if we can accept that the tasks upon which participants are measured provide reliable, relatively process-pure measures of implicit learning. The results in this thesis strongly suggest that they do not.

Nevertheless, let us leave aside the question of task reliability and validity for the moment and consider the results of the studies in the thesis from a multiple systems standpoint. If the multiple memory systems view of a domain-general procedural memory system is correct, one would expect to find correlations between implicit learning measures on different tasks and across verbal and non-verbal modalities, yet the current results failed to show any evidence of this at all. Not only was there no relationship between Hebb serial order learning, serial reaction time and contextual cueing tasks in Study 1 (Chapter 5), which could conceivably be explained by different task demands, there was also no relationship between verbal and non-verbal analogue versions of the same tasks. Indeed, in the case of the contextual cueing task, there was no correlation between verbal and non-verbal conditions on the same task. Neither was there a relationship between performance on the weather prediction task and the extended serial reaction time tasks administered to the same children in Study 3 and 4

(Chapters 7 and 8). Although very little previous research has examined relationships between different implicit learning paradigms at an individual level, a recent study reports results that are consistent with the findings in this thesis, with low correlations between implicit learning measures across different types of task (serial reaction time and statistical learning tasks), as well as across visual and auditory modalities (Siegelman & Frost, 2015). Therefore, at first glance this task specificity clearly sits in opposition to the procedural deficit view of implicit learning as a domain-general mechanism, adding support to the argument that implicit learning is not a unitary construct.

However, returning once more to the issue of reliability, the low test-retest reliability of the tasks in the thesis makes it difficult to attribute the low correlations between tasks (ie: Hebb, serial reaction time, contextual cueing and weather prediction) to the specificity of implicit learning or to draw any conclusions about whether implicit learning differs across verbal and non-verbal modalities. There seems to be an inevitable Catch 22 with the poor reliability of the existing tasks. While the lack of supporting evidence for domain-generality using these tasks in previous research and the low correlations between the tasks in this thesis cast doubt on the unitary nature of procedural learning, the poor task reliability makes it difficult to rule out entirely whether procedural learning as a stable domain-general ability does exist, since the tasks as they currently exist may simply be incapable of measuring it.

Here, another recent study reporting low correlations between closely related statistical learning tasks in adults may be helpful (Erikson, Kaschak, Thiessen, & Berry 2016). Testing participants on such closely related tasks makes it less likely that differing task demands and the resulting recruitment of divergent cognitive functions are responsible for the lack of correlation. This is the same rationale behind testing participants on SRT tasks that were identical in all but sequence in Chapters 6 and 7. Erikson et al. (2016) showed that their tasks had low test-retest reliability, but critically they were able to show that improving the test-retest reliability of their measures (with composite implicit learning measures and a longer test phase) did not improve the correlations between tasks. It seems logical to suppose that if highly similar tasks with

reasonable levels of test-retest reliability do not correlate, then implicit learning may not be a unitary ability.

On the other side of the multiple memory systems debate, connectionist accounts aim to demonstrate how the apparent dissociations between declarative and implicit memory in amnesia, for example, can be predicted by simple, single-system connectionist models of learning (e.g., Kinder & Shanks, 2003). This apparently more parsimonious view suffers from its own challenge, however. Poldrack and Foerde, (2008) point out that single systems views can suffer from "parameter proliferation", whereby the fit of any given connectionist model is improved by adding parameters that relate to mechanisms that belong to other cognitive systems entirely. Such single system explanations are then arguably no more parsimonious than multiple systems views. While the findings in this thesis do suggest that cognitive processes unrelated to a "pure" construct of implicit learning may systematically influence performance on the tasks (i.e., attention, declarative learning or motor learning), the findings also make clear that such processes, along with many others may affect the stability of the tasks in unsystematic ways too. It is primarily the reliability, rather than validity, of implicit learning tasks that is questioned by the results in this thesis.

Accounts that seek to reconcile the ubiquity of implicit learning across multiple domains with evidence of considerable task specificity can be visualised as sitting somewhere between the multiple memory systems and connectionist positions. For example, Frost, Armstrong, Siegelman, and Christiansen (2015) suggest that the domain-general, unitary nature of procedural learning extends only as far as a set of computational principles. However, these principles are constrained by modality, as a result of the different combinations of cortical areas that must be recruited to perfom each task. Different tasks will recruit different combinations of cortical areas and the result will be considerable specificity for implicit learning. The studies in this thesis showed significant procedural learning on tasks across a wide range of cognitive domains and across verbal and non-verbal modalities (from sequence learning and perceptual priming to category learning), alongside a lack of any relationship between

them and this combination is not inconsistent with Frost et al's (2015) explanation, yet the question of poor task reliability remains.

Arciuli (2017) adds that since other cognitive processes have a systematic influence on implicit learning, implicit learning itself must be viewed as a multi-component ability made up of many functions from processing speed and memory to attention. The results in the thesis are not inconsistent with this position also, but once again the poor reliability of the tasks upon which procedural learning is measured precludes any definitive conclusions.

10.2.8 Procedural learning as an individual difference

The above clearly has implications for procedural learning from an individual differences perspective, since two of the key criteria that procedural learning needs to fulfill in order to be considered an individual difference variable is that it needs to be both a unitary ability and a stable characteristic of the individual (Siegelman & Frost, 2015).

It seems that multiple memory systems views are more likely to see procedural learning as a legitimate individual difference variable, while single system connectionist views are not. The positions held by Frost et al. (2015) and Arciuli (2017) fall somewhere in between. For Frost et al. (2015), individual differences in procedural learning reflect relative strengths and weaknesses at a neural level (such as white matter density, for example), while for Arciuli (2017) individual differences in implicit learning would reflect both how well individual component functions of implicit learning perform, as well as how optimally they are connected. Given the poor reliability of the implicit learning paradigms currently available, it can be argued that from a psychometric view, these instantiations of procedural learning as an individual difference variable may be of limited practical use. It remains that any questions about the nature of procedural learning and the relationship between procedural learning and language ability must wait until there are reliable tasks with which to investigate them.

10.3 Final words

To conclude, the studies in this thesis have found no evidence that poorer procedural learning is related to lower levels of language-related attainment. Moreover, these studies have uncovered a number of possible reasons for the pattern of inconsistent findings in literature relating procedural learning deficits to developmental language disorder and dyslexia. These reasons hinge on the both the reliability and the validity of the tasks used to investigate procedural learning. The results of the studies in this thesis seriously question the suggestion that the construct of a "procedural learning system" can be reliably measured and cast strong doubt on claims from earlier studies that deficits in such a system are related to language learning difficulties.

Chapter 11   References

Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, *23*(6), 1358-1367.

Adams A. M., Gathercole., S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders*, *35*(1), 95-116.

Aguilar, J. M., & Plante, E. (2014). Learning of grammar-like visual sequences by adults with and without language-learning disabilities. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1394-1404.

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*(1), 64-88.

Alloway, T. P., & Archibald, L. (2008). Working memory and learning in children with developmental coordination disorder and specific language impairment. *Journal of Learning Disabilities*, *41*(3), 251-262.

Allen, R., & Reber, A. S. (1980). Very long term memory for tacit knowledge. *Cognition*, *8*(2), 175-185.

Andrade, J., & Baddeley, A. (2011). The contribution of phonological short-term memory to artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, *64*(5), 960-974.

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675-693.

Archibald, L. M., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory & Cognition*, *41*(2), 268-280.

Arciuli, J. (2017). The multi-component nature of statistical learning. *Phil. Trans. R. Soc. B*, *372*(1711), 20160058.

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464-473.

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

Baillieux, H., Vandervliet, E. J., Manto, M., Parizel, P. M., De Deyn, P. P., & Mariën, P. (2009). Developmental dyslexia and widespread activation across the cerebellar hemispheres. *Brain and Language*, *108*(2), 122-132.

Badcock, N. A., Bishop, D. V., Hardiman, M. J., Barry, J. G., & Watkins, K. E. (2012). Co-localisation of abnormal brain structure and function in specific language impairment. *Brain and Language*, *120*(3), 310-320.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). Memory. East Sussex.

Bahl, M., Plante, E., & Gerken, L. (2009). Processing prosodic structure by adults with language-based learning disability. *Journal of Communication Disorders*, *42*(5), 313-323.

Baird, G., Dworzynski, K., Slonims, V., & Simonoff, E. (2010). Memory impairment in children with language impairment. *Developmental Medicine & Child Neurology*, *52*(6), 535-540.

Barbiero, C., Lonciari, I., Montico, M., Monasta, L., Penge, R., Vio, C., ... & Carrozzi, M. (2012). The submerged dyslexia iceberg: how many school children are not diagnosed? Results from an Italian study. *PloS One*, *7*(10), e48082.

Bartsch, T., & Butler, C. (2013). Transient amnesic syndromes. *Nature Reviews Neurology*, *9*(2), 86-97.

Bavin, E. L., Wilson, P. H., Maruff, P., & Sleeman, F. (2005). Spatio-visual memory of children with specific language impairment: evidence for generalized processing problems. *International Journal of Language & Communication Disorders*, *40*(3), 319-332.

Beckers, G. J., Berwick, R. C., Okanoya, K., & Bolhuis, J. J. (2016). What do animals learn in artificial grammar studies?. *Neuroscience & Biobehavioral Reviews*.

Belleville, S., Caza, N., & Peretz, I. (2003). A neuropsychological argument for a processing view of memory. *Journal of Memory and Language*, *48*(4), 686-703.

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*(1), 184-198.

Ben-Yehudah, G., & Fiez, J. A. (2008). Impact of cerebellar lesions on reading and phonological processing. *Annals of the New York Academy of Sciences*, *1145*(1), 260-274.

Bernal, B., & Altman, N. R. (2003). Speech delay in children: a functional MR imaging study. *Radiology*, *229*(3), 651-658.

Berry, D. C., & Dienes, Z. (1991). The relationship between implicit memory and implicit learning. *British Journal of Psychology*, *82*(3), 359-373.

Berry, D., & Dienes, Z. P. (1993). *Implicit learning: Theoretical and empirical issues*. Psychology Press.

Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. (2012). Models of recognition, repetition priming, and fluency: exploring a new framework. *Psychological review*, *119*(1), 40.

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology*, *12*(3), 205-221.

Bishop, D. V. M. (1982). Test for Reception of Grammar (Medical Research Council, Oxford).

Bishop, D. V. (1994). Is specific language impairment a valid diagnostic category? Genetic and psycholinguistic evidence. *Philosophical Transactions: Biological Sciences*, 105-111.

Bishop, D. V. (2002). Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics Part A*, *114*(1), 56-63.

Bishop, D. V. (2003). *Test for reception of grammar: TROG-2 version 2*. Pearson Assessment.

Bishop, D. V. (2006). What causes specific language impairment in children?. *Current Directions in Psychological Science*, *15*(5), 217-221.

Bishop, D. V. (2013). Cerebral asymmetry and language development: cause, correlate, or consequence?. *Science*, *340*(6138), 1230531.

Bishop, D. V. M., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory Temporal Processing Impairment: Neither Necessary nor Sufficient for Causing Language Impairment in Children. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1295-1310.

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858.

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of child Psychology and Psychiatry*, *37*(4), 391-403.

Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016b). CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. Terminology. *PeerJ Preprints*, *4*, e2484v1.

Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain*, *133*(3), 868-879.

Bloom, J. S., Garcia-Barrera, M. A., Miller, C. J., Miller, S. R., & Hynd, G. W. (2013). Planum temporale morphology in children with developmental dyslexia. *Neuropsychologia*, *51*(9), 1684-1692.

Bo, J., Jennett, S., & Seidler, R. D. (2012). Differential working memory correlates for implicit sequence performance in young and older adults. *Experimental Brain Research*, *221*(4), 467-477.

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, *43*, 106-122.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis version 2. *Englewood, NJ: Biostat*, *104*.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, *22*(3), 441-469.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., ... & Hulme, C. (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*(4), 422-432.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature*, *301*(5899), 419-421.

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 798.

Breznitz, Z., & Meyler, A. (2003). Speed of lower-level auditory and visual processing as a basic factor in dyslexia: Electrophysiological evidence. *Brain and Language*, *85*(2), 166-184.

Norbury, C. F., Bishop, D. V., & Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module?. *Applied Psycholinguistics*, *23*(2), 247-268.

Brookman, A., McDonald, S., McDonald, D., & Bishop, D. V. (2013). Fine motor deficits in reading disability and language impairment: same or different?. *PeerJ*, *1*, e217.

Brooks, D. N., & Baddeley, A. D. (1976). What can amnesic patients learn?. *Neuropsychologia*, *14*(1), 111-122.

Brosnan, M., Demetre, J., Hamill, S., Robson, K., Shepherd, H., & Cody, G. (2002). Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia*, *40*(12), 2144-2155.

Brown, J., Aczel, B., Jiménez, L., Kaufman, S. B., & Grant, K. P. (2010). Intact implicit learning in autism spectrum conditions. *Quarterly Journal of Experimental Psychology*, *63*(9), 1789-1812.

Brown, R. M., & Robertson, E. M. (2007). Off-line processing: reciprocal interactions between declarative and procedural memories. *Journal of Neuroscience*, *27*(39), 10468-10475.

Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., & Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for Wernicke's Wortschatz?. *Brain*, *122*(10), 1901-1917.

Buchner, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 631-663.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, *40*(3), 227-259.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Bussy, G., Krifi-Papoz, S., Vieville, L., Frenay, C., Curie, A., Rousselle, C., ... & Herbillon, V. (2011). Apprentissage procédural implicite dans la dyslexie de surface et la dyslexie phonologique. *Revue de neuropsychologie*, *3*(3), 141-146.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376.

Cabeza, R, & Nyberg, L. (2000a). Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies. *Journal of Cognitive Neuroscience, 12*(1), 1-47.

Cabeza, R., & Nyberg, L. (2000b). Neural bases of learning and memory: Functional neuroimaging evidence. *Current Opinion in Neurology, 13*(4), 415-421.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., ... & Seidlová-Málková, G. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, 0956797611434536.

Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: Evidence for comorbidity. *Journal of Child Psychology and Psychiatry*, *46*(5), 524-532.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*(2), 149-180.

Castles, A., Datta, H., Gayan, J., & Olson, R. K. (1999). Varieties of developmental reading disorder: Genetic and environmental influences. *Journal of Experimental Child Psychology*, *72*(2), 73-94.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378-1396.

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities*, *35*(6), 510-525.

Cermak, L. S., Lewis, R., Butters, N., & Goodglass, H. (1973). Role of verbal mediation in performance of motor tasks by Korsakoff patients. *Perceptual and Motor Skills*, *37*(1), 259-262.

Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *The Journal of General Psychology*, *130*(3), 290-304.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: a usage-based account of natural language recursion. *Frontiers in Psychology*, *6*.

Christiansen, M. H., Dale, R. A., Ellefson, M. R., & Conway, C. M. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In *Simulating the evolution of language* (pp. 165-187). Springer London.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28-71.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 224.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*(9).

Clark, G. M., & Lum, J. A. (2017). First-order and higher order sequence learning in specific language impairment. *Neuropsychology*, *31*(2), 149.

Clark, G. M., Lum, J. A., & Ullman, M. T. (2014). A meta-analysis and meta-regression of serial reaction time task performance in Parkinson's disease. *Neuropsychology*, *28*(6), 945.

Clark, A., O'hare, A., Watson, J., Cohen, W., Cowie, H., Elton, R., ... & Seckl, J. (2007). Severe receptive language disorder in childhood—familial aspects and long-term outcomes: results from a Scottish study. *Archives of disease in childhood*, *92*(7), 614-619.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406-416.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*(3), 235.

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders–a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, *46*(2), 128-149.

Cohen, M. (1997). Children's memory scale (CMS). *San Antonio: The Psychological Corporation*.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17.

Cohen, M., Campbell, R., & Yaghmai, F. (1989). Neuropathological abnormalities in developmental dysphasia. *Annals of Neurology*, *25*(6), 567-570.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207-210.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1195-1204.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry, 42,* 741-748.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language and literacy: developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry*, *48*(2), 147-156.

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539-546.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.

Corkin, S. (2013). *Permanent present tense: The unforgettable life of the amnesic patient, HM* (Vol. 1000). Basic Books (AZ).

Couperus, J. W., Hunt, R. H., Nelson, C. A., & Thomas, K. M. (2011). Visual search and contextual cueing: differential effects in 10-year-old children and adults. *Attention, Perception, & Psychophysics*, *73*(2), 334-348.

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, *3*(1).

Cronbach, L. J., & Furby, L. (1970). How we should measure" change": Or should we?. *Psychological Bulletin*, *74*(1), 68.

Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research*, *60*(1), 24-41.

Dauer, W., & Przedborski, S. (2003). Parkinson's disease: mechanisms and models. *Neuron*, *39*(6), 889-909.

DeCoster, J., Iselin, A. M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*(4), 349.

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, *7*.

Degonda, N., Mondadori, C. R., Bosshardt, S., Schmidt, C. F., Boesiger, P., Nitsch, R. M., ... & Henke, K. (2005). Implicit associative learning engages the hippocampus and interacts with explicit associative learning. *Neuron*, *46*(3), 505-520.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., ... & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359-1364.

Démonet, J. F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *The Lancet*, *363*(9419), 1451-1460.

Deroost, N., Zeischka, P., Coomans, D., Bouazza, S., Depessemier, P., & Soetens, E. (2010). Intact first-and second-order implicit sequence learning in secondary-school-aged children with developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *32*(6), 561-572.

Desmottes, L., Maillart, C., & Meulemans, T. (2017). Memory consolidation in children with specific language impairment: Delayed gains and susceptibility to

interference in implicit sequence learning. *Journal of Clinical and Experimental Neuropsychology*, *39*(3), 265-285.

Desmottes, L., Meulemans, T., & Maillart, C. (2016b). Implicit spoken words and motor sequences learning are impaired in children with specific language impairment. *Journal of the International Neuropsychological Society*, *22*(5), 520-529.

Desmottes, L., Meulemans, T., & Maillart, C. (2016a). Later learning stages in procedural memory are impaired in children with specific language impairment. *Research in developmental disabilities*, *48*, 53-68.

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., ... & Maquet, P. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. *Learning & Memory*, *12*(5), 480-490.

Di Betta, A. M., & Romani, C. (2006). Lexical learning and dysgraphia in a group of adults with developmental dyslexia. *Cognitive Neuropsychology*, *23*(3), 376-400.

Dixon, M. L., Zelazo, P. D., & De Rosa, E. (2010). Evidence for intact memory-guided attention in school-aged children. *Developmental Science*, *13*(1), 161-169.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.

Dominey, P. F., & Georgieff, N. (1997). Schizophrenics learn surface but not abstract structure in a serial reaction time task. *NeuroReport*, *8*(13), 2877-2882.

Dominey, P. F., Hoen, M., Blanc, J. M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. *Brain and Language*, *86*(2), 207-225.

Doyon, J., Gaudreau, D., Laforce Jr, R., Castonguay, M., Bedard, P. J., Bedard, F., & Bouchard, J. P. (1997). Role of the striatum, cerebellum, and frontal lobes in the learning of a visuomotor sequence. *Brain and Cognition*, *34*(2), 218-245.

Du, W., & Kelly, S. W. (2013). Implicit sequence learning in dyslexia: a within-sequence comparison of first-and higher-order information. *Annals of Dyslexia*, *63*(2), 154-170.

Duinmeijer, I. (2013). Persistent problems in SLI: which grammatical problems remain when children grow older?. *Linguistics in Amsterdam*, *6*, 28-48.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7-to 10-year-olds. *Journal of Experimental Child Psychology*, *91*(2), 113-136.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.

Dyck, M. J., Piek, J. P., & Patrick, J. (2011). The validity of psychiatric diagnoses: The case of 'specific'developmental disorders. *Research in Developmental Disabilities*, *32*(6), 2704-2713.

Eckert, M. A., Leonard, C. M., Possing, E. T., & Binder, J. R. (2006). Uncoupled leftward asymmetries for planum morphology and functional language processing. *Brain and Language*, *98*(1), 102–111.

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management*, *20*(3), 683-689.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109-120.

Erickson, L., Kaschak, M., Thiessen, E., & Berry, C. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra: Psychology*, *2*(1).

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*(2), 321-335.

Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., ... & Zorzi, M. (2010). Multisensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of Cognitive Neuroscience*, *22*(5), 1011-1025.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, *2*(4), 460-493.

Fawcett, A. J., & Nicolson, R. I. (1992). Automatisation deficits in balance for dyslexic children. *Perceptual and Motor Skills*, *75*(2), 507-529.

Feldman, J., Kerr, B., & Streissguth, A. P. (1995). Correlational analyses of procedural and declarative learning performance. *Intelligence*, *20*(1), 87-114.

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *52*(4), 915-929.

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society*, *15*(4), 501-508.

Foerde, K., Poldrack, R. A., & Knowlton, B. J. (2007). Secondary-task effects on classification learning. *Memory & Cognition*, *35*(5), 864-874.

Ford, A., Triplett, W., Sudhyadhom, A., Gullett, J. M., McGregor, K., FitzGerald, D., ... & Crosson, B. (2013). Broca's area and its striatal and thalamic connections: a diffusion-MRI tractography study. *Frontiers in Neuroanatomy*, *7*, 8.

Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory & Cognition*, *42*(1), 112-125.

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology*, *22*(9), 814-819.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*(1), 13-18.

Friend, A., DeFries, J. C., & Olson, R. K. (2008). Parental education moderates genetic influences on reading disability. *Psychological Science*, *19*(11), 1124-1130.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177-182.

Gabay, Y., Schiff, R., & Vakil, E. (2012a). Dissociation between online and offline learning in developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *34*(3), 279-288.

Gabay, Y., Schiff, R., & Vakil, E. (2012b). Dissociation between the procedural learning of letter names and motor sequences in developmental dyslexia. *Neuropsychologia*, *50*(10), 2435-2441.

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934-945.

Gabay, Y., Vakil, E., Schiff, R., & Holt, L. L. (2015). Probabilistic category learning in developmental dyslexia: Evidence from feedback and paired-associate weather prediction tasks. *Neuropsychology*, *29*(6), 844.

Gabriel, A., Maillart, C., Stefaniak, N., Lejeune, C., Desmottes, L., & Meulemans, T. (2013). Procedural learning in specific language impairment: effects of sequence complexity. *Journal of the International Neuropsychological Society*, *19*(3), 264-271.

Gabriel, A., Maillart, C., Guillaume, M., Stefaniak, N., & Meulemans, T. (2011). Exploration of serial structure procedural learning in children with language impairment. *Journal of the International Neuropsychological Society*, *17*(02), 336-343.

Gabriel, A., Meulemans, T., Parisse, C., & Maillart, C. (2015). Procedural learning across modalities in French-speaking children with specific language impairment. *Applied Psycholinguistics*, *36*(3), 747-769.

Gabriel, A., Stefaniak, N., Maillart, C., Schmitz, X., & Meulemans, T. (2012). Procedural visual learning in children with specific language impairment. *American Journal of Speech-Language Pathology*, *21*(4), 329-341.

Gabrieli, J. D. (2009). Dyslexia: a new synergy between education and cognitive neuroscience. *Science*, *325*(5938), 280-283.

Gabrieli, J. D., Corkin, S., Mickel, S. F., & Growdon, J. H. (1993). Intact acquisition and long-term retention of mirror-tracing skill in Alzheimer's disease and in global amnesia. *Behavioral Neuroscience*, *107*(6), 899.

Gagnon, S., Foster, J., Turcotte, J., & Jongenelis, S. (2004). Involvement of the hippocampus in implicit learning of supra-span sequences: The case of SJ. *Cognitive Neuropsychology*, *21*(8), 867-882.

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, *9*(10), 1213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language*, *29*(3), 336-360.

Gathercole, S. E., Tiffany, C., Briscoe, J., & Thorn, A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of child Psychology and Psychiatry*, *46*(6), 598-611.

Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills*, *79*(1), 55-64.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 34.

Geschwind, N., & Levitsky, W. (1968). Human brain: left-right asymmetries in temporal speech region. *Science*, *161*(3837), 186-187.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*(6), 408-418.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.

Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry*, *57*(2), 180-187.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*(1), 1-50.

Gori, S., Seitz, A. R., Ronconi, L., Franceschini, S., & Facoetti, A. (2016). Multiple causal links between magnocellular–dorsal pathway deficit and developmental dyslexia. *Cerebral Cortex*, *26*(11), 4356-4369.

Goswami, U., & Bryant, P. (1989). The interpretation of studies using the reading level design. *Journal of Reading Behavior*, *21*(4), 413-424.

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, *19*(9), 524-533.

Goujon, A., & Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (Papio papio). *Behavioural Brain Research*, *247*, 101-109.

Gould, J. H., & Glencross, D. J. (1990). Do children with a specific reading disability have a general serial-ordering deficit?. *Neuropsychologia*, *28*(3), 271-278.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 501.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 164.

Graham, K. S., Barense, M. D., & Lee, A. C. (2010). Going beyond LTM in the MTL: a synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, *48*(4), 831-853.

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2009). The role of the basal ganglia in learning and memory: neuropsychological studies. *Behavioural Brain Research*, *199*(1), 53-60.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*(2), 129-151.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197.

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships*, *6*(4), 505-518.

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology*, *94*(1), 34.

Guérard, K., Saint-Aubin, J., Boucher, P., & Tremblay, S. (2011). The role of awareness in anticipation and recall performance in the Hebb repetition paradigm: implications for sequence learning. *Memory & Cognition*, *39*(6), 1012-1022.

Habib, M. (2000). The neurological basis of developmental dyslexia: an overview and working hypothesis. *Brain*, *123*(12), 2373-2399.

Hampson, M., Tokoglu, F., Sun, Z., Schafer, R. J., Skudlarski, P., Gore, J. C., & Constable, R. T. (2006). Connectivity–behavior analysis reveals that functional connectivity between left BA39 and Broca's area varies with reading ability. *Neuroimage*, *31*(2), 513-519.

Hartman, M., Knopman, D. S., & Nissen, M. J. (1989). Implicit learning of new verbal associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1070.

Hatcher, P. J., 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*(1), 41-57.

Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., ... & Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *47*(8), 820-827.

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry*, *45*(2), 338-358.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64.

Hayman, C. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 228-240.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. *Brain Mechanisms and Learning*, 37-46.

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing*, *12*(1), 99-128.

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., ... & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities*, *32*(6), 2362-2375.

Hedenius, M. (2013). Procedural and declarative memory in children with developmental disorders of language and literacy (Doctoral thesis, Uppsala University, Sweden). Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A638188&dswid=-5383

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

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1-21.

Henderson, L. M., & Warmington, M. (2017). A sequence learning impairment in dyslexia? It depends on the task. *Research in Developmental Disabilities*, *60*, 198-210.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews. Neuroscience*, *11*(7), 523.

Hensler, B. S., Schatschneider, C., Taylor, J., & Wagner, R. K. (2010). Behavioral genetic approach to the study of dyslexia. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *31*(7), 525.

Hick, R., Botting, N., & Conti-Ramsden, G. (2005). Cognitive abilities in children with specific language impairment: consideration of visuo-spatial skills. *International Journal of Language & Communication Disorders*, *40*(2), 137-149.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149-171.

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, *61*(1), 97-111.

Howard Jr, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging*, *12*(4), 634.

Howard, J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia*, *44*(7), 1131-1144.

Howes, N. L., Bigler, E. D., Burlingame, G. M., & Lawson, J. S. (2003). Memory performance of children with dyslexia: A comparative analysis of theoretical perspectives. *Journal of Learning Disabilities*, *36*(3), 230-246.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352-365.

Hulme, C., Brigstocke, S., & Moll, K. (2016). *Test of Basic Arithmetic and Numeracy Skills*. Oxford University Press.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685-701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology-Learning Memory and Cognition*, *23*(5), 1217-1232.

Hulme, C., & Snowling, M. J. (2009). *Developmental Disorders of Language Learning and Cognition*. John Wiley & Sons.

Hulme, C., & Roodenrys, S. (1995). Practitioner review: Verbal working memory development and its disorders. *Journal of Child Psychology and Psychiatry*, *36*(3), 373-398.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640-648.

Irannejad, S., & Savage, R. (2012). Is a cerebellar deficit the underlying cause of reading disabilities?. *Annals of Dyslexia*, *62*(1), 22-52.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*(3), 366-371.

Ise, E., & Schulte-Körne, G. (2012). Implizites Lernen und LRS: Spielen Defizite im impliziten Lernen eine Rolle bei der Entstehung von Schwierigkeiten im Lesen und Rechtschreiben?. *Lernen und Lernstörungen*.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and language*, *30*(5), 513-541.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306.

James, W. (1890). The Principles of. *Psychology*, *2*.

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated?. *Cortex*, *49*(8), 2001-2006.

Jernigan, T. L., Hesselink, J. R., Sowell, E., & Tallal, P. A. (1991). Cerebral structure on magnetic resonance imaging in language-and learning-impaired children. *Archives of Neurology*, *48*(5), 539-545.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology: Section A*, *54*(4), 1105-1124.

Jiménez, L. (Ed.). (2003). *Attention and implicit learning* (Vol. 48). John Benjamins Publishing.

Jiménez, J. E., Rodríguez, C., & Ramírez, G. (2009). Spanish developmental dyslexia: Prevalence, cognitive profile, and home literacy experiences. *Journal of Experimental Child Psychology*, *103*(2), 167-185.

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

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, *27*(3), 443-463.

Johnstone, T., Shanks, D., & Neely, James H. (1999). *Two Mechanisms in Implicit Artificial Grammar Learning? Comment on Meulemans and Van Der Linden (1997), 25*(2), 524-531.

Justus, T. (2006). The cerebellum and English grammatical morphology: evidence from production, comprehension, and grammaticality judgments. *The Cerebellum*, *16*(7).

Kahta, S., & Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia*, *66*(2), 235-250.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2), 199-225.

Kail, R., Hale, C.A., Leonard, L.B., & Nippold, M.A. (1984). Lexical storage and retrieval in language-impaired children. *Applied Psycholinguistics, 5,* 37-49.

Kaplan, B. J., Dewey, D. M., Crawford, S. G., & Wilson, B. N. (2001). The term comorbidity is of questionable value in reference to developmental disorders data and theory. *Journal of Learning Disabilities*, *34*(6), 555-565.

Karmiloff-Smith, A. (1992). Nature, Nurture and PDP: Preposterous Developmental Postulates?. *Connection Science*, *4*(3-4), 253-269.

Kaufmann, W. E., & Galaburda, A. M. (1989). Cerebrocortical microdysgenesis in neurologically normal subjects A histopathologic study. *Neurology*, *39*(2), 238-238.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321-340.

Keele, S. W., & Ivry, R. (1990). Does the cerebellum provide a common computation for diverse tasks? A timing hypothesis. *Annals of the New York Academy of Sciences*, *608*(1), 179-211.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316.

Kelly, D. J. (2010). Language acquisition challenges for preschoolers residing in low-SES households: Implications for speech-language pathologists and developmental researchers. *SIG 1 Perspectives on Language Learning and Education*, *17*(2), 41-48.

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

Kelly, S. W., Jahanshahi, M., & Dirnberger, G. (2004). Learning of ambiguous versus hybrid sequences by patients with Parkinson's disease. *Neuropsychologia*, *42*(10), 1350-1357.

Kemény, F., & Lukács, Á. (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 249-258.

Kibby, M. Y. (2009). Memory functioning in developmental dyslexia: an analysis using two clinical memory measures. *Archives of Clinical Neuropsychology*, *24*(3), 245-254.

Kibby, M. Y., & Cohen, M. J. (2008). Memory functioning in children with reading disabilities and/or attention deficit/hyperactivity disorder: a clinical investigation of their working memory and long-term memory functioning. *Child Neuropsychology*, *14*(6), 525-546.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171.

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184-193.

Kidd, E., & Kirjavainen, M. (2011). Investigating the contribution of procedural and declarative memory to the acquisition of past tense morphology: Evidence from Finnish. *Language and Cognitive Processes*, *26*(4-6), 794-829.

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review*, *110*(4), 728.

Kirchner, D.M., & Klatzky, R.L. (1985). Verbal rehearsal and memory in language-disordered children. *Journal ofSpeech and Hearing Research, 28,* 556-565.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in Huntington's disease: Evidence from the serial reaction time task. *Neuropsychologia*, *29*(3), 245-254.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*(3), 172-179.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106-120.

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*(1), 169.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399.

Korsakoff, S. S. (1889). Etude médico-psychologique sur une forme des maladies de la mémoire. *Revue Philosophique*, *28*, 501-530.

Kramer, J. H., Knee, K., & Delis, D. C. (2000). Verbal memory impairments in dyslexia. *Archives of Clinical Neuropsychology*, *15*(1), 83-93.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2016). Neurobiological basis of language learning difficulties. *Trends in Cognitive Sciences*, *20*(9), 701-714.

Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2006). Contextual cuing by global features. *Attention, Perception, & Psychophysics*, *68*(7), 1204-1216.

Kuppuraj, S., Rao, P., & Bishop, D. V. (2016). Declarative capacity does not trade-off with procedural capacity in children with specific language impairment. *Autism & Developmental Language Impairments*, *1*, 2396941516674416.

Laasonen, M., Väre, J., Oksanen-Hennah, H., Leppämäki, S., Tani, P., Harno, H., ... & Cleeremans, A. (2014). Project DyAdd: Implicit learning in adult dyslexia and ADHD. *Annals of Dyslexia*, *64*(1), 1-33.

Laforce, R., & Doyon, J. (2001). Distinct contribution of the striatum and cerebellum to motor learning. *Brain and Cognition*, *45*(2), 189-211.

Lahey, M., Edwards, J., & Munson, B. (2001). Is processing speed related to severity of language impairment?. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1354-1361.

Lamy, D., Goshen-Kosover, A., Aviani, N., Harari, H., & Levkovitz, H. (2008). Implicit memory for spatial context in depression and schizophrenia. *Journal of Abnormal Psychology*, *117*(4), 954.

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. *BMJ: British Medical Journal*, *333*(7568), 597.

Lawrence, A. D., Hodges, J. R., Rosser, A. E., Kershaw, A., Ffrench-Constant, C., Rubinsztein, D. C., ... & Sahakian, B. J. (1998). Evidence for specific cognitive deficits in preclinical Huntington's disease. *Brain: A Journal of Neurology*, *121*(7), 1329-1341.

Lee, J. C., Mueller, K. L., & Tomblin, J. B. (2016). Examining procedural learning and corticostriatal pathways for individual differences in language: testing endophenotypes of DRD2/ANKK1. *Language, Cognition and Neuroscience*, *31*(9), 1098-1114.

Lee, J. C., & Tomblin, J. B. (2015). Procedural learning and individual differences in language. *Language Learning and Development*, *11*(3), 215-236.

Leggio, M. G., Tedesco, A. M., Chiricozzi, F. R., Clausi, S., Orsini, A., & Molinari, M. (2008). Cognitive sequencing impairment in patients with focal or atrophic cerebellar damage. *Brain*, *131*(5), 1332-1343.

Leonard, L. B. (2014). *Children with specific language impairment*. MIT press.

Leonard, C. M., & Eckert, M. A. (2008). Asymmetry and dyslexia. *Developmental Neuropsychology*, *33*(6), 663-681.

Leonard, C. M., Eckert, M. A., Lombardino, L. J., Oakland, T., Kranzler, J., Mohr, C. M., ... & Freeman, A. (2001). Anatomical risk factors for phonological dyslexia. *Cerebral Cortex*, *11*(2), 148-157.

Leonard, C. M., Voeller, K. K., Lombardino, L. J., Morris, M. K., Hynd, G. W., Alexander, A. W., ... & Agee, O. F. (1993). Anomalous cerebral structure in dyslexia revealed with magnetic resonance imaging. *Archives of Neurology*, *50*(5), 461-469.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*(1), 24-37.

Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M., & Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nature Neuroscience*, *6*(11), 1230.

Logan, J. A., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child Development*, *84*(6), 2131-2144.

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology*, *49*(3), 150.

Lovegrove, W. (1991). Spatial frequency processing in dyslexic and normal readers. *Vision and Visual Dysfunction: Vision and Visual Dyslexia*, *13*, 148-154.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4-to 12-year-old children. *Developmental Neuropsychology*, *22*(3), 595-624.

Lukács, Á., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, *28*(3), 472.

Lum, J. A., & Bleses, D. (2012). Declarative and procedural memory in Danish speaking children with specific language impairment. *Journal of Communication Disorders*, *45*(1), 46-58.

Lum, J. A., & Conti-Ramsden, G. (2013). Long-term memory: A review and meta-analysis of studies of declarative and procedural memory in specific language impairment. *Topics in Language Disorders*, *33*(4), 282.

Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, *51*, 1-10.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138-1154.

Lum, J. A., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and declarative memory in children with and without specific language impairment. *International Journal of Language & Communication Disorders*, *45*(1), 96-107.

Lum, J. A., & Kidd, E. (2012). An examination of the associations among multiple memory systems, past tense, and vocabulary in typically developing 5-year-old children. *Journal of Speech, Language, and Hearing Research*, *55*(4), 989-1006.

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities*, *34*(10), 3460-3476.

Lund, T. (1988). Some metrical issues with meta-analysis of therapy effects. *Scandinavian Journal of Psychology*, *29*(1), 1-8.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19.

Mainela-Arnold, E., & Evans, J. L. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*(4), 897-909.

Mainela-Arnold, E., & Evans, J. L. (2014). Do statistical segmentation abilities predict lexical-phonological and lexical-semantic abilities in children with and without SLI?. *Journal of Child Language*, *41*(2), 327-351.

313

Majerus, S., Leclercq, A. L., Grossmann, A., Billard, C., Touzin, M., Van der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: No evidence for a causal association. *Cortex*, *45*(6), 708-720.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252.

Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus*, *11*(6), 776-782.

Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., da Guarda, S. N. F., Gerwig, M., ... & Molinari, M. (2012). Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. *The Cerebellum*, *11*(2), 457-487.

Mareschal, D., Butterworth, B., & Tolmie, A. (Eds.). (2013). *Educational neuroscience*. John Wiley & Sons.

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C. H., Beaton, A., Desmond, J., ... & Leggio, M. (2014). Consensus paper: language and the cerebellum: an ongoing enigma. *The Cerebellum*, *13*(3), 386-410.

Marien, P., Engelborghs, S., Fabbro, F., & De Deyn, P. P. (2001). The lateralized linguistic cerebellum: a review and a new hypothesis. *Brain and Language*, *79*(3), 580-600.

Marsolek, C. J., & Bowers, J. S. (Eds.). (2003). *Rethinking implicit memory*. Oxford University Press.

Matthews, G., Warm, J. S., Reinerman-Jones, L. E., Langheim, L. K., Washburn, D. A., & Tripp, L. (2010). Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *Journal of Experimental Psychology: Applied*, *16*(2), 187.

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual Review of Neuroscience*, *23*(1), 649-711.

Martin, N., Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*(5), 641-682.

Mayor-Dubois, C., Zesiger, P., Van der Linden, M., & Roulet-Perez, E. (2014). Nondeclarative learning in children with specific language impairment:

predicting regularities in the visuomotor, phonological, and cognitive domains. *Child Neuropsychology*, *20*(1), 14-22.

McAnally, K. I., & Stein, J. F. (1997). Scalp potentials evoked by amplitude-modulated tones in dyslexia. *Journal of Speech, Language, and Hearing Research*, *40*(4), 939-945.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry*, *41*(7), 869-874.

McCoach, D. B., Black, A. C., & O'Connell, A. A. (2007). Errors of inference in structural equation modeling. *Psychology in the Schools*, *44*(5), 461-470.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills. *Journal of Experimental Child Psychology*, *58*(1), 112-133.

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., ... & Olson, R. K. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry*, *52*(5), 547-557.

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *The Journal of General Psychology*, *114*(1), 75-88.

Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, *23*(10), 1092-1098.

Melby-Lervag, M., Lyster, S., & Hulme, C. (2012) Phonological Skills and their Role in Learning to Read: A Meta-Analytic Review. *Psychological Bulletin, 138,* 322-352

Menghini, D., Carlesimo, G. A., Marotta, L., Finzi, A., & Vicari, S. (2010). Developmental dyslexia and explicit long-term memory. *Dyslexia*, *16*(3), 213-225.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., ... & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. *Neuropsychologia*, *48*(4), 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*(4), 1218-1226.

Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., & Vicari, S. (2008). Structural correlates of implicit learning deficits in subjects with developmental dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 212-221.

Merrill, E. C., Conners, F. A., Roskos, B., Klinger, M. R., & Klinger, L. G. (2013). Contextual cueing effects across the lifespan. *The Journal of Genetic Psychology*, *174*(4), 387-402.

Merrill, E. C., Conners, F. A., Yang, Y., & Weathington, D. (2014). The acquisition of contextual cueing effects by persons with and without intellectual disability. *Research in Developmental Disabilities*, *35*(10), 2341-2351.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*, 3.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*(3), 199-221.

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*(2), 416-433.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215-234.

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

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*.

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding?. *Journal of Experimental Child Psychology*, *64*(2), 199-231.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386.

Moscovitch, M. (1985). Memory from Infancy to Old Age: Implications for Theories of Normal and Pathological Memorya. *Annals of the New York Academy of Sciences*, *444*(1), 78-96.

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *The Quarterly Journal of Experimental Psychology*, *61*(4), 505-514.

Murphy, R. A., Mondragón, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, *319*(5871), 1849-1851.

Muthén, B. O., & Muthén, L. (1998-2016). Mplus user's guide (7th ed). *Los Angeles, CA: Muthén & Muthén*.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, *73*(2), 139-158.

Neale, M., Christophers, Una, & Whetton, Chris. (1988). *Neale analysis of reading ability* (Revised British ed.). Windsor: NFER-NELSON.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, *31*(5), 723.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*(3), 505-511.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85-117.

Nichols, S., Jones, W., Roman, M. J., Wulfeck, B., Delis, D. C., Reilly, J., & Bellugi, U. (2004). Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain and Language*, *88*(2), 180-189.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research?. *Cognition*, *35*(2), 159-182.

Nicolson, R. I., & Fawcett, A. J. (1999). Developmental dyslexia: The role of the cerebellum. *Dyslexia*, *5*(3), 155-177.

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

Nicolson, R. I., & Fawcett, A. J. (2010). *Dyslexia, learning, and the brain*. MIT press.

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

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: the cerebellar deficit hypothesis. *Trends in Neurosciences*, *24*(9), 508-511.

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

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

Nissen, M. J., Willingham, D., & Hartman, M. (1989). Explicit and implicit remembering: When is learning preserved in amnesia?. *Neuropsychologia*, *27*(3), 341-352.

Norbury, C.F., Bishop, D. V. M., & Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module? *Applied Psycholinguistics. 23,* 247-268.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247-1257.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd edition). MacGraw-Hill, New York.

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*.

Ofen, N., Kao, Y. C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S., & Gabrieli, J. D. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience*, *10*(9), 1198-1205.

Olson, R. K., Hulslander, J., Christopher, M., Keenan, J. M., Wadsworth, S. J., Willcutt, E. G., ... & DeFries, J. C. (2013). Genetic and environmental influences on writing and their relations to language and reading. *Annals of Dyslexia*, *63*(1), 25-43.

Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *Journal of Neuroscience*, *26*(17), 4596-4601.

Ostergaard, A. L. (1992). A method for judging measures of stochastic dependence: Further comments on the current controversy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 413.

Ostergaard, A. L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory & Cognition*, *26*(1), 40-60.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin*, *82*(1), 85.

Oyler, J. D., Obrzut, J. E., & Asbjornsen, A. E. (2012). Verbal learning and memory functions in adolescents with reading disabilities. *Learning Disability Quarterly*, *35*(3), 184-195.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. *Journal of Neuroscience*, *9*(5), 1465-1472.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*(1), 563-593.

Page, M. P., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Page, M. P., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506-526.

Page, M. P., & Norris, D. (2008). Is there a common mechanism underlying word-form learning and the Hebb repetition effect? Experimental data and a modelling framework. *Interactions between short-term and long-term memory in the verbal domain*, 136-156.

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J. S., & Hallett, M. (1993). Procedural learning in Parkinson's disease and cerebellar degeneration. *Annals of Neurology*, *34*(4), 594-602.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R. S., & Frith, C. D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain*, *119*(1), 143-157.

Pavlidou, E. V., & Williams, J. M. (2010). Developmental dyslexia and implicit learning: evidence from an AGL transfer study. *Procedia-Social and Behavioral Sciences*, *2*(2), 3289-3296.

Pavlidou, E. V., & Williams, J. M. (2014). Implicit learning and reading: Insights from typical children and children with developmental dyslexia using the artificial grammar learning (AGL) paradigm. *Research in Developmental Disabilities*, *35*(7), 1457-1472.

Pavlidou, E. V., Kelly, L. M., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning?. *Dyslexia*, *16*(2), 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*(1), 55-77.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413.

Perez, T. M., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia*, *18*(2), 94-109.

Perlant, A. S., & Largy, P. (2011). Are implicit learning abilities sensitive to the type of material to be processed? Study on typical readers and children with dyslexia. *Journal of Research in Reading*, *34*(3), 298-314.

Pernet, C., Andersson, J., Paulesu, E., & Demonet, J. F. (2009). When all hypotheses are right: a multifocal account of dyslexia. *Human Brain Mapping*, *30*(7), 2278-2292.

Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 785-800.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology*, *22*(4), 493-516.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General*, *119*(3), 264.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263.

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307.

Petrill, S. A., Deater-Deckard, K., Schatschneider, C., & Davis, C. (2005). Measured environmental influences on early reading: Evidence from an adoption study. *Scientific Studies of Reading*, *9*(3), 237-259.

Pinker, S. (1994). The language instinct. New York: William Morrow and Company.

Plante, E., Bahl, M., Vance, R., & Gerken, L. (2010). Children with specific language impairment show rapid, implicit learning of stress assignment rules. *Journal of Communication Disorders*, *43*(5), 397-406.

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, *35*(5), 453-462.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56.

Plomin, R., Haworth, C. M., Meaburn, E. L., Price, T. S., Wellcome Trust Case Control Consortium 2, & Davis, O. S. (2013). Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science*, *24*(4), 562-568.

Poldrack, R. A. (1996). On testing for stochastic dissociations. *Psychonomic Bulletin & Review*, *3*(4), 434-448.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546-550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews*, *32*(2), 197-205.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245-251.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, *13*(4), 564.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *Journal of Cognitive Neuroscience*, *13*(5), 687-697.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25-42.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological bulletin*, *133*(2), 227.

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

Powell, R. P., & Bishop, D. V. (1992). Clumsiness and perceptual problems in children with specific language impairment. *Developmental Medicine & Child Neurology*, *34*(9), 755-765.

Preacher, K. J. (2015). Extreme groups designs. In R. L. Cautin & S. O. Lilienfeld (Eds.), The encyclopedia of clinical psychology (Vol. 2, pp. 1189-1192). Hoboken, NJ: John Wiley & Sons, Inc.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178.

Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2010). The effects of mood, cognitive style, and cognitive ability on implicit learning. *Learning and Individual Differences*, *20*(3), 215-219.

Price, A. L. (2009). Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task. *Memory & Cognition*, *37*(2), 210-222.

Putzke, J. D., Williams, M. A., Adams, W., & Boll, T. J. (1998). Does memory test performance in children become more consistent with age? Cross-sectional comparisons using the WRAML. *Journal of Clinical and Experimental Neuropsychology*, *20*(6), 835-845.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Rabiner, D., Coie, J. D., & Conduct Problems Prevention Research Group. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(7), 859-867.

Ramus, F. (2014). Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends in Cognitive Sciences*, *18*(6), 274-275.

Ramus, F. (2004). Neurobiology of dyslexia: a reinterpretation of the data. *TRENDS in Neurosciences*, *27*(12), 720-726.

Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model. *Brain*, *136*(2), 630-645.

Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry*, *44*(5), 712-722.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short-and long-term memory. *Trends in Cognitive Sciences*, *9*(8), 374-380.

Raschle, N. M., Chang, M., & Gaab, N. (2011). Structural brain alterations associated with dyslexia predate reading onset. *Neuroimage*, *57*(3), 742-749.

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size?. *Journal of Consulting and Clinical Psychology*, *64*(6), 1316.

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

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219.

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

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. *Advances in Psychology*, *139*, 113-123.

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*(10), 2026-2042.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, *15*(4), 574-583.

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*(5), 888.

Rekkas, P. V., & Constable, R. T. (2005). Evidence that autobiographic memory retrieval does not become independent of the hippocampus: an fMRI study contrasting very recent with remote events. *Journal of Cognitive Neuroscience*, *17*(12), 1950-1961.

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech, Language, and Hearing Research*, *38*(4), 850-863.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1239-1257.

Richlan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping*, *30*(10), 3299-3308.

Rimrodt, S. L., Peterson, D. J., Denckla, M. B., Kaufmann, W. E., & Cutting, L. E. (2010). White matter microstructural differences linked to left perisylvian language network in children with dyslexia. *Cortex*, *46*(6), 739-749.

Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning?. *Journal of Neuroscience*, *27*(38), 10073-10075.

Rochelle, K. S., & Talcott, J. B. (2006). Impaired balance in developmental dyslexia? A meta-analysis of the contending evidence. *Journal of Child Psychology and Psychiatry*, *47*(11), 1159-1166.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. *Handbook of Neuropsychology*, *8*, 63-63.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*(3), 726.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability the difference score in the measurement of change. *Journal of Educational Measurement*, *20*(4), 335-343.

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

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217-236.

Rüsseler, J., Gerth, I., & Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology*, *28*(5), 808-827.

Rutter, M. (2004). Pathways of genetic influences on psychopathology. *European Review*, *12*(1), 19-33.

Ryle, G. (1984). The concept of mind (1949). *London: Hutchinson*.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110-114.

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*(2), 101-105.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 1926-1928.

Salthouse, T. A., McGuthry, K. E., & Hambrick, D. Z. (1999). A framework for analyzing and interpreting differential aging patterns: Application to three measures of implicit learning. *Aging, Neuropsychology, and Cognition*, *6*(1), 1-18.

Schacter, D. L. (1994). In DL Schacter & E. Tulving. In Memory Systems. MIT Press. pp. 233--268.

Schacter, D. L., & Moscovitch, M. (1984). Infants, amnesics, and dissociable memory systems. In *Infant memory* (pp. 173-216). Springer US.

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3-8.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736-1747.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*(6), 1013-1025.

Schmalz, X., Altoè, G., & Mulatti, C. (2016). Statistical learning and dyslexia: a systematic review. *Annals of Dyslexia*, 1-16.

Schneider, W. (2010). Memory development in childhood. *The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition*, 347-376.

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. *Handbook of Metamemory and Memory*, *14*, 391-409.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research*, *61*(3), 175-190.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163.

Seger, C. A. (2006). The basal ganglia in human learning. *The Neuroscientist*, *12*(4), 285-290.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203-219.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523.

Selst, M. V., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, *47*(3), 631-650.

Semel, E., Wiig, E., Secord, W., & Sabers, D. (1987). CELF-R: Clinical evaluation of language functions-revised (technical manual). *New York: Psychological Corporation*.

Sengottuvel, K., & Rao, P. K. (2013). Aspects of grammar sensitive to procedural memory deficits in children with specific language impairment. *Research in Developmental Disabilities*, *34*(10), 3317-3331.

Sengottuvel, K., & Rao, P. K. (2014). Sequence learning pattern in children with specific language impairment. *International Journal on Disability and Human Development*, *13*(1), 55-62.

Shanks, D. R., 2005. Implicit learning. In Lamberts, K., & Goldstone, R. L. *Handbook of cognition*, 202-220. London: Sage.

Shanks, D. R., & Channon, S. (2002). Effects of a secondary task on" implicit" sequence learning: learning or performance?. *Psychological Research*, *66*(2), 99-109.

Shanks, D. R., Green, R. E., & Kolodny, J. A. (1994). A critical examination of the evidence for unconscious (implicit) learning. In C. Umilta & M. Moscovitch (Eds.), Attention and performance XV (pp. 837-860). Cambridge, MA: MIT Press.

Shanks, D. R., & John, M. F. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367-395.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1435.

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 248.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145-150.

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 219-236.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B*, *372*(1711), 20160059.

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

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

Siegert, R. J., Taylor, K. D., Weatherall, M., & Abernethy, D. A. (2006). Is implicit sequence learning impaired in Parkinson's disease? A meta-analysis. *Neuropsychology, 20(4)*, 490-495

Sigurdardottir, H. M., Danielsdottir, H. B., Gudmundsdottir, M., Hjartarson, K. H., Thorarinsdottir, E. A., & Kristjánsson, Á. (2017). Problems with visual statistical learning in developmental dyslexia. *Scientific Reports*, *7*(1), 606.

Simmons, F. R., & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia*, *14*(2), 77-94.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666-681.

Skottun, B. C. (2000). The magnocellular deficit theory of dyslexia: the evidence from contrast sensitivity. *Vision Research*, *40*(1), 111-127.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657-1661.

Smith, J., Siegert, R. J., McDowall, J., & Abernethy, D. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. *Brain and Cognition*, *45*(3), 378-391.

Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403-415.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, *43*(2), 219-234.

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia*, *7*(1), 37-46.

Snowling, M. J. (2013). Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs*, *13*(1), 7-14.

Snowling, M., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence?. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 587-600.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(3), 369-373.

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, *57*(12), 1360-1369.

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development*, *74*(2), 358-373.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, *81*(1), 1-23.

Snowling, M. J., Adams, J. W., Bishop, D. V. M., & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language & Communication Disorders*, *36*(2), 173-183.

Song, S. B., Howard, J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory, 14*, 167-176.

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, *28*(4), 467-490.

Sperling, A. J., Lu, Z. L., & Manis, F. R. (2004). Slower implicit categorical learning in adult poor readers. *Annals of Dyslexia*, *54*(2), 281-303.

Squire, L. R. (1987). Memory and brain. New York: Oxford University Press.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171-177.

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: A new hypothesis. *Memory Consolidation: Psychobiology of Cognition*, 185-210.

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*(3), a021667.

Staels, E., & Van den Broeck, W. (2015). No solid empirical evidence for the solid (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 650.

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

Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition*, *21*(6), 819-827.

Stark, C. E., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 945.

Stein, J., Talcott, J., & Walsh, V. (2000). Controversy about the visual magnocellular deficit in developmental dyslexics. *Trends in Cognitive Sciences*, *4*(6), 209-210.

Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-Impaired PreschoolersA Follow-Up Into Adolescence. *Journal of Speech, Language, and Hearing Research*, *41*(2), 407-418.

Stoodley, C. J., Harrison, E. P., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, *44*(5), 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*(1), 173-183.

Stoodley, C. J., & Stein, J. F. (2013). Cerebellar function in developmental dyslexia. *The Cerebellum*, *12*(2), 267-276.

Stromswold, K. (1998). Genetics of spoken language disorders. *Human Biology*, 297-324.

Stuebing, K. K., Barth, A. E., Molfese, P. J., Weiss, B., & Fletcher, J. M. (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children*, *76*(1), 31-51.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological Review*, *112*(1), 159.

Szalkowski, C. E., Fiondella, C. F., Truong, D. T., Rosen, G. D., LoTurco, J. J., & Fitch, R. H. (2013). The effects of Kiaa0319 knockdown on cortical and subcortical anatomy in male rats. *International Journal of Developmental Neuroscience*, *31*(2), 116-122.

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *The Quarterly Journal of Experimental Psychology*, *62*(3), 435-443.

Szmalec, A., Loncke, M., Page, M., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1270.

Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1275-1278.

Tallal, P., Sainburg, R. L., & Jernigan, T. (1991). The neuropathology of developmental dysphasia: Behavioral, morphological, and physiological evidence for a pervasive temporal processing disorder. *Reading and Writing*, *3*(3-4), 363-377.

Tallal, P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*, *241*(5390), 468-469.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, *13*(1), 69-74.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of Cognitive Neuroscience*, *16*(8), 1339-1351.

Thomas, K. M., & Nelson, C. A. (2001). Serial reaction time learning in preschool-and school-age children. *Journal of Experimental Child Psychology*, *79*(4), 364-387.

Tijms, J. (2004). Verbal memory and phonological processing in dyslexia. *Journal of Research in Reading*, *27*(3), 300-310.

Tisak, J., & Smith, C. S. (1994). Defending and extending difference score methods. *Journal of Management*, *20*(3), 675-682.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*(4), 269-293.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260.

Torgesen, J. K. (2005). Recent discoveries from research on remedial interventions for children with dyslexia. *The Science of Reading*, 521-537.

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). TOWRE–2 Test of Word Reading Efficiency. *Austin, TX: Pro-Ed*.

Treiman, R. (1985). Phonemic awareness and spelling: Children's judgments do not always agree with adults'. *Journal of Experimental Child Psychology*, *39*(1), 182-201.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374-391.

Tulving, E. (1985). How many memory systems are there?. *American Psychologist*, *40*(4), 385.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*(1), 1-25.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301. (PRS and priming

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 336-342.

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

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature reviews. Neuroscience*, *2*(10), 717.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231-270.

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit?. *Cortex*, *42*(4), 480-485.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*(2), 266-276.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *51*, 205-222.

Vaidya, C. J., Huger, M., Howard, D. V., & Howard Jr, J. H. (2007). Developmental differences in implicit learning of spatial context. *Neuropsychology*, *21*(4), 497.

Vakil, E., Lowe, M., & Goldfus, C. (2015). Performance of Children With Developmental Dyslexia on Two Skill Learning Tasks—Serial Reaction Time and Tower of Hanoi Puzzle: A Test of the Specific Procedural Learning Difficulties Theory. *Journal of Learning Disabilities*, *48*(5), 471-481.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., ... & Gadian, D. G. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences*, *95*(21), 12695-12700.

Vellutino, F. R. (1979). Dyslexia: Theory and research. Cambridge, MA: MIT Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2-40.

Vellutino, F. R., & Scanlon, D. M. (1985). Free recall of concrete and abstract words in poor and normal readers. *Journal of Experimental Child psychology*, *39*(2), 363-380.

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 & Psychiatry*, *76*(10), 1392-1397.

Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia*, *41*(1), 108-114.

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57-63.

Vinter, A., & Perruchet, P. (2000). Implicit learning in children is not related to age: Evidence from drawing behavior. *Child Development*, 1223-1240.

Vlachos, F., Papathanasiou, I., & Andreou, G. (2007). Cerebellum and reading. *Folia Phoniatrica et Logopaedica*, *59*(4), 177-183.

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., & Curran, T. (2003). Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits?. *Journal of Experimental Child Psychology*, *84*(4), 338-354.

Wadsworth, S. J., Olson, R. K., & DeFries, J. C. (2010). Differential genetic etiology of reading difficulties as a function of IQ: an update. *Behavior Genetics*, *40*(6), 751-758.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256.

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, *217*(5132), 972-974.

Warrington, E. K., & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, *12*(4), 419-428.

Warrington, E. K., & Weiskrantz, L. (1970). Amnesic syndrome: Consolidation or retrieval?. *Nature*, *228*(5272), 628-630.

Wasserman, E. A., Teng, Y., & Castro, L. (2014). Pigeons exhibit contextual cueing to both simple and complex backgrounds. *Behavioural Processes*, *104*, 44-52.

Webster, R. I., & Shevell, M. I. (2004). Topical review: Neurobiology of specific language impairment. *Journal of Child Neurology*, *19*(7), 471-481.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Psychological Corporation.

Weismer, S. E., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1249-1260.

Weitz, D., O'Shea, G., Zook, N., & Needham, W. (2011). Working memory and sequence learning in the Hebb digits task: Awareness is predicted by individual differences in operation span. *The American Journal of Psychology*, *124*(1), 49-62.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2017). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*.

Westerberg, C. E., Miller, B. B., Reber, P. J., Cohen, N. J., & Paller, K. A. (2011). Neural correlates of contextual cueing are modulated by explicit learning. *Neuropsychologia*, *49*(12), 3439-3447.

Whitehouse, A. J., & Bishop, D. V. (2008). Cerebral dominance for language function in adults with specific language impairment or autism. *Brain*, *131*(12), 3193-3200.

Whitehouse, A. J., Line, E. A., Watt, H. J., & Bishop, D. V. (2009). Qualitative aspects of developmental language impairment relate to language and literacy outcome in adulthood. *International Journal of Language & Communication Disorders*, *44*(4), 489-510.

Wilkinson, G. S. (1993). *WRAT-3: Wide range achievement test administration manual*. Wide Range, Incorporated.

Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities*, *33*(2), 179-191.

Willcutt, E. G., Pennington, B. F., Boada, R., Ogline, J. S., Tunick, R. A., Chhabildas, N. A., & Olson, R. K. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *110*(1), 157.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1047.

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*(05), 766-780.

Winograd, T. (1975). Frame representations and the declarative/procedural controversy. *Representation and Understanding: Studies in Cognitive Science*, 185-210.

Wise, S. P., Murray, E. A., & Gerfen, C. R. (1996). The frontal cortex-basal ganglia system in primates. *Critical Reviews™ in Neurobiology*, *10*(3-4).

Yang, Y., & Hong-Yan, B. (2011). Unilateral implicit motor learning deficit in developmental dyslexia. *International Journal of Psychology*, *46*(1), 1-8.

Yang, Y., Bi, H. Y., Long, Z. Y., & Tao, S. (2013). Evidence for cerebellar dysfunction in Chinese children with developmental dyslexia: an fMRI study. *International Journal of Neuroscience*, *123*(5), 300-310.

Yap, R. L., & Leij, A. V. D. (1994). Testing the automatization deficit hypothesis of dyslexia via a dual-task paradigm. *Journal of Learning Disabilities*, *27*(10), 660-665.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, *43*(5), 635-645.

Zellin, M., von Mühlenen, A., Müller, H. J., & Conci, M. (2014). Long-term adaptation to change in implicit contextual learning. *Psychonomic Bulletin & Review*, *21*(4), 1073-1079.

Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1986). Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience*, *6*(10), 2950-2967.

Appendices

Appendix A: Search terms for the literature search for the meta-analyses

Database	Search Terms
PsychINFO, Medline via Ovid	<u>(OR between all the terms)</u> Implicit learning (entered as a subject heading) Implicit adj2 learn$* Implicit adj2 memory Procedur$ adj2 learn$ Procedur$ adj2 memory Probabili$ adj2 learn$ Probabili$ adj2 memory Statistic$ adj2 learn$ Statistic$ adj2 memory Sequence adj2 learn$ Serial adj2 learn$ Serial reaction time Hebb$ adj2 learn$ contextual cueing Artificial grammar finite state grammar Weather prediction task AND (OR between all the terms) language disorders (subheading) language development disorders (subheading) specific language impairment (subheading) dyslexia (subheading) Language adj2 impair$ Language adj2 problem$ Language adj2 disorder$ Language adj2 deficit$ Language adj2 difficult$ Language adj2 abilit$ Language adj2 fluen$ Read$ adj2 abilit$ Read$ adj2 fluen$ Read$ adj2 impair$ Read$ adj2 difficult$ Verbal adj2 impair$ Verbal adj2 deficit$ Verbal adj2 abilit$ Phonolog$ adj2 impair$ Phonolog$ adj2 deficit$ Gramma$ adj2 impair$ Gramma$ adj2 deficit$

Appendix B: Correlations between components elements of the implicit learning tasks

Correlations between the individual components of the difference scores for the implicit learning tasks were high (see. High correlations between component scores of the contextual cueing and SRT tasks, however, simply demonstrate that participants who are fast at pressing buttons on one task are fast on the others and are not suggestive of any other commonality between tasks. The RT component totals do not correlate with random or Hebb trial recall component scores, which just means that if participants were fast at pressing buttons they were not necessarily good at recalling sequences and these correlations are not relevant to implicit learning.

Appendix B 1 Correlations for the RT and recall component scores for the implicit tasks

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. CCNV Predictable																
2. CCV Predictable	.73*															
3. CCNV Unpredictable	.91*	.76*														
4. CCV Unpredictable	.74*	.86*	.74*													
5. SRT1 Probable	.62*	.48*	.63*	.48*												
6. SRT2 Probable	.53*	.42*	.54*	.40*	.88*											
7. VRT1 Probable	.63*	.59*	.62*	.60*	.73*	.70*										
8. VRT2 Probable	.56*	.54*	.58*	.47*	.64*	.68*	.81*									
9. SRT1 Improbable	.59*	.45*	.60*	.43*	.95*	.86*	.64*	.55*								
10. SRT2 Improbable	.44*	.29*	.42*	.28	.80*	.90*	.59*	.58*	.81*							
11. VRT1 Improbable	.59*	.54*	.59*	.57*	.72*	.66*	.93*	.78*	.63*	.61*						
12. VRT2 Improbable	.51*	.50*	.51*	.41*	.61*	.65*	.76*	.94*	.54*	.59*	.73*					
13. NVHebb Hebb	-.14	-.02	-.09	-.26	-.11	.05	-.02	.07	-.13	.03	-.08	.06				
14. NV Hebb Random	-.18	-.32	-.19	-.33	-.17	-.05	-.12	-.05	-.21	-.03	-.10	-.04	.62*			
15. V Hebb Hebb	-.15	-.15	-.16	-.14	-.03	-.16	-.11	-.19	-.02	-.24	-.08	-.24	.24	.22		
16. V Hebb Random	-.16	-.08	-.16	-.05	-.10	-.00	-.01	-.03	-.11	-.12	-.04	-.07	.32	.23	.48*	

$*p < .05$