Cognitive Predictors of Individual Differences in Language Acquisition

by

Ashley Rae Blake

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Supervised by:

Professor Ewa Dąbrowska and Professor Dagmar Divjak

Department of Linguistics and Communication

School of English, Drama and Creative Studies

College of Arts and Law

University of Birmingham

July 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

There are vast individual differences in the way in which people acquire, process and use language. While most people learn their native language with ease, some people have difficulty with understanding and producing language in their day-to-day lives, a condition known as Developmental Language Disorder (DLD). Research to date has focused on cognitive and linguistic difficulties associated with DLD, but the underlying causes are not well-understood. A series of empirical studies were conducted with the objective of developing a deeper understanding regarding the interrelationship between cognitive abilities and language in children and adults, with and without DLD. There are two novel approaches to this research, firstly, the speed of automatization is explored using the Multiple-trial Tower of Hanoi paradigm, showing that automatization is relevant to language in children with DLD, and LT and DLD adults. Secondly, this thesis investigates the role of (foreign) language aptitude/analytic ability, with results suggesting that language aptitude is relevant to linguistic outcomes in language typical (LT) children and adults, with and without DLD. The overall results presented in this programme of research suggest that specific cognitive abilities are important for language. In children with DLD (study one), a cumulative risk model is presented, demonstrating the combinatory effects of nonverbal IQ, working memory, and speed of automatization. Study two investigates language and cognition in adulthood, showing that language aptitude is relevant to linguistic outcomes in adults with and without DLD. There are small to no differences in other cognitive abilities assessed in this study, yet the results show that adults with DLD require more time to process linguistic and non-linguistic tasks. Study three demonstrates that analogical reasoning is a key predictor of language in LT children and children with DLD. Lastly, a key relationship emerges between nonverbal intelligence and grammar, challenging the widely accepted view that grammar develops purely implicitly.

Language difficulties have been studied since the 19th century, and although great strides have been made in gaining a deeper understanding of DLD, there is still much to learn. Collectively, the results from the studies detailed in this thesis provide evidence that individuals with DLD process language differently to their LT peers and show that certain cognitive abilities support language in DLD.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Ewa Dąbrowska, Professor Dagmar Divjak, and Dr Nick Riches, for your constant expertise, guidance, and support throughout my PhD journey. Thank you especially to Ewa for all you have taught me, for your time, unwavering energy, and enthusiasm – you constantly foster a love of language and learning in our lab and it is infectious!

I am grateful for scholarship funding from an Alexander von Humboldt Professorship grant, awarded to Professor Ewa Dąbrowska (grant number ID-1195918). Thank you to the schools and organisations that supported my work and special thanks to the many children and adults that took part in my studies, without you this doctoral research would not have been possible. It has been a true privilege to learn about DLD and I hope to continue on this journey in the future.

I have been fortunate to share this journey with a wider research community and learn alongside wonderful colleagues and friends in the Language and Cognition lab at Friedrich-Alexander-Universität, Erlangen-Nürnberg. It has been a rewarding few years – special mention goes to Richenda, Magda, Leonarda, Panos, Armine, Miquel, and Elodie. Thank you also to my colleagues in the department of Linguistics and Communication at the University of Birmingham, and also to the Out of Our Minds team for adopting me and always making me feel welcome.

Thank you (more than I can express) to my mother, Jan, my siblings and family, and to many close friends who have supported me over the past few years. This thesis is for my husband, Dean, whose support and belief has made it all possible. It is also for my children, J, C, D and M, and in memory of my father, Mike Michell, with whom I have always wished to share this journey.

List of papers and publications

This thesis incorporates several manuscripts that have either been published or will be submitted to a scientific journal. These are detailed below, together with information regarding author contributions.

Published book chapter:

Dąbrowska, E., & Blake, A. (2022). Speed of automatization predicts performance on "decorative" grammar in second language learning. In Piske, T., & Steinlen, A.K. (Eds.) (2022). *Cognition and second language acquisition: Studies on pre-school, primary school and secondary school children*. Narr Francke Attempto. pp. 285-309

(Permission to include this chapter has been granted from the editors and publisher)

Author contribution:
ED: conceptualization, writing: original draft outline, editing and reviewing, data analysis, responding to reviewers comments.
AB: conceptualization, writing: original draft from outline, editing and reviewing, data analysis, responding to reviewers comments.

Articles submitted to scientific journals (or in preparation for submission):

The following articles are in preparation for submission to academic journals:

Study one:
Blake, A., Dąbrowska, E, and Riches, N. (under review). Cognitive predictors of children's linguistic abilities.

Blake, A., Dąbrowska, E, and Riches, N. (under review). Exploring cognitive predictors of language in children with developmental language disorder: the case for nonverbal IQ, working memory, implicit statistical learning, and the speed of automatization.

Author contribution (Study 1):

ED, NR: conceptualization, writing: reviewing and editing, guidance on data analysis. ED: supervision.

AB: conceptualization, design of online task, data collection, transcription and analyses of narrative data, writing: initial draft, reviewing and editing, data processing and analysis.

Study two:

Blake, A., Dąbrowska, E, and Llompart, M. (under review). Towards an understanding of developmental language disorder in adults: investigating the relationship between cognitive abilities and linguistic outcomes.

Author contribution: (Study 2)

ED, ML: conceptualization, writing: reviewing and editing, guidance on data analysis. ED: supervision.

AB: conceptualization, design of online tasks, data collection, writing: initial draft, reviewing and editing, data processing and analyses.

Study three:

Blake, A., Wright, R., Dąbrowska, E, Riches, N, and Winckel, E. (in preparation). Non-verbal cognition as a predictor of linguistic outcomes in children with and without developmental language disorder. Pre-registered study: (https://osf.io/uc2qa).

Author contribution: (Study 3)

ED, NR, EW: conceptualization, writing: reviewing and editing, guidance on data analyses. ED: supervision.

AB: conceptualization, pre-registration, design of online experimental tasks, data collection, writing: initial draft, reviewing and editing, contribution to data processing and analyses.

RW: conceptualization, pre-registration, design of online experimental tasks, data collection, writing: contribution to initial draft, reviewing and editing, data processing and analyses.

The following journal article was published as part of the research conducted for this thesis. The article is a methodological paper providing an account of online data research during the COVID-19 pandemic. The paper details the methodology described in study one of this thesis and therefore is provided for reference only.

Blake, A., & Dąbrowska, E. (2024). Investigating the relationship between the speed of automatization and linguistic abilities: data collection during the COVID-19 pandemic. *Linguistics Vanguard*. link to paper

Author contribution:
ED: conceptualization, writing: reviewing and editing, supervision, responding to reviewers comments.
AB: conceptualization, writing: initial draft, reviewing and editing, responding to reviewers comments.

Table of Contents

Abstract ii
Acknowledgements iv
List of papers and publications v
Table of Contents viii
List of figures x
List of tables xi
List of abbreviations xii
Thesis overview xiv

PART ONE: LITERATURE REVIEW 1

1.0 LANGUAGE 1

1.1 Usage-based theories of language acquisition 1
1.2 Developmental language disorder (DLD): a definition 6
1.3 Causes of DLD: biology 10
1.3.1 Risk factors for DLD 10
1.3.2 Behavioural genetic research 11
1.3.3 Neurobiological research 13
1.3.4 Transdiagnostic approach to neurodevelopmental disorders 14
1.4 Language characteristics of DLD 15
1.4.1 Word learning and vocabulary 17
1.4.2 Grammar 19
1.4.3 Literacy 20
1.5 Clinical markers of DLD 22
1.6 DLD in adulthood 23
1.6.1 Language and literacy 24
1.6.2 Word learning and memory 25
1.6.3 Education and employment 26

2.0 COGNITIVE ABILITIES 27

2.1 Intelligence / Nonverbal IQ 27
2.2 Language aptitude/language analytic ability 32
2.3 Analogical reasoning 35
2.4 Processing speed 39
2.5 Memory 42
2.5.1 Working memory 42
2.5.2 Declarative and procedural memory 47
2.5.3 Limitations regarding dual-processing models 51
2.6 Implicit learning 54
2.6.1 Implicit learning: procedural learning deficits 55
2.6.2 Implicit learning: statistical learning 62
2.6.3 Evidence regarding implicit learning from meta-analyses 65
2.6.4 Limitations regarding implicit learning 69
3.0 Cognitive skill learning 71
3.1 Skill acquisition theory and automatization 71

PART TWO: EMPIRICAL STUDIES 78

4.0 Introduction to empirical studies and overall research questions 78

4.1 Pilot study: Dąbrowska and Blake (2022) 80
Abstract....81
Introduction 82
Method....93
Results 99
Discussion....104
Conclusion....107
4.2 Study one (a): Blake, Dąbrowska, and Riches (under review). 109
Abstract....110
Introduction 110
Method....120
Results 127
Discussion....137
Conclusion....141
4.3 Study one (b): Blake, Dąbrowska, and Riches (under review). 142
Abstract....143
Introduction 143
Method....154
Results 163
Discussion....174
Conclusion....179
4.4 Study two: Blake, Dąbrowska, and Llompart (under review). 180
Abstract....181
Introduction 182
Method....192
Results 200
Discussion....212
Conclusion....217
4.5 Study three: Blake, Wright, Dąbrowska, Riches, and Winckel (in preparation) 218
Abstract....219
Introduction 220
Method....227
Results 238
Discussion....251
Conclusion....258

5.0 Discussion 260
5.1 Summary of studies....260
5.1.1 Pilot study: Dąbrowska and Blake (2022)....260
5.1.2 Study one: Blake, Dąbrowska, and Riches (1a and 1b: under review) 261
5.1.3 Study two: Blake, Dąbrowska, & Llompart (under review) 262
5.1.4 Study three: Blake, Wright, Dąbrowska, Riches, and Winckel (in preparation)....263
5.2 Overall summary of research findings....264
5.3 Clinical implications 274
5.4 Strengths and limitations....276
5.5 Conclusion 279

References....281
Appendices....356
Appendix 1. Study 2: Pearson's correlation matrix for the LT group....357
Appendix 2. Study 2: Pearson's correlation matrix for the DLD group....358
Appendix 3. Study 2: Questions used to determine DLD status 359
Appendix 4. Response time data 360

List of figures

Figure 1. Illustration of areas of difficulty commonly seen in DLD 16
Figure 2. Physical version of the MToH task ..... **Error! Bookmark not defined.**
Figure 3. Online version of the MToH task ..... **Error! Bookmark not defined.**
Figure 4. Generalisation as a proportional analogy problem 87
Figure 5. The Tower Of Hanoi puzzle 90
Figure 6. An example of an item from the Picture Selection task 97
Figure 7. Average no. of moves per trial 100
Figure 8. Average time per trial 100
Figure 9. ROC curve demonstrating the predictive ability of the logistic regression model 173
Figure 10. Classification accuracy of cognitive predictors 174
Figure 11. Example of a pattern recognition trial in the VSL task (pair recognition) 196
Figure 12. Example of a pattern recognition trial in the VSL task (triplet recognition) 196
Figure 13. Example of a pattern completion trial in the VSL task 197
Figure 14. Multiple-trial Tower of Hanoi task 198
Figure 15. Example of a test item from the Pictures and Sentences task 198
Figure 16. Test item from the Pictures and Sentences task 229
Figure 17. Sample item from the verbal analogical reasoning task. 232
Figure 18. Example of a sample item from the Scene Analogy task 233
Figure 19. Example trials from the Go/No-Go task. 234
Figure 20. Example trial from 'Where's Nellar?' 236
Figure 21. Distribution by group on the language tasks 240
Figure 22. Distribution by group on the cognitive measures 240
Figure 23. Error rates by trial type in the analogical reasoning task 241
Figure 24. Accuracy rates by construction for the receptive grammar task 242
Figure 25. Reaction time on the ISL task by block for all participants 242
Figure 26. Reaction time on the ISL task by block for LT and DLD participants 243
Figure 27. Motor learning distribution 243
Figure 28. Accuracy by collocation type 244

List of tables

Table 1. Example of items used in the Grammaticality Judgment task....94
Table 2. Example of items used in the Picture Selection task....96
Table 3. Descriptive statistics....102
Table 4. Pairwise correlations between all variables....103
Table 5. Descriptive statistics....130
Table 6. Pearson's correlations....132
Table 7. Regression analyses: predictors of linguistic measures....135
Table 8. Means, standard deviations and group comparisons....167
Table 9. Pearson's correlations for the LT group....169
Table 10. Pearson's correlations for the DLD group....170
Table 11. Logistic regression model showing cognitive predictors of group membership....171
Table 12. The number of DLD children falling above or below the median values for the LT group172
Table 13. Confusion matrix....173
Table 14. Examples of constructions in the Pictures and Sentences task....199
Table 15. Examples of constructions in the Grammaticality Judgment task....199
Table 16. Means, standard deviations and group comparisons....204
Table 17. Regression models for the whole sample (LT and DLD groups combined)....209
Table 18. Regression models for the LT group....210
Table 19. Regression models for the DLD group....**Error! Bookmark not defined.**
Table 20. Constructions in the Pictures and Sentences task....230
Table 21. Mean, median, standard deviation (SD), reliability, *t*-tests....239
Table 22. Correlations for the DLD group (n = 34)....245
Table 23. Correlations for the LT group (n = 39)....245
Table 24. Model outputs for Vocabulary....247
Table 25. Model outputs for Receptive Grammar....248
Table 26. Model outputs for Collocations....249
Table 27. Model outputs for Recalling Sentences....250

List of abbreviations

AUC	Area under the curve receiver operating characteristic
AutoCV	Coefficient of variation (measure of the speed of automatization)
AutoDiff	Difference in moves (measure of the speed of automatization)
ACT	Adaptive Control of Thought
ADHD	Attention deficit hyperactivity disorder
AFC	Alternate forced choice
ASD	Autism Spectrum Disorder
ASRT	Alternate serial reaction time task
BPVS	British Picture Vocabulary Scale
CATALISE	Criteria and terminology applied to language impairments: synthesising the evidence
CE	Central executive
CELF	Clinical Evaluation of Language Fundamentals
CLAN	Computerized Language Analysis program
CV	Coefficient of variation
CPM	Coloured Progressive Matrices
DLD	Developmental language disorder
DP	Declarative / Procedural Model
FOC	First-order conditional (sequence)
HM	Henry Molaison
ISL	Implicit statistical learning
GJT	Grammaticality Judgment Task
GNG	Go/No-Go
GWAS	Genome-wide association study
IQ	Intelligence quotient
ISL	Implicit statistical learning
LAssoc1	Length of associative phase 1 (4/5 trails with optimal solution on the MToH)
LAssoc2	Length of associative phase 2 (5/5 trials with optimal solution on the MToH)
LgAnal	Language Analysis
LT	Language typical
MLAT	Modern Language Aptitude Test
MLAT-E	Modern Language Aptitude Test-Elementary

MLS	Manchester Language Study
MLTU	Mean length of utterances in T-units
MRI	Magnetic resonance imaging
MToH	Multiple-trial tower of Hanoi
NAD	Nonadjacent dependencies
NLI	Non-specific language impairment
PDH	The Procedural Circuit Deficit Hypothesis
PLAB	Pimsleur Language Aptitude Battery
PST	Picture Selection Task
ROC	Receiver Operating Characteristics curve
RT	Response time
SLI	Specific Language Impairment
SLP	Speech and language pathologist
SOC	Second-order conditional (sequence)
SRT	Serial reaction time
SICR	Statistically Induced Chunking Recall
TEDS	Twins Early Developmental Study
TROG	Test for the Reception of Grammar
ToH	Tower of Hanoi
VOCD	Vocabulary diversity

Thesis overview

The overall goal of this doctoral research project is to investigate cognitive predictors of individual differences in language acquisition, in children and adults, with and without DLD. This thesis is organised as follows:

Part One comprises a literature review. Firstly, a brief background is provided regarding children's language development according to usage-based theories of language acquisition. Then this section provides an overview of DLD, discussing the terminology, potential underlying causes and presentation of the condition in children and adults. Thereafter, discussion focuses on the cognitive abilities which are examined in this thesis.

Part Two incorporates the academic papers that comprise the thesis. These comprise a pilot study and three empirical studies. As this doctoral research is presented as an 'alternative format thesis', the studies have been written for publication. Because of this, there is unavoidable repetition of theoretical understanding and literature between the papers and the introductory section of this thesis. Reference lists have been removed from each of the manuscripts and amalgamated into a combined reference list at the end of the thesis.

Part Three provides a general discussion and conclusion which integrates the overall findings generated by the research conducted for this doctoral thesis.

PART ONE: LITERATURE REVIEW

1.0 LANGUAGE

1.1 Usage-based theories of language acquisition

Language is a remarkable phenomenon; the ability to use language is part of what makes us human and separates human cognition from that of other species. Yet communication is not unique to the human species, nor is cognition. Many mammalian species and primates use cognitive skills in their daily lives – for example, to enable them to navigate their surroundings, categorize objects and in some cases, using tools to retrieve food (Tomasello, 1999; Tomasello & Herrmann, 2010). Likewise, many animals communicate, yet in animals, communication and cognition tends to serve individual needs, (for example, in primates), whereas humans engage in collaborative activity with a shared purpose (Tomasello, 2023). As such, Tomasello suggests that the fundamental difference between human communication and that of animal species is that humans engage in cultural learning, using linguistic symbols, a system which is a unique 'invention' of the human species (Tomasello, 1999, 2003). This introductory section provides a broad, basic overview as to how language develops in humans, according to usage-based models of language acquisition.

Language typically emerges in the first year of a child's life and is something that is often taken for granted. But how do children acquire language? This question underlies different theories and debates which account for the way in which language develops. For many years, the dominant theory of language acquisition was Chomsky's theory of Universal Grammar (1965), which (broadly) proposes that grammatical knowledge is genetically

determined and that infants are born with an innate knowledge of language. Influenced by Chomsky, modular accounts suggest that language is domain specific (Fodor, 1983), and that language is biologically determined and separate from cognition (Pinker, 1994, 1999). However, recent research regarding individual differences in language acquisition has challenged these theories; we now know that there are vast differences in the way in which people acquire and use language (Dąbrowska, 2012, 2015, 2018; Hulstijn, 2015; Kidd, Donnelly, & Christiansen, 2018; Kidd & Donnelly, 2020). This thesis does not attempt to address differences between the theories (see Ambridge and Lieven, 2011, for a detailed account of the different approaches), but aligns with usage-based models of language acquisition.

Bybee (2010, p.8) states that grammar is 'the cognitive organization of one's experience with language'. According to usage-based theories, children's language does not develop through innate representations but emerges gradually through general cognitive processes and linguistic experience (Behrens, 2009; Bybee, 2010; Ellis, Römer, & O'Donnell, 2016; Goldberg, 2006; Tomasello, 2003). The process of communication starts from birth - infants are born with perceptual abilities that enables them to learn language, and over time these general capacities become more language-specific (Mattys, Juscyzk, Luce, & Morgan, 1999). Soon after birth, newborn infants show a preference for their mother's voice over other female voices, having already becoming attuned to the rhythm and sound of the maternal voice in utero (DeCasper & Spence, 1986). Additionally, infants show sensitivity to their native tongue from as young as four days old (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988). From two months of age, infants engage in 'protoconversation' by exchanging sounds, gestures and facial expressions with their caregiver in a turn-taking manner (Trevarthen & Aitken, 2001). Evidence shows that infants are able to detect patterns in language early on in

development and can identify syllabic boundaries from a continuous speech stream from as young as eight-months of age (Saffran, Aslin, & Newport, 1996). Infants are also sensitive to phonotactic regularities and by nine-months of age, they can identify consonant clusters that occur frequently within words of their native language (Mattys et al., 1999). According to Tomasello (1999, 2003), it is at this age that infants start to understand communicative intentions in others and that they can share in these interactions. This is known as 'intention reading', or 'shared intentionality' (Tomasello & Carpenter, 2007). This stage is referred to as 'the nine-month revolution', a phase that certainly distinguishes human communication from primates (Tomasello, 1999). From nine-to-twelve months of age, infants share in joint attentional behaviour such as shared attention, gaze following, pointing or holding up objects, and imitative learning. These joint attention frames scaffold children's learning and create what Tomasello (2003) refers to as 'common ground' which is important in the context of shared cultural learning.

Human language is inherently pattern based. According to usage-based models, '…grammar is the result of the conventionalization of commonly used patterns: linguistic structure is formed through the repetition of linguistic patterns in language use.' (Divjak, 2019, p.17). This is important for language development as children have a remarkable capacity to detect patterns that occur in language, and through exposure, they begin to incorporate them into their own understanding and use of language. The process of language acquisition occurs gradually and in different phases which are (broadly) described as follows:

1. The initial stage of children's language development is somewhat formulaic and begins with constructions or form-meaning pairings which vary in size and specification (Goldberg, 1995). Children's early one-word utterances are holophrases such as *more*,

up, *down* (Tomasello, 1999, 2003), or simple phrases such as *thank-you*. These simple holophrases are often reproduced without being specifically analysed, but usually utterances undergo some form of categorisation. To do this, the child extracts complex forms, and through observing patterns that occur across utterances, these forms are broken down into chunks (words, morphemes and phrases), which are matched with similar structures and functions (Tomasello, 2003). In this way, children build up a database of analysed exemplars which vary in size and specification.

2. The second phase involves generalisation of stored exemplars to create novel utterances. To do this, the child uses analogy to infer items of a similar meaning and form new constructions. This process involves schematic representations (see Dabrowksa, 2010b), or slot-and-frame patterns, e.g., *It's X-ing*, *I want a Y*, which initially comprise of low-semantic function but increase in terms of abstraction (Lieven, 2016). For example, if the child has learned a phrase: *I want a* [*Y*], then a simple transitive such as: *I want a* [*banana*] could be aligned to: *I want a* [*strawberry*], where the slot for the noun, [*banana*], is replaced by a new noun representing a different piece of fruit, that is, *[strawberry*]. In this way, the child schematizes linguistic units which can then be generalised to new constructions, through analogy. The process of generalisation is slow and difficult at first but through considerable practice, linguistic units become routinized.
3. In the final stage of language development, new constructions are 'mentally encoded and committed to memory', a term referred to as entrenchment (Divjak, 2019, p.131). In this way, linguistic information is automatized for future use and fluent processing. Linguistic entrenchment relies on frequency of input (Divjak, 2019), although can occur from a singular representation, especially if the situation is salient to the learner (see Divjak & Caldwell-Harris, 2015). However, routinization strengthens associations,

so the more frequently an utterance is produced (or heard), the more likely that it will become entrenched (see Divjak, 2019; Divjak & Caldwell-Harris, 2015; Schmid, 2020).

During the initial stages of acquisition, children learn the structures that they hear and use most frequently. As a result, children's early language is item-based and partly formulaic (Behrens, 2009; Dąbrowska, 2010b). This is demonstrated in an early study conducted by Lieven, Pine and Baldwin (1997), who found that 60% of children's utterances were made up of 25 lexically based patterns. Similarly, in a later study of child-directed speech of second language (L2) English-speaking mother-child dyads and their two-year-old children, Cameron-Faulkner, Lieven, and Tomasello (2003) reported that that approximately half of the utterances that children are exposed to consist of highly frequent, lexically specific frames, usually comprising two morphemes. Of these, 45% began with one or two of 17 words, the five most common were: *what*, *that*, *it, you*, *are/aren't*. Furthermore, 32% of the utterances were questions. Children used many of these frequent frames in their own speech, often correlating with frequency of use in their mother's speech. Expanding upon this, Dąbrowska and Lieven (2005) investigated whether children of two and three years of age use constructions based on lexically specific frames, by juxtaposing and superimposing items in their sample. 90% of utterances produced by the children were based upon previously recorded utterances and the remaining 10% were based on extensions of previously encountered utterances. It has been demonstrated that 58–92 percent of multiword utterances in two-year-old children can linked back to the child's own corpora as exact repetitions of their own utterances, or those that they have heard only once or twice before (Lieven, Salomo, & Tomasello, 2009). As children are exposed to a wider variety of linguistic representations (from what they hear and from their own day-to-day usage), their constructions grow in complexity and become more abstract.

This, in turn, increases the capacity of available slots and expands their productivity (Dąbrowska, 2004).

Domain general cognitive abilities underlie usage-based models of language acquisition, and this thesis aims to explore certain cognitive abilities which are known to be important to language learning. These specific abilities will be discussed later in this introductory section.

1.2 Developmental language disorder (DLD): a definition

Language acquisition is a robust process, and most children acquire their native language (or languages) without apparent difficulty. However, for some children, language does not progress at the expected rate despite typical development in other areas. For these children, language difficulties are due to a condition known as Developmental Language Disorder (DLD). This section will discuss identification, potential causes and characteristics of DLD in childhood and adulthood.

DLD is a neurodevelopmental disorder affecting the way in which children understand and produce spoken language (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 consortium, 2017; Leonard, 2014a). DLD is a heterogeneous condition and people with DLD typically have strengths and weaknesses across language and cognition. Language problems are usually persistent and children who begin school with language difficulties rarely 'catch up' with their language typical (LT) peers, with a two-to-three year gap in the first few years of primary school which does not diminish with time (Norbury, Vamvakas, Gooch, Baird, Charman, Simonoff, & Pickles, 2017). For many children with DLD, the difficulties that they experience have a detrimental effect on their educational, social, and emotional progress, with

ongoing problems that continue to middle childhood and later life (Bishop et al., 2017; Clegg, Hollis, Mawhood, & Rutter, 2005; Conti-Ramsden, 2003; Johnson, Beitchman, & Brownlie, 2010).

Various studies estimating the prevalence of DLD provide similar percentage occurrences in the general population, for example, 7.58% in the United Kingdom (Norbury, Gooch, Wray, Baird, Charman, Simonoff, Vamvakas, & Pickles, 2016), 7.4% in the USA (Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien, 1997), 8.5% in China (Wu et al., 2023), and 6.4% in Australia (Calder, Brennan-Jones, Robinson, Whitehouse, & Hill, 2022). However, DLD often remains largely undetected in individuals (McGregor, 2020; Norbury et al., 2016; Tomblin et al., 1997). One reason is that DLD is not very well-known and DLD receives less research funding than other developmental conditions such as Autism Spectrum Disorder (ASD) (Bishop, 2010; McGregor, 2020), despite DLD being seven times more prevalent than ASD (McGregor, 2020). Identification of DLD is another problematic area with previous research indicating that many parents and carers are not able to confidently determine whether their child has oral language difficulties (Adlof, Scoggins, Brazendale, Babb, & Petscher, 2017; Hendricks, Adlof, Alonzo, Fox & Hogan, 2019; Tomblin et al., 1997). This situation is further exacerbated at school-going age, as although children with DLD often have immature language skills in comparison to their age related peers, by the time they reach school they can usually hold a conversation and follow simple instructions (McGregor, 2020). As a result, their language difficulties may go unnoticed, especially as many educational professionals lack adequate training to identify DLD, or to differentiate speech problems from language difficulties (Dockrell, Howell, Leung, & Fugard, 2017; Messer & Dockrell, 2006). Even when language difficulties are detected, there is inequity in terms of speech and language therapy services, screening facilities and support offered to pupils in school (in the U.K.: see

Dockrell, Lindsay, Letchford, & Mackie, 2006, Dockrell; Ricketts, Palikara, Charman, & Lindsay, 2019, and in the USA: see McGregor, 2020).

Another reason for poor identification of DLD is due to inconsistency regarding terminology in the past. Historically, DLD was widely referred to as Specific Language Impairment (SLI), a term that gained popularity in the 1980s. Diagnostic criteria for SLI are determined by language assessment scores that fall more than 1.25 standard deviations below the mean, combined with nonverbal IQ scores of 85 or higher (Reilly et al., 2014). However, because diagnosis of SLI stipulated a mismatch between language and cognition (i.e., poor language yet strong cognition), this ultimately led to many children being ineligible for diagnosis and intervention support. In response to the lack of consensus regarding terminology and diagnostic criteria, an international panel called 'The CATALISE consortium' was formed. This group comprised 59 professionals from numerous disciplines, including education, psychology, speech and language therapy, paediatrics, and child psychiatry. Two Delphi consensus studies were conducted to discuss recommendations regarding identification and diagnosis of DLD (Bishop et al., 2016, 2017). As a result, the term Developmental Language Disorder is now recommended for children who have persistent difficulties learning, using or producing language. These difficulties are not explained by biomedical conditions such as sensorineural hearing loss, Down Syndrome, or ASD, and inclusive of children with lower nonverbal IQ (Bishop et al., 2017). Importantly, according to the new diagnostic criteria for DLD, there is no longer a mismatch between nonverbal IQ and language, hence DLD encompasses a broader profile of children.

Comorbidity is common and DLD is often diagnosed alongside co-occurring difficulties such as Attention Deficit Hyperactivity Disorder (ADHD), developmental dyslexia,

speech difficulties, motor skills, and sleep problems (see Nitin et al., 2022). However, according to the new diagnostic criteria for DLD (Bishop et al., 2017), biomedical conditions such as ASD are considered a differentiating factor, and in these cases a language disorder is seen as associative rather than a co-occurring difficulty. In such instances, an example of recommended terminology would be: 'Language disorder associated with ASD'. This is because the biomedical condition (in this case, ASD), has a potential genetic or neurological origin and additional factors need to be considered when planning language interventions (Bishop et al. 2017, p.1071).

The variability typically seen in individuals with DLD, and the clinical interpretation of the condition, has led to the suggestion that DLD should be reconceptualised as a continuum/spectrum disorder (Lancaster & Camarata, 2019). Based on data from an epidemiological sample of children with and without DLD (505 children, data from the US Epidemiological Study of Language Impairment database: Tomblin, 2010), clustering techniques were used to ascertain whether DLD is best classified according to subtypes, individual differences, and/or a continuum/spectrum model. Their results did not provide sufficient evidence for a subtype model of DLD, and the authors report mixed evidence for an individual differences model. However, a continuum/spectrum model was supported by their analyses in recognising '…differential diagnosis of severity *and* individual patterns of strengths and weaknesses' (Lancaster & Camarata, 2019, p.13). It should be noted that this study focused on a population of children based in the United States and it would be interesting to see results from additional population-based cohorts in different countries.

Although the recommended term of DLD is widely used in research and diagnosis, the term SLI is still used in research (see Rice, Taylor, Zubrick, Hoffman, & Earnest, 2020b).

Many studies referred to in this thesis will use the term SLI, especially if research was conducted prior to the CATALISE consortium study. To maintain consistency with the current recommended terminology, the term DLD will be used throughout this thesis.

1.3 Causes of DLD: biology

DLD is considered a polygenic disorder (Mountford, Braden, Newbury, & Morgan, 2022; Newbury, Gibson, Conti-Ramsden, Pickles, Durkin, & Toseeb, 2019), meaning that there are several genes involved in the condition. Yet, despite extensive research, we do not fully understand the biological factors that contribute to DLD. This section provides a basic overview of research regarding environmental risk factors and evidence from genetic and neurobiological studies.

1.3.1 Risk factors for DLD

A meta-analysis by Rudolph (2017) reviewed eight epidemiological studies to ascertain potential risk factors for DLD. Those risk factors that were identified as being clinically significant include: maternal education level (lower), Apgar score at birth (lower), birth order (later), biological sex (male), and later speech development. Similar risk factors were identified in a recent review by investigating neurobiological mechanisms of risk factors in DLD. Among others, identified risks include: being male, later birth order, alcohol and nicotine exposure in-utero, pre-term birth, and reduced sleep quality (see Boerma, ter Haar, Ganga, Wijnen, Blom, & Wierenga, 2023). Specific to Australia, data from The Raine Study cohort were explored to ascertain prevalence of DLD and risk factors (Calder et al., 2022). Based on a population sample of 1,506 children, potential risk factors were assessed by parental report and include biological sex, gestational age at birth (below 37 weeks), maternal ethnicity, second language

other than English, SES, smoke and alcohol exposure in utero, father's residence at birth, day care attendance, and whether the child was read to at home. Their results revealed that the only significant risk was found to be smoking in utero, at 18 weeks' gestation (Calder et al., 2022).

1.3.2 Behavioural genetic research

From a genetic perspective, twin studies provide evidence that identical (monozygotic) twin pairs have a much higher trait concordance rate in comparison to non-identical (dizygotic) twins (Bishop et al., 1995; 2006; Lewis & Thompson, 1992; Tomblin & Buckwalter, 1998). The significance of genetic factors has been examined in several studies using data from the longitudinal Twins Early Development Study (TEDS) (Oliver & Plomin, 2007). Hayiou-Thomas, Oliver, and Plomin (2005) analysed data from the TEDs cohort to explore genetic and environmental factors contributing to specific (DLD), and non-specific (NLI) language difficulties. DLD status was determined by scores of one standard deviation or more below the control group's mean in language and nonverbal scores, whereas the NLI group was determined by language and nonverbal scores of minus one standard deviation below the mean of the control group. Interestingly, the results revealed a higher group heritability for NLI compared to DLD, with group heritability (h^2_g) estimates of: $h^2_g = .18$ for DLD, compared to $h^2_g = .52$ for NLI. Because these results were considerably lower than expected, further investigation was conducted by Bishop and Hayiou-Thomas (2008), where it was found that twins identified by population screening did not show a high heritability, but those identified clinically (i.e., through speech and language services), had a substantially higher heritability rate of .97. In a more recent study of TEDS data (210 children), Hayiou-Thomas, Smith-Woolley, and Dale (2021) report that a higher number of risk factors present in early childhood (i.e., those present at four-years of age), increase the likelihood of poor language and literacy outcomes in later childhood or early adolescence. Risk factors include family history of

language or reading difficulties, and poor language, speech, emerging literacy and nonverbal skills. It is suggested that these factors increase the likelihood of poor oral language, reading fluency, and reading comprehension at the age of 12.

A longitudinal, population-based, twin study investigated the heritability of speech, language and nonverbal cognition of twins, from two to six years of age. At two years of age, assessments were via parent report, and at four and six years of age, they were based on direct child assessments. Overall results established that there is an inheritable basis for language impairment, with a persistent twinning effect across language and speech phenotypes (Rice, 2020a; Rice, Zubrick, Taylor, Gayán, & Bontempo, 2014, Rice, Zubrick, Taylor, Hoffman, & Gayán, 2018, Rice, Taylor, Zubrick, Hoffman, & Earnest, 2020b). It is interesting to point out that twinning effects predicted DLD status, but no twinning effect was observed for nonverbal IQ, suggesting that nonverbal IQ does not have the same causal pathway as that of DLD (Rice et al., 2020b).

Mountford et al. (2022) highlight the fact that genetic studies usually focus on monogenic disorders (variation in a single gene, hence studies are typically conducted within families), and complex disorders (genetic association studies). An example of a monogenic inheritance pattern is the *FOXP2* gene, identified in a three-generation family referred to as the 'KE family'. A genome-wide linkage search by Fisher, Vargha-Khadem, Watkins, Monaco, and Pembrey (1998) identified a monogenic trait in the KE family, caused by a region on chromosome area 7q31. Fine mapping of this region in a later study showed that mutation in the *FOXP2* gene causes impaired neural structures that support speech and language (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). The KE family has been of interest to in the context of DLD research as approximately half of the KE family members have childhood

apraxia of speech, and some members also have expressive and receptive language difficulties and nonverbal cognitive deficits (Watkins, Dronkers, & Vargha-Khadem, 2002). However, it was found that the *FOXP2* gene does not pose a risk factor for DLD. Whilst triggering language deficits similar to the deficits observed in DLD, the *FOXP2* gene does not appear to be linked to the onset of disorders such as autism or DLD (Newbury et al., 2002). Therefore, as described by Bishop (2006), *FOXP2* is not specifically considered as a gene for language but it regulates other gene activity, including areas in the brain which are central to speech and language.

Mountford et al. (2022) discuss advances in recent genetic sequencing technology which provide information regarding significant gene variants contributing to DLD (see results from Chen et al., 2017; Plug et al., 2021). In addition, recent advances in genome-wide association studies (GWAS) are promising for the field of genetic research as they employ new technology, and utilise extremely large cohorts, hence generating a great deal of statistical power (for example, see Doust et al., 2022 for dyslexia, Eising et al., 2022, for reading and literacy skills, and Lin et al., 2024, for language phenotypes in Chinese L1 and Chinese-English bilinguals).

1.3.3 Neurobiological research

Research shows that people with DLD have differences in brain regions which are important for language. In particular, irregularities have been observed in the basal ganglia. Krishnan, Watkins, and Bishop (2016, p. 703), define the basal ganglia as encompassing the '…striatum (caudate nucleus, putamen, and nucleus accumbens), globus pallidus, substantia niagra, and subthalamic nucleus', and suggest that the basal ganglia are involved in skill learning and sequential learning. The authors propose that corticostriatal dysfunctions underlie language difficulties in DLD which could explain poor performance on tasks involving

procedural learning (e.g., implicit learning of complex rules, probabilistic and sequential learning). A systematic review encompassing 18 studies, conducted by Mayes, Reilly, and Morgan (2015), revealed that children with language difficulties have atypical brain structure and function in principal language regions: specifically, the frontal gyrus and posterior superior temporal gyrus, and caudate nucleus. A later study by Krishnan et al. (2022) used semi-quantitative structural magnetic resonance imaging (MRI), to explore neural differences in children between 10 and 15 years of age, with and without DLD. Their results showed that children with DLD have less myelin in certain areas of the brain (specifically, reduced myelin in the left caudate nucleus), in comparison to LT controls. Myelin enables quick and reliable communication; therefore, reduced myelin may contribute to speech and language difficulties in children with DLD (Krishnan et al., 2022). However, both Mayes et al. (2015) and Krishnan et al. (2016) suggest that hemispheric asymmetry should be interpreted with caution. Mayes et al. (2015) propose that there is the possibility of inconsistencies in the direction of results (i.e., increased or decreased structural volume, as well as functional activation), the fact that studies often involve small sample sizes, different approaches to data analyses, and also differences due to the heterogeneity of DLD samples. Lastly, a recent systematic review and meta-analysis by Ullman et al. (2024) examined structural brain data of 577 participants (including 250 with DLD). The results show that individuals with DLD have structural abnormalities in the basal ganglia and particularly in the neostriatum. The authors suggest that the evidence is robust enough to warrant the classification of basal ganglia abnormalities as an early biomarker of DLD.

1.3.4 Transdiagnostic approach to neurodevelopmental disorders

A new approach to the understanding of neurodevelopmental disorders (including DLD), has been put forward in a comprehensive review by Astle, Holmes, Kievit, and

Gathercole (2022). The authors raise issues regarding the dominant approach in research design which relies on selective, binary diagnostic criteria and group designs, i.e., where a neurodivergent group is compared to typically developing controls. Instead, they propose a transdiagnostic approach which involves identifying dimensions and clusters of cognitive, behavioural and neural characteristics between individuals, which occur in a multidimensional space. This approach is considered effective as most neurodevelopmental disorders have overlapping characteristics, for example, pragmatic skills, hyperactive behaviour, phonological processing, and executive function difficulties (Astle et al., 2022; Astle, Bassett, & Viding, 2022). Therefore, the aim of a transdiagnostic approach is to increase understanding of neurodevelopmental disorders by encompassing a broader range of individual needs (rather than the general current methodology of comparative group designs), and ultimately translate this knowledge into supportive interventions and teaching practice.

1.4 Language characteristics of DLD

As shown in Figure 1, language deficits in DLD fall under the umbrella of speech, language and communication needs, and typically involve areas of syntax, morphology, semantics, word finding, pragmatics, phonology, discourse, and verbal learning/memory (Bishop et al., 2017).

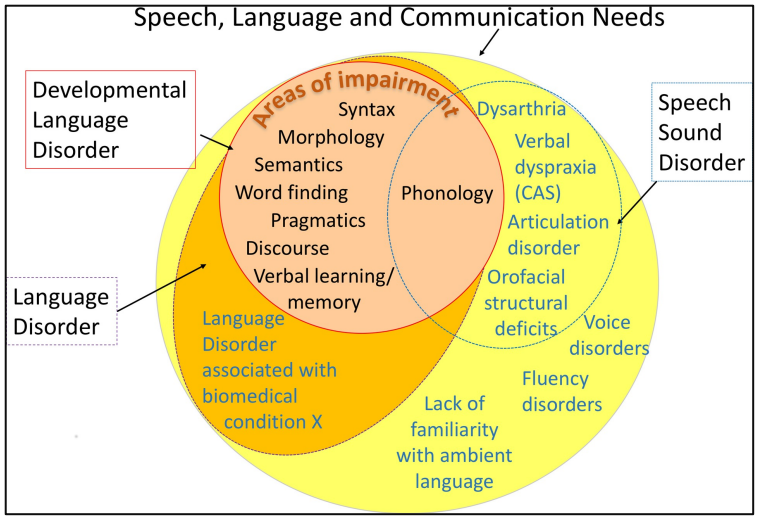


Figure 1. Illustration of areas of difficulty commonly seen in DLD

(Bishop, Snowling, Thompson, Greenhalgh, and the CATALISE-2 consortium, 2017)

Children with DLD typically have weak language skills but due to the heterogeneity in DLD, there is no typical language profile and some areas are suggested to be more challenging than others. As explained by Lancaster and Camarata (2019), previous research has explored DLD from the viewpoint of subtypes or clusters (see Conti-Ramsden et al., 1997; Conti-Ramsden & Botting, 1999), or related to specific individual differences (Leonard, 2014a). More recent evaluation (as discussed in the previous section), views DLD as a continuum/spectrum condition (Lancaster & Camarata, 2019).

It is beyond the scope of this thesis to explore the different areas of language that are implicated in DLD but this section provides a brief overview of some of the literature related to word learning and vocabulary, grammar, and writing.

1.4.1 Word learning and vocabulary

The onset of language development in children with DLD is frequently delayed and children with DLD may begin to talk later than expected, with their first words often not appearing until two years or later (Bishop, 2006; Leonard, 2014a). Children with DLD have difficulty learning new words (Alt, Plante, & Creusere, 2004; Alt & Plante, 2006; Gray, 2003; Kan & Windsor, 2010), and typically have vocabularies that are not as large and without the same complexity as language typical peers (McGregor, Oleson, Bahnsen, & Duff, 2013, McGregor, Van Horne, Curran, Cook, & Cole, 2021; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

Word learning has been explored from the context of 'fast mapping', which suggests that children typically learn new words quickly, sometimes after a single exposure (Carey & Bartlett, 1978). One of the earliest studies to explore fast mapping in children with language difficulties found that pre-school children with DLD were able to associate a non-word with a referent, yet their retention of the new word was below LT peers (Dollaghan, 1987). Prior research suggests that verbs may be a particular area of difficulty for children with DLD in early stages of word learning (Leonard, 2014a). An example of this is demonstrated in a study by Riches, Tomasello, and Conti-Ramsden (2005), where they investigated verb learning in a study with pre-school children with DLD (mean age 5;6 years). Children were taught four novel verbs which were presented in the same order: e.g., *dack, tam, meek, gorp*, although exposure varied, with mass exposure on day one, versus spaced exposure over four days. Children were tested immediately after the presentation of a novel verb and a week after the final training session, to assess retention. In comparison to the LT group (language matched, with a mean age of 3;4 years), the children with DLD showed a greater dependence upon frequency and they benefited from more widely spaced presentation of the verbs. These

findings are in line with a much earlier study of children of a similar age range, by Rice, Oetting, Marquis, Bode, and Pae (1994), where children were shown a video which contained target words embedded into the story (verbs and nouns). A strong effect of frequency was found in LT children, but this was not sufficient enough to facilitiate retention in children with DLD, who required a higher input frequency (ten presentations) to learn new words. In comparison, LT controls were able to retain learning after just three presentations of words.

Jackson, Leitão, and Claessen (2021) investigated word learning deficits in children with DLD to ascertain whether difficulties arise early on in encoding and retaining word forms in memory, or whether problems stem from not being able to establish a link between word forms and their respective meanings. Children with and without DLD (mean age 6;11 years), were taught eight novel words (pseudowords) and were assessed over four days to measure encoding (referring to the initial memory trace linking a word form to its referent), re-encoding, and retention (involving consolidation of the memory trace to long-term memory). In comparison to the LT group, the children with DLD showed significant problems in learning new words and their meanings. However, they were able to establish form-referent links of novel words (for example, by determining the word form and meaning and creating a link between these two elements). The authors suggest that difficulties potentially arise from encoding word forms, that is, the initial memory trace of a word and it's referent, prior to integration to long-term memory, especially for children with DLD who have poor verbal working memory. These findings align with further studies which provide evidence of encoding deficits in children with DLD (Gordon, Storkel, Lowry, & Sultani, 2024; Haebig et al., 2019; Leonard et al., 2019).

1.4.2 Grammar

Grammar is a specific area of difficulty in DLD, with evident weaknesses in syntax (Delage & Frauenfelder, 2020; Hsu & Bishop, 2014; Montgomery, Evans, Fargo, Schwartz, & Gillam, 2018, Montgomery, Gillan, & Plante, 2024; van der Lely, 2005), and inflectional morphology (van der Lely, 2005). Children with DLD tend to produce sentences that are shorter and structurally more simplistic, with less subordination than those produced by LT children (Nippold, Mansfield, Billow, & Tomblin, 2008, 2009).

In DLD, the development of tense and agreement is delayed and persists in an inconsistent manner beyond pre-school age (for a review, see Leonard, 2014a). Examples of specific morphemes include finiteness markers such as third-person singular present -*s*, past-tense -*ed*, *BE* (is, are, am, was, were), and *DO* forms (do, does, did) (Rice, Wexler, & Cleave, 1995; Rice, Wexler, & Hershberger, 1998; Rice & Wexler, 1996). In pre-school children, omission rates for -*s* and -*ed* are approximately 75%, and between 40 and 60% for *BE* and *DO.* Tense-marking is considered a phenotype marker of DLD (Conti-Ramsden, Botting, & Faragher, 2001; Rice & Wexler, 1996), with suggested clinical reliability continuing into adolescence (Dale, Rice, Rimfield, & Hayiou-Thomas, 2018).

By examining the natural language input that children with DLD receive, a recent review article by Leonard, Deevy, Bredin-Oja, and Schroeder (2024), explores why tense and agreement marking is particularly difficult for preschool children with DLD. They suggest that subject + nonfinite verb sequences in auxiliary-fronted questions are problematic for children with DLD because they cannot link them to the input that they receive, and so they are treated as stand-alone utterances. For example, in response to: *Can [the horse run fast]?*, children with DLD will often produce utterances which omit a tense and agreement morpheme, such as *The*

horse run fast. Leonard et al. (2024) suggest that children do not link the fronted modal auxiliary (*can*) as providing the relevant agreement and/or tense information. In early development, children frequently hear bare verb stems (e.g., *I run, you run, we run, they run*), and are not often exposed to the third-person singular *-s*. Accordingly, children with DLD tend to adopt bare stems in their speech (Leonard et al., 2024; Leonard & Kueser, 2019) . Further grammatical errors are highlighted by Leonard and Kueser (2019), for example, confusion with grammatical case, as seen in incorrect use of pronouns, such as *her eating* and *him do that* (p.351). Children with DLD may also have weaknesses with prosody, grammatical aspect (such as the progressive *-ing* inflection), and lexical aspect (for example, telic and atelic predicates, where telic predicates suggest that the event is completed, such as *he ate the apple*, whereas atelic predicates do not provide this information, e.g., *she swam fast*). In addition, some children with DLD have difficulties with certain sentence structures that diverge from canonical word order (e.g., Subject-Verb-Object word order in English). On the other hand, it has been observed that English speaking children with DLD do not typically produce word order errors in simple declarative sentences, but tend to make errors in passives, among other structures. Word order errors vary according to the language spoken (see Leonard & Kueser, 2019 for examples of Dutch, German, Swedish, Hebrew. Italian, and Danish languages).

Lastly, whilst a wide body of research has focused on English grammar, grammatical weaknesses are attested across languages, such as Romance, Germanic, Uralic, and Chinese languages (for a review, see Leonard, 2014b).

1.4.3 Literacy

In addition to oral language difficulties, many people with DLD face challenges with reading and written language, and there is a high rate of comorbidity between DLD and

dyslexia despite the fact that the two conditions are separable, with different aetiologies (Bishop & Snowling, 2004; Snowling et al., 2019; 2020). The overlap between the two conditions is not surprising considering that oral language skills provide the foundation for children's reading and writing development (Adlof & Hogan, 2019), and reading comprehension skills (Nation & Snowling, 2004). This is illustrated in a study by Snowling, Hayiou-Thomas, Nash, and Hulme (2020). They examined reading comprehension outcomes in a large cohort of 8- and 9-year-old children who were part of a longitudinal study, recruited during preschool. The children were divided into three groups: those with dyslexia, DLD, or comorbid dyslexia and DLD. Their results showed that approximately 48% of children with DLD also met the criteria for dyslexia. For these children, the combination of oral language difficulties and poor decoding skills leads to severe problems with reading comprehension.

There is limited research regarding the relationship between DLD and writing skills. A scoping review by Tucci and Choi (2023) highlights several areas of weakness in writing skills of preschool children (aged four to six years), childhood (defined as less than ten years of age), through to adolescence (10-17 years of age), and adulthood (18 years and over). In pre-school children, problems include difficulties with emergent writing skills, such as letter writing and name writing. Older children and adolescents have difficulty with spelling, grammar in written texts (for example, morphology, sentence structure and grammatical accuracy). Writing samples at this stage of development are generally shorter and lexically less complex than written texts of LT peers. Writing processes are also affected, with evidence of planning difficulties, problems with cohesion, and slower writing speeds. Adults with DLD may have spelling difficulties - however, the authors could not draw firm conclusions regarding other areas of written language due to a general lack of research in adult populations (out of 72

studies included in the review, 12 included adults with DLD, but of these only six were focused on adults).

1.5 Clinical markers of DLD

Prior research has identified three recommended clinical markers of DLD which comprise: 1) non-word repetition as a measure of phonological short-term memory, 2) sentence repetition, using tasks which assess children's ability to repeat sentences of increasing length and syntactic complexity, and 3) tense-marking ability. Of these three, it is suggested that sentence repetition is the most reliable phenotypic marker in children with DLD (Archibald & Joanisse, 2009; Conti-Ramsden et al., 2001), and in adulthood (Poll, Betz, & Miller, 2010). In Sentence repetition tasks have high levels of sensitivity (90%), specificity (85%), and overall accuracy (88%) in identifying DLD (Conti-Ramsden et al., 2001). A recent meta-analysis examining performance in sentence repetition tasks in LT children and DLD populations (46 studies, 103 effect sizes), reports an overall effect size of Hedges' g = -2.08 (Ward, Polišenská, & Bannard, 2024). Furthermore, sentence repetition has been established as a reliable clinical marker in many languages besides English: for example, in Spanish speaking children with DLD (Pratt, Peña and Bedore, 2021), Mandarin-speaking children (Wang, Zheng, Lin, Zhang, & Sheng, 2022), Arabic (Taha, Stojanovik, & Pagnamenta, 2021), Italian (Guasti, White, Bianco, Arosio, Camilleri, & Hasson, 2021), and Danish (Vang Christensen, 2019).

The accuracy of clinical markers of DLD in differentiating between individuals with and without DLD has been examined in a meta-analysis by Pawłowska (2014). Reporting data from 13 studies, the meta-analysis assessed proposed clinical markers in L1 English. Overall findings conclude that performance on sentence repetition, nonword repetition, and tense marking should be interpreted as suggestive of the presence or absence of DLD. Several

limitations are raised, including the use of different assessment tools between studies and the use of different diagnostic cut off scores, as well as the accuracy in defining DLD/LT groups (for example, some studies may compare clinical DLD groups with high ability LT groups).

However, more recently, Redmond, Ash, Cristopulos, and Pfaff (2019), examined two clinical markers, sentence repetition and tense marking ability, in a community-based sample of 1,060 children (mean age 7;1 years). Several reference standards were used to assess validity, including clinically significant levels of parental concern, behavioural measures, whether the child was in receipt of language therapy, and composite language scores. Based upon these criteria, the results showed valid diagnostic accuracy and substantiate the use of sentence repetition and tense marking as clinical markers of DLD.

1.6 DLD in adulthood

A small but growing body of literature has investigated the trajectory of language abilities in DLD beyond childhood. While language difficulties are sometimes resolved in development, for many people with DLD, problems are persistent (Bishop & Edmundson, 1987; Stothard et al., 1998). Language problems can have a negative impact on confidence and self-esteem (Durkin, Toseeb, Botting, Pickles, & Conti-Ramsden, 2017), and several longitudinal studies report the risk of social, emotional and mental health difficulties in adults with DLD (Botting, Durkin, Toseeb, Pickles & Conti-Ramsden, 2016; Botting, Toseeb, Pickles, Durkin, & Conti-Ramsden, 2016; St Clair et al., 2022; Johnson, Beitchman, & Brownlie, 2010; Law, Rush, Schoon, & Parsons, 2009; Whitehouse, Watt, Line, & Bishop, 2009). Additionally, language difficulties have an impact upon other areas of life, including education, employment, and socio-emotional difficulties (see Wilmot, Boyes, Sievers, Leitão,

& Norbury, 2024, for first-hand perspectives regarding the lived experiences of adults with DLD).

1.6.1 Language and literacy

Longitudinal studies show that adults with DLD have continued language and literacy difficulties (Botting, 2020; Clegg, Hollis, Mawhood, & Rutter, 2005; Conti-Ramsden, Durkin, Toseeb, Botting, & Pickles, 2018; Law et al., 2009, 2009; Whitehouse et al., 2009). Botting (2020) evaluated language, literacy, and cognition in a group of adults from the Manchester Language Study (MLS) cohort. At the age of 24, standardized language assessment scores were lower for the DLD group in comparison to age matched peers. This was particularly evident in areas that require functional literacy in everyday life, such as reading bank statements and bills, writing phone messages, letters, cheques, and completing forms. A positive finding was that the majority of the DLD group reported reading books, magazines and newspapers, despite their literacy difficulties. As pointed out by Del Tufo and Earle (2020), where children with DLD typically have problems with grammar and syntax, in adulthood, difficulties tend to occur in areas requiring 'higher order language', such as comprehending legal language - the authors give an example of this by referring to Miranda warnings which are legal rights read to individuals at the time of arrest (see Rost & McGregor, 2012). Additionally, adults with DLD have difficulty with both spoken and written narratives (Suddarth, Plante, & Vance, 2012), and problems with spelling (Tucci & Choi, 2023).

Diagnosis and awareness of DLD in adult populations is problematic. As discussed by Del Tufo and Earle (2020), there are many factors at play. There are no normed diagnostic tools post adolescence, and child diagnostic assessments are not sensitive enough to detect higher-order language difficulties that are often seen in people with DLD later on in life. This

potentially creates a situation where adolescents have a language profile that no longer meets diagnostic criteria for DLD which may place them at risk of being discharged from services prior to adulthood (Del Tufo & Earle, 2020; also see Krassowski and Plante, 1997)).

1.6.2 Word learning and memory

A series of studies examining word learning and memory in college students with DLD, shows significant word learning deficits in adulthood (McGregor, Licandro, Arenas, Eden, Stiles, Bean, & Walker, 2013; McGregor, Arbisi-Kelm, & Eden, 2017; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017; McGregor, Arbisi-Kelm, Eden, & Oleson, 2020). In these studies, participants were trained on a set of novel words and their referents, with assessments conducted immediately after training and then again a week later. In the 2013 study, when participants were tested after the training phase, the DLD group remembered fewer words and meanings in comparison to the LT group. The authors interpreted this as relevant to encoding problems. When tested one week after the initial training session, the gap between the DLD and LT group remained stable for recalling meanings, yet memory for word forms was problematic (McGregor et al., 2013). Later studies showed a deficit in encoding but not retention (McGregor et al., 2017; McGregor et al., 2017, 2020), with adults with DLD requiring as much as 39% more exposure than LT adults, to encode new words into long-term memory (McGregor et al., 2020). Overall, the authors conclude that encoding new word forms is a barrier for word learning in adults with DLD, yet they have relative strengths in retention. As highlighted earlier, is important to note that the encoding of lexical items refers to the initial memory trace, where a word form is connected to its referent. However, consolidation is essential for the integration of these memory traces into long-term memory (see Davis & Gaskell, 2009; Henderson, Weighall, Brown, & Gaskell, 2013; James, Gaskell, & Henderson, 2020).

Further studies have explored the implications of poor memory in college students. Del Tufo & Earle (2020) found that college students with DLD have problems with nonword repetition and verbal working memory, which affected their reading comprehension under timed conditions (for example, in examination conditions). Similarly, it has been shown that college students with DLD find it challenging to learn from spoken content – learning 20% less than their LT peers when assessed by recognition and recall assessments immediately after lectures. These students also achieved lower scores than their LT peers in phonological short-term memory and nonword repetition (Becker & McGregor, 2016).

1.6.3 Education and employment

Unresolved language difficulties affect other areas of life such as education, employment prospects, and independent living (see Dubois, St-Pierre, Desmarais & Guay, 2020 for a recent systematic review). Although educational performance varies between individuals, a systematic review of academic performance in students of primary and secondary school (three to 19 years of age), shows that individuals with DLD are at risk of lower academic achievement (Ziegenfusz, Paynter, Flückiger, & Westerveld, 2022). This may result in young people with DLD choosing to leave formal education earlier than LT peers (Conti-Ramsden et al., 2018; Schoon, Parsons, Rush, & Law, 2010).

Several studies have shown that people with DLD are often employed in less-skilled roles in comparison to their LT peers (Clegg et al., 2005; Conti-Ramsden et al., 2018; Dubois et al., 2020; Johnson et al., 2010; Law et al., 2009; Whitehouse et al., 2009). On the other hand, Conti-Ramsden et al. (2018) found that the number of adults with DLD gaining accredited qualifications was higher than previously shown in longitudinal studies during the 1990s and

2000s (where studies reported that many young adults with DLD left school with no qualifications and did not pursue further education in adulthood, see e.g., Clegg et al. 2005). It has also been noted that people with DLD approach the job market in a less traditional way, sometimes not using curriculum vitaes, and reporting challenges in attending assessment centres or interviews (Dubois, Guay, & St-Pierre, 2023).

Evidence to date establishes the crucial significance of exploring both cognition and language in people with DLD across the lifespan. The sections that follow discuss specific cognitive abilities that are of interest to this thesis, with an overview of prior research and how this relates to DLD.

2.0 COGNITIVE ABILITIES

2.1 Intelligence / Nonverbal IQ

Intelligence has been defined as two separable constructs: crystallised intelligence (Gc), which refers to acquired knowledge (for example, vocabulary), and fluid intelligence (Gf), which involves reasoning and problem-solving skills (Cattell, 1943, 1963; Cattell, Feingold, & Sarason, 1941). Gc peaks in late adulthood, at around 60 years, before declining, whereas Gf has been found to peak in the twenties and then decline from middle to late adulthood (Salthouse, 2010; Tucker-Drob, 2019).[1] Nonverbal intelligence (IQ) scores are usually stable

[1] Research regarding intelligence has resulted in considerable debate spanning many decades, discussion of which is not central to this thesis. However, extensive reviews are provided by Neisser et al. (1996), Nisbett et al. (2012), and more recently, by Deary, Cox and Hill (2022).

during development and throughout the lifespan (Deary, Cox, & Hill, 2022; Neisser et al., 1996).

The relationship between cognitive abilities and education is well-established (Jencks, 1979; Jensen, 1998; Neisser et al., 1996). For example, there is a strong correlation between intelligence measured at the age of 11 years (using a cognitive abilities reasoning test), and academic attainment in standardized national examinations at 16 years of age *(r* = .81: Deary, Strand, Smith, & Fernandes, 2007). A bidirectional relationship is suggested, with the possibility that intelligence motivates a longer period of education, and vice versa, in that education increases IQ test scores (Deary et al., 2022; Ritchie & Tucker-Drob, 2018). One aspect to bear in mind, however, is that intelligence is typically measured by standardized IQ tests, and it should be considered that these standardized assessments do not capture certain elements such as creativity and culture (Neisser et al., 1996).

A robust relationship has been observed between nonverbal IQ and working memory (Kane & Engle, 2002), and between nonverbal IQ and vocabulary (Raven, 2000). Furthermore, prior research accounts for a relationship between IQ and explicit learning but not between IQ and implicit learning (Feldman, Kerr, & Streissguth, 1995; Reber, Walkenfeld, & Hernstadt, 1991). To this extent, IQ is not considered relevant for grammar as it is generally acknowledged that grammar relies exclusively on implicit learning mechanisms (see, for example, DeKeyser, Alfi-Shabtay, & Ravid, 2010; Ellis, 1996; Ullman, 2001). However, recent studies provide evidence of a relationship between nonverbal IQ and grammar in adults. For example, Dąbrowska (2018) examined individual differences in a group of L1 speakers, reporting a marginally higher relationship between nonverbal IQ and grammar (r = .46), than that between nonverbal IQ and vocabulary (r = .43). Similarly, in Lancaster and Camarata's (2019) analyses

of a population-based sample of children with DLD, it is noted that simple correlations reveal an association between nonverbal IQ and grammar (r = .25, as measured by the Grammatical Understanding subtest of the Test of Language Development: Newcomer and Hammill, 1988), whereas the relationship between nonverbal IQ and vocabulary is reported at r = .15.[2]

Nonverbal IQ has played a significant role in DLD research with discussion centred around diagnostic criteria for SLI. As discussed earlier in this thesis, SLI diagnosis required that an individual should have impaired language ability, yet normal scores on standardised measures of nonverbal IQ (i.e., no more than one standard deviation below the mean: e.g., > 85). According to the new terminology of DLD, a difference between nonverbal and verbal ability is no longer required for diagnostic purposes (Bishop et al., 2017). Therefore, the term DLD encompasses a wider group of children and ultimately means that more children with language difficulties can access the services that they need. However, it has been observed that diagnosed children frequently score below LT peers in nonverbal IQ tests (Cole, Mills, & Kelley, 1994; Miller & Gilbert, 2008). Gallinat and Spaulding (2014) investigated the performance of children with DLD and LT children on tests measuring nonverbal cognition. A meta-analysis of 131 studies (138 children with DLD and 138 LT children) found that children with DLD scored .69 standard deviations below their LT peers on norm-referenced tests of nonverbal IQ, equating to approximately 10 points lower than LT children (Gallinat & Spaulding, 2014). Recently, Lancaster, Milne, Gordon, Emerson, and Selin (2024) conducted an analysis of the nonverbal IQ tests measured in Gallinat and Spaulding's meta-analysis. By categorising explicit and implicit elements of various IQ tests, they identified four main constructs: attention, pattern learning, receptive language and working memory. The authors

[2] The relationship between nonverbal IQ and the Grammatical Completion subtest was r = .13 (Lancaster & Camarata, 2019). This subtest assesses morphological recognition and understanding.

suggest that the emphasis on these constructs differs between nonverbal IQ tests and they suggest that researchers should be careful to select the right test for their needs.

Several longitudinal studies report unpredictable patterns of nonverbal IQ scores in children and adults with DLD. For example, Botting (2005) examined nonverbal IQ scores of 82 children from the MLS cohort (based on assessment data from several time points: seven years, eight years, 11 years and 14 years of age). Analysis showed a reduction of almost 23 points between mean nonverbal IQ scores assessed at the age of seven (mean score 105.6), and when participants were reassessed at 14 years of age (mean score 83.0).

There is limited research regarding nonverbal IQ in adults, yet the evidence to date demonstrates a different pattern to that seen in childhood. Tomblin, Freese and Records (1992) report a decline in mean nonverbal IQ scores in a group of 20 adults with DLD. When assessed in adulthood, their mean performance IQ was 90, in comparison to 98.5 in childhood. Returning to the MLS cohort, Botting (2020) assessed language, literacy and cognitive skills at the age of 24 years. Results revealed higher nonverbal IQ scores from those assessed in adolescence (mean nonverbal IQ scores of 86.4 at the age of 16 years and mean score of 99.9 at the age of 24). These results are similar to those reported by Clegg et al. (2005) in their longitudinal study following individuals with DLD from childhood to middle adulthood. Nonverbal IQ scores were similar in early childhood (9;11 years) and middle childhood (13;4 years). However, between middle childhood and early twenties (24;3 years), there was a drop of 14 points in l IQ scores (from 92.3 to 78). In middle adulthood (mean age 36;2), IQ scores were very similar to those observed in childhood. The authors account for the fluctuation in scores by offering several explanations including missing data in middle childhood assessments, or the possibility of practice effects, as the same nonverbal IQ measure was used at the last two time points.

Whilst Clegg et al.'s sample was much smaller (n = 17) than those in Botting's 2005 and 2020 studies, these results indicate a changeable pattern of nonverbal IQ in DLD.

Another issue that possibly contributes to inconsistencies in the nonverbal IQ scores reported above (and acknowledged by both Clegg et al. and Botting), is that different nonverbal IQ assessments were used in these studies at different time points. Lancaster et al. (2024) suggest that nonverbal IQ tests vary in the demands that they place on the individual, that is, some tests may have a strong working memory component whereas others have a high attentional component. If different tests are administered at different time points this could give rise to fluctuation in scores. Another possibility, highlighted by Botting (2020), it that cognitive skills in individuals with DLD could vary over time, in comparison to LT peers. Botting (2005) refers to an argument raised by Krassowski and Plante (1997) who suggest that fluctuating IQ scores in DLD could be representative of their ability at *a point in time* in that it reflects their *current* ability, arising from the inconsistent nature of the condition.

Lastly, the fact that children with DLD typically have lower scores in tests measuring cognitive abilities has been explored according to a theory of cognitive development known as 'mutualism'. Griffiths, Kievet, and Norbury (2022) analysed data from a longitudinal cohort study (n = 501), over three time points, between seven and 13 years of age. The theory of mutualism holds that nonverbal reasoning and language should operate in a reciprocal fashion, whereby good nonverbal reasoning skills potentially enhances vocabulary growth, and vice versa. Accordingly, Griffiths et al. (2022) suggest that poor language in DLD (in this case, vocabulary), may be the reason that children with DLD have low nonverbal reasoning scores. As predicted, their results showed that children with higher vocabulary scores exhibited more growth in nonverbal abilities, and vice versa. Interestingly, they found that the strength of

mutualistic coupling did not differ between LT children and those with DLD. These findings could be beneficial for children with DLD who have low nonverbal IQ scores, as the authors suggest that language interventions could support reciprocal development of both language and nonverbal reasoning skills. It is worth noting that this study assessed language in the context of vocabulary and the extent to which mutualistic coupling affects grammar in DLD is not known.

2.2 Language aptitude/language analytic ability

There is a logical transition from nonverbal IQ, as discussed in the previous section, and (foreign) language aptitude, which is discussed in this section. For example, a meta-analysis by Li (2016) reveals a significant correlation between performance on language aptitude tests and IQ (r = .64). Furthermore, it is noted that IQ and aptitude tests often measure the same components, such as working memory and vocabulary (Winckel and Dąbrowska, 2024). However, aptitude and nonverbal IQ are considered to be separate constructs and rather operate independently from each other (Li, 2016).

Foreign language aptitude (hereafter referred to as language aptitude), refers to specific cognitive abilities which are considered relevant to learning a foreign language (Carroll, 1981; Carroll & Sapon, 2002). The most frequently used measures of language aptitude are test batteries such as the Modern Language Aptitude Test (MLAT: Carroll and Sapon, 1959), and the Pimsleur Language Aptitude Battery (PLAB: Pimsleur, 1966). The MLAT comprises several subtests measuring distinct abilities, these are: phonetic coding (the ability to distinguish between sounds), grammatical sensitivity (grammatical functions between sentences), rote learning, and inductive learning (involving analogy) (Carroll, 1964, 1990). Of

these four abilities, two measures are considered relevant for grammar, these are grammatical sensitivity and inductive learning, which can be referred to as 'language analytic ability' (Li, 2015; Skehan, 2002).[3]

The 'Words in Sentences' subtest of the MLAT (Carroll and Sapon, 1959), is a grammatical sensitivity task which involves determining the relationship between words and how a word fits into a sentence. Participants are presented with two sentences, the key sentence has a word in capital letters, and their task is to select the word in the second sentence that plays the same role as the word in capitals in the key sentence. The Words in Sentences subtest does not use explicit grammatical terminology such as noun, verb, or adjective, but tests the ability to recognize the grammatical function of words between two sentences (essentially involving analogical mapping and metalinguistic awareness). The Language Analysis subtest of the PLAB (1966) consists of words and sentences in an artificial language and assesses the ability to translate novel sentences into the new language.

Language aptitude involves explicit language learning abilities (see meta-analysis by Li, 2015). For this reason, it is not considered relevant to L1 language – as discussed earlier, it is widely assumed that L1 grammar is acquired implicitly (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). However, recent studies in adult monolinguals demonstrate a relationship between language aptitude and grammar. For example, Dąbrowska (2018) reports a correlation of $r = .46$ between language aptitude (measured by the Language Analysis subtest from the PLAB), and sentence comprehension. Similar results are presented by Llompart and

[3] In this thesis, studies that explore the relationship between language aptitude and linguistic outcomes use the Language Analysis subtest from the PLAB, in the pilot study. In study two: the 'Sentence Pairs' task, modelled on the 'Words in Sentences' (grammatical sensitivity) subtest of the MLAT. In study three, the 'Matching Words' subtest of MLAT-Elementary is used for children.

Dąbrowska (2023), who assessed performance on two language aptitude tests in their study; the Sentence Pairs task, and Language Analysis. Their results are very similar to those reported by Dąbrowska (2018), with correlations ranging from r = .43 to r = .48 between language aptitude (Sentence Pairs and Language Analysis), and grammatical comprehension (measured by a picture selection task). In a more recent paper, Winckel and Dąbrowska (2024), report a robust relationship of r = .68 between the Sentence Pairs task and grammatical comprehension (a forced choice sentence comprehension task). A slightly smaller association was found by Prela, Dąbrowska, and Llompart (2022), who examined the relationship between language aptitude, using the Sentence Pairs task, and a grammaticality judgement task, with a correlation of r = .37. It is interesting to note that the aforementioned correlations are higher than the mean effect size reported in Li's (2015) meta-analysis (r = .31) between language analytic ability and L2 grammatical attainment).

Winckel and Dąbrowska (2024) propose a logical link between language analytic ability and grammatical knowledge. According to usage-based theories of language acquisition, the ability to understand and produce novel utterances involves processes of generalization and analogy. As such, grammatical sensitivity relies on explicit, metalinguistic knowledge, where individuals are required to compare form and meaning between two sentences. Analogical reasoning plays a significant role, and in fact, Carroll (1990) refers to the grammatical sensitivity task as a 'grammatical analogies test'. Furthermore, language analytic ability potentially aligns with usage-based models on the proviso that rule induction relies on identifying similarities between structures. This is aptly summarised by Goldwater (2017, p. 793):

'…grammatical constructions are relational categories in that they are represented by the abstraction of exemplar sentences' semantics and syntactic relations'.

If (as per the findings introduced above), language aptitude/language analytic ability is relevant to L1 grammar learning, then there is a possibility that it is also relevant to individuals with DLD.

2.3 Analogical reasoning

Usage-based models propose that children's language development relies on the ability to observe and deduce patterns, rules, and regularities from the input that they receive. In doing so, children generalise across grammatical constructions to form new abstract constructions of their own (Goldberg, 1995). This involves the process of analogy. In this section, discussion is firstly centred on terminology and prior research regarding analogical reasoning before focusing on language specifics of analogical reasoning and the role that it plays in DLD.

Analogical reasoning is a domain general cognitive ability which involves the ability to observe and identify: a) the relational similarities between two situations, and, b) use this information to generate further inferences (Gentner, 1988; Gentner & Smith, 2012). It is suggested that by comparing two entities, children can identify common relational information that may not be obvious to them if observed in isolation (Gentner & Namy, 2006). The process of analogical reasoning begins early on in development and it has been observed that children aged between two and four years of age are capable of solving basic problems through analogy (Crisafi & Brown, 1986). Successful analogical problem solving is defined as the ability to solve higher-order relations, for example, in a traditional 'A is to B' as 'C is to D' (A::B::C:D)

problem, this would refer to the process of solving the higher-order relation between A and B, applying this principle to C, and arriving at the solution of D (Goswami, 1989). Behrens (2017) explains that analogical mapping involves either perceptual similarity (that is, similarity between objects), or relational similarity (for example, when presented with several objects, by noticing that two items share similarities without being physically similar). It is proposed that in the early stages of development, children respond to objects through featural and perceptual factors, usually shape similarity (Gentner & Namy, 2006), yet later on in development they tend to emphasise relational similarity. This is referred to as a relational shift (Gentner, 1988).

Richland, Morrison, and Holyoak (2006) examined the underlying developmental factors contributing to analogical reasoning through a series of experiments with four groups of children aged: 3 and 4 years, 6 and 7 years, 9 to 11 years, and 13 and 14 years. To assess analogical reasoning, they designed a scene analogy task comprising analogy problems involving illustrated pictures which differed in terms of relational complexity and featural distraction. Here, the authors refer to relational complexity as a number of related 'arguments' or 'entities' (see Andrews & Halford, 2002). Featural distraction is described as the ability to inhibit a response to a feature that is possibly superficial, or one that potentially interferes due to semantic similarity (Richland et al., 2006). In their study, Richland et al. (2006) found that children as young as three and four years of age could identify simple analogical relations, yet their performance was affected by increased relational complexity and featural distraction. On the other hand, children in the nine to eleven age group were less likely to be affected by distraction, and very few children in the 13 to 14 age group made featural errors. Richland et al. (2006) interpret their results in accordance with Gentner's (1988) relational shift, suggesting that analogical reasoning ability improves with development. A later study using the same scene analogy task measured performance alongside working memory and inhibitory control,

both of which are considered to place demands on analogical reasoning performance. Results showed that working memory capacity predicted analogical reasoning performance, yet inhibitory control was not significant (Simms, Frausel, & Richland, 2018).

As highlighted earlier, analogy plays a key role in usage-based models of language acquisition (Behrens, 2009; Bybee, 2010; Gentner & Christie, 2010; Gentner & Namy, 2006; Tomasello, 2003). In their everyday lives, children's language exposure (and use of language) involves similar utterances with minimal variation – for example, item based schemas such as: *I wanna X*, *Let's X*, *Where's-the-X* (Tomasello, 2003, p. 297). In observing patterns across utterances that have a similar structure (in both linguistic form and the level of communicative function), children create new abstract constructions of their own (Tomasello, 2003). This process of generalisation involves analogical reasoning (Behrens, 2017; Tomasello, 2003). For a visual example of a proportional analogy problem, see Figure 2. Generalisation as a proportional analogy problem(Dąbrowska and Blake, 2022), in part two of this thesis.

Gentner & Christie (2010) propose that the relationship between analogical reasoning and language is bidirectional. This is interesting from the perspective of language difficulties - if language depends on analogical reasoning ability, as described above, then poor analogical reasoning ability could potentially have an effect upon language development. Similarly, stronger analogical reasoning skills should support language ability. There is some evidence to suggest that analogical reasoning could be impaired in people with DLD. Nippold, Erskine, and Freed (1998) explored analogical reasoning in children with and without DLD using a verbal analogical reasoning task (using an A::B::C:D format), a perceptual reasoning task (involving selecting the correct design), and functional analogical reasoning (story and object transfer problem). In this study, children with DLD performed below LT controls on all three

analogical reasoning tasks. However, when nonverbal intelligence was controlled for, the differences between the groups were not maintained, therefore highlighting the close relationship between nonverbal intelligence and analogical reasoning.

Aside from Nippold et al. (1998), most research regarding analogical reasoning in DLD has been conducted in a series of comprehensive studies by Leroy, Krzemien and colleagues. Collectively, these studies provide evidence of analogical reasoning deficits in children with DLD in both linguistic and non-linguistic domains (Krzemien, Jemel, & Maillart, 2017, Krzemien, Maillart, Parisse, & Leroy, 2019, Krzemien, Thibaut, & Maillart, 2020; Leroy, Parisse, & Maillart, 2012, Leroy, Maillart, & Parisse, 2014). Overall results suggest that children with DLD are dependent upon similarities between structures and rely on perceptual cues to aid analogical mapping. As explained by Leroy et al. (2012), children may find it easier to create a new utterance when they can draw upon a common linguistic cue between utterances, for example (‘*He sees a X’* construction), whereas it is more difficult when the linguistic cue is less obvious, such as in a Subject-Verb-Object construction (e.g., ‘*I want a X’*, Leroy et al., 2012 p. 393). Krzemien et al. (2019) suggest that the reason children with DLD find analogical reasoning difficult could be attributed to limitations in processing capacity. When perceptual cues are present, then processing load is reduced (see Im-Bolter, Johnson, and Pascual-Leone, 2006). In addition, Krzemien et al. (2020) propose that poor inhibition may play a contributory factor to analogical reasoning deficits in DLD, as children with DLD are unable to inhibit responses to perceptual distraction.

2.4 Processing speed

Processing speed is defined as the speed at which people perform basic cognitive processes and it is significantly related to performance in cognitive tasks (Kail & Ferrer, 2007). A considerable amount of literature has shown that people with DLD have limitations in processing speed, as will be discussed in this section.

Early research by Kail (1994) presents a 'generalised slowing' hypothesis, which suggests that response times (RTs) in children with DLD increases in a linear fashion across tasks (assessed by picture naming tasks). In a meta-analysis, Kail analysed response time data from five studies, comparing performance in children with and without DLD. The results demonstrated that children with DLD (age 6;1 - 13;1 years) were 33% slower in RTs than age-matched LT children. Kail's meta-analysis was replicated by Windsor and Hwang (1999), investigating data from three studies (age 6;10 – 11;6 years), revealing that children were one fifth slower than their peers. As a further extension of these results, Miller, Kail, Leonard, and Tomblin (2001) examined RTs in children (9 years of age), with and without DLD. Children with DLD were slower across linguistic and non-linguistic tasks. The participants were assessed again when they reached 14 years of age. Results showed slower RTs in the DLD group in comparison to LT peers, demonstrating that slower processing speeds continue into adolescence (Miller, Leonard, Kail, Zhang, Tomblin, & Francis, 2006).

Research indicates that processing speed is associated with working memory; if one is able to process information quickly, then this will ultimately free up working memory capacity (Leonard, 2014a). Leonard, Ellis Weismer, Miller, Francis, Tomblin, and Kail (2007) investigated processing speed, working memory, and language in 14-year-olds (126 LT

participants and 78 with DLD). They found that working memory was the most significant predictor of language but processing speed also played an important role (these two predictors together accounted for 62% of the variance in composite language test scores). Although closely related, it is interesting to note that processing speed and working memory were not found to be interchangeable.

A more recent study by Witherstone (2024) adds to this body of evidence, showing that children with DLD require more time to process information. This study examined processing speed related to inflectional morphology in English in children (mean age 8;5 years), with and without DLD. Children's sensitivity to inflectional morphology (past tense, third person singular and plural), was measured in real time by the child listening to a target word and then pressing a button when they heard the word within a sentence. The words were either grammatical or ungrammatical and the speed of sentences was presented at a baseline speed (166 words per minute), or 30% slower than baseline. When children with DLD listened to the sentences at typical conversational rate, their RTs were the same regardless of whether the sentence was grammatical or not. However, when given longer to process the sentences, response times and accuracy improved in children with DLD. However, they showed persistent deficits with regular past tense inflections, regardless of the speed of presentation.

Research regarding processing speed in adults with DLD is less prolific. Miller and Poll (2009) recruited a small group of college students to ascertain whether processing speed in adults with DLD differs to that in children. In this study, they found that the DLD group had slower response times in linguistic and non-linguistic tasks, and individuals with higher language scores also had faster RTs. Furthermore, processing speed has been linked to working memory in a study investigating memory trace decay, indicating that adults with DLD take

longer than their LT peers to detect target words in a sentence (Poll, Watkins, & Miller, 2014). Lastly, Poll, Miller, and Van Hell (2016) investigated processing speed limitations in adults with DLD, in conjunction with sentence repetition accuracy. The sentence repetition task comprised of short or long sentences that were argument laden (e.g., long argument sentence: *The gentleman amused the earl with his description and impressed his servant but his cat frightened the mouse*), or adjunct laden (e.g. short adjunct sentence: *the author cheated the agency at first and bragged in public*). Processing speed was measured by a truth value judgement task and a word detection task. The results revealed an interaction between processing speed, argument status, and sentence length. The authors suggest that because the long argument condition contained more complex verb argument structures than other sentences, they required more cognitive resources. Consistent with Kail's generalised slowing hypothesis (Kail, 1994), adults with DLD required more processing time overall.

Evidence to date has been integrated in a meta-analysis by Zapparrata, Brooks, and Ober (2023). They examined whether slower processing speeds typically seen in DLD are domain general. Prior studies suggest that inhibitory control and cognitive flexibility are potentially a contributory factor (Pauls & Archibald, 2016), and due to this, the authors did not include complex response time (RT) tasks. Their analyses focused on studies that used simple RT tasks (for instance, motor responses to a stimulus), choice RT tasks (forced choice evaluation of simple stimuli), naming tasks (verbal naming of visual stimuli), and interference control tasks (attentional control with congruent or baseline conditions to reduce inhibitory demands). Overall results revealed an effect size of $g = .47$, with longer RTs in DLD groups in comparison to age-matched neurotypical peers. This response was evident across task types, stimulus types, stimulus modalities and response modalities, although larger effects were shown in tasks requiring a verbal response.

Collectively, the overall evidence regarding slow processing speed in people with DLD has led to the recommendation that it should be classified as a clinical marker for DLD (Park, Miller, & Mainela-Arnold, 2015; Park, Miller, Sanjeevan, Van Hell, Weiss, & Mainela-Arnold, 2020).

2.5 Memory

This section outlines research regarding working memory, declarative, and procedural memory and how these processes support language in individuals with and without DLD.

2.5.1 Working memory

Working memory is a limited capacity system enabling the temporary storage and processing of information necessary for complex cognition (Baddeley, 2000; Baddeley & Hitch, 1974; Baddeley, Hitch, & Allen 2021). Although there are a number of different models and theories of working memory (for example - but not limited to - Cowan, 2005, 2022; Engle, Tuholski, Laughin, & Conway, 1999), this thesis focuses on the model by Baddeley and colleagues.

The multi-component model of working memory formed by Baddeley and Hitch (1974) incorporates the central executive (CE), which is central to the model, and coordinates performance between tasks, attention control, and inhibition (Baddeley, 1996). The central executive is supported by two so-called slave systems, the phonological loop (which processes verbal and acoustic information), and the visuospatial sketchpad (for processing visuospatial information, i.e., visual, spatial and kinaesthetic components). In 2000, a further component

was added to the model, the episodic buffer (Baddeley, 2000), providing a limited capacity interface between the phonological loop, the visuospatial sketchpad, and long-term memory. The episodic buffer is controlled by the central executive (CE) which can retrieve information through conscious awareness, reflect upon that information, and manipulate and modify it. It is suggested that the main structure of working memory model, i.e., the central executive and two slave systems, are developed by at least six years of age, and increase in capacity from four years of age to early adolescence (Gathercole, Pickering, Ambridge, & Wearing, 2004).

Previous research shows that working memory is a significant predictor of academic success, and individual differences in working memory capacity influence long-term educational outcomes. For example, working memory capacity measured at the age of five years of age is a significant predictor of literacy (reading comprehension and spelling), and numeracy skills (mathematical reasoning and number operations), six years later (Alloway & Alloway, 2010). Whilst it has been proposed that intelligence and working memory are related constructs (Ackerman, Beier, & Boyle, 2002, 2005; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004, Kane, Hambrick, & Conway, 2005), an interesting finding from Alloway and Alloway's study is that working memory accounted for more statistical variance than IQ, with the authors suggesting that "…working memory is not a proxy for IQ, but rather represents a dissociable cognitive skill with unique links to learning outcomes" (Alloway & Alloway, 2010, p.26).

Evidence shows that there are adverse effects of poor working memory capacity upon educational achievement. This is demonstrated in a longitudinal study in the United Kingdom, assessing working memory and academic attainment. Assessment was based on national curriculum tests in a group of children at the ages of seven and fourteen years, over

two time points. At seven years of age, those children who did not reach the required attainment levels in mathematics and English (reading, reading comprehension and spelling), performed poorly in working memory assessments. When reassessed at the age of fourteen, working memory ability correlated with maths and science assessments, but only weakly with English curriculum assessments (Gathercole, Pickering, Knight, & Stegmann, 2004; Gathercole & Pickering, 2000). Similarly, Alloway, Gathercole, Kirkwood, and Elliott (2009) found that low working memory skills in primary school children correlate with poor outcomes in reading and mathematics. Ultimately, children with working memory difficulties struggle to catch up with their peers without early intervention and are therefore at risk of educational underachievement.

The association between working memory and language has been examined extensively in the context of DLD (for reviews, see: Archibald, 2017; Henry & Botting, 2017; Montgomery, Magimairaj, & Finney, 2010). The relationship between verbal short-term memory (or verbal working memory, referring to the function of the phonological loop as per Baddeley et al.'s model), and impaired working memory in children with DLD is typically assessed by simple-span tasks such as nonword repetition tasks. Nonword repetition involves a child listening to and repeating a 'pseudoword' (Baddeley, 2003), to assess learning and retention of novel words (Baddeley, Gathercole, & Papagno, 1998). An early, well-known study by Gathercole and Baddeley (1990), examined the memory capacity of a small group of children with DLD (six children), and two small control groups matched on vocabulary, reading ability and nonverbal intelligence. Working memory was assessed by performance in a nonword repetition task which revealed that children with DLD had reduced phonological working memory. The DLD group were poorer at repeating three-syllable and four-syllable nonwords, relative to one and two syllable nonwords, in comparison to the control groups.

There is extensive literature reporting phonological working memory deficits in DLD (Archibald, Oram Cardy, Joanisse, & Ansari 2013; Archibald & Gathercole, 2006a, 2007; Bishop, North, and Donlan, 1996; Gathercole, Willis, Emslie, & Baddeley, 1992; Gathercole & Baddeley, 1990; Jackson, Leitão, & Claessen, 2016, Jackson, Leitão, Claessen, & Boyes, 2019a). Additional to this evidence, a meta-analysis examining nonword repetition in children found that children with DLD score 1.27 standard deviations below LT children on nonword repetition tasks. Performance was not associated with age, representing significant and persistent phonological memory deficits in children with DLD (Graf Estes, Evans, & Else-Quest, 2007).

Montgomery and colleagues have carried out a series of studies to investigate the role of working memory in sentence comprehension, focused on children with DLD (Montgomery, 1995, 2000a, 2000b,; Montgomery & Evans, 2009). Collectively, these studies highlight that children with DLD have difficulties in comprehending long, complex sentences due to the demands of working memory. Therefore, difficulties are apparent when processing demands are high. Montgomery, Gillam, and Evans (2021) propose new model of working memory and sentence comprehension to explain language difficulties in DLD. Their model suggests that certain abilities (fluid reasoning, controlled attention, and language in long-term memory), are mediated by working memory which supports children's sentence comprehension. The authors review data from 234 children (mean age of 9;6 years), with and without DLD (Montgomery, Gillam, Evans, & Sergeev, 2017; Montgomery, Evans, Fargo, Schwartz, & Gillam, 2018), to determine the relationship between cognitive processes and sentence comprehension. Their overall results show that working memory plays a significant role in syntactic comprehension in children with DLD. However, Montgomery et al. (2021) suggest that there is not a clear-cut distinction between LT children having good working memory capacity, and vice versa in

DLD. Rather, they suggest that the situation is more complex. Where LT children use pattern recognition and chunking of multiword templates (stored in long term memory), to comprehend both canonical sentences and noncanonical sentences[4], in the DLD group, pattern recognition did not support comprehension. Instead, children with DLD used chunking strategies to comprehend canonical sentences, whereas noncanonical sentences were processed on a word-by-word basis. Montgomery et al. (2021) propose that poor (or non-existent) language knowledge in long-term memory means that children with DLD were not able to use chunking strategies to support their comprehension. To process difficult sentences, they resort to controlled attention, which places significant strain on their working memory capacity.

Most studies examining working memory difficulties in DLD are focussed on the verbal domain and there is less research exploring other elements such as visuospatial working memory (see Henry & Botting, 2017; Montgomery et al., 2010). However, research findings to date are mixed. Evidence for visuospatial working memory deficits is shown in a meta-analysis by Vugs, Cuperus, Hendriks, and Verhoeven (2013). Differences between LT and DLD children are reported in visuospatial storage (d = .49), and in visuospatial CE (d = .63). A later longitudinal study by Vugs, Hendriks, Cuperus, Knoors, and Verhoeven (2017) found that children with DLD performed below LT children in all areas of working memory (verbal storage, verbal central executive, visuospatial storage and visuospatial CE), at time point one, with effect sizes ranging from d = 1.03 to d = 2.30 (at four and five years of age). Yet, when they were reassessed at time point two (at seven to eight years of age), visuospatial components of working memory were not significantly different when age and intelligence were factored

[4] The children heard two canonical sentences, for example, subject-verb-object: (*The square had changed the bed under the very new dry key*), and subject relatives: (*The watch that had hugged the truck behind the kite was bright*). They also heard two noncanonical structures: verbal *be* passives (*The watch was bumped by the wheel near the very bright clock),* and object relatives: (*The chair that the bread had splashed under the square was new)* (Montgomery et al., 2017, p. 2607).

in. Contrary to these results, some studies do not find evidence of visuospatial memory deficits in DLD (Alloway & Archibald, 2008; Archibald & Gathercole, 2006b; Ellis Weismer, Davidson, Gangopadhyay, Sindberg, Roebuck, & Kaushanskaya., 2017).

2.5.2 Declarative and procedural memory

Two long-term memory systems support language development: declarative memory and non-declarative memory, of which procedural memory is a sub-system (Squire, 2004; Squire, & Zola-Morgan, 1988). There is broad consensus that both memory systems are central to language learning: declarative memory supports explicit, knowledge-based learning, described as 'knowing that', whereas nondeclarative procedural memory underlies implicit motor, perceptual and cognitive skills, described as 'knowing how' (Eichenbaum, 2012; Gabrieli, 1998; Squire, 1982). The distinction between the two memory systems has been explored through clinical studies, such as that of Henry Molaison, (HM), who developed anterograde amnesia after the removal of his temporal lobes to treat severe epilepsy. Consequently, HM was unable to retrieve declarative memories, although his procedural memory ability remained unimpaired. HM showed progress in motor tasks such as the pursuit rotor task and bimanual tracking task but had no memory of having done so (Corkin, 1968). Similarly, evidence from clinical studies of amnesiac Korsakoff syndrome patients, with neural damage following severe alcoholism, show normal ability in tasks such as mirror-image reading, but with no recollection of the words read (Cohen & Squire, 1980). Procedural memory impairment is found in patients with neurodegenerative basal ganglia diseases such as Huntington's disease and Parkinson's disease (Saint-Cyr, Taylor, & Lang, 1988).

Some cognitive psychologists have interpreted previous research findings regarding declarative and non-declarative memory as being central to specific neural regions. As such,

the declarative memory system involves the hippocampus, perirhinal, entorhinal and parahippocampal cortices (Eichenbaum, 2012; Squire & Wixted, 2011). The medial-temporal region interacts with the neocortex for the encoding, storage and retrieval of declarative memories (Gabrieli, 1998), and this region is central to acquiring, retaining, and retrieving factual based knowledge (Cohen & Squire, 1980). Procedural memory is subserved by the frontal/basal-ganglia circuits (Eichenbaum, 2012a; Squire, 2004). The basal ganglia are a set of interconnected, sub-cortical nuclei, which include the striatum (caudate nucleus and putamen in primates), globus pallidus, subthalamic nucleus, and substantia nigra (Albin, Young, & Penney, 1989; Konopka & Roberts, 2016; Redgrave, Prescott, & Gurney, 1999). The basal ganglia receive input from the cortex and feedback to this region is provided by a series of pathways that loop through the thalamus (Konopka & Roberts, 2016). The function of the putamen is to regulate motor functions whilst the caudate nucleus is fundamental to certain areas of cognition (Ullman, 2004). The basal ganglia region supports motor, perceptual and cognitive skills (Gabrieli, 1998), which include skill and habit-based learning as well as the control of established sensory to motor adaptations, such as adjusted and conditioned reflexes (Eichenbaum, 2012).

Procedural learning of a motor skill or cognitive skill occurs following repeated exposure and practice (Cohen & Squire, 1980), which results in autonomous execution (Graybiel & Grafton, 2015). For example, procedural memory underlies tasks such as riding a bike, playing the piano, driving a car, or learning to touch type on a keyboard. All skills begin as an explicit, controlled processes, but with repetition and practice, they become automatic and implicit, and not available to conscious access (Schmidt, 1990). (This will be discussed further in section *3.1 Skill acquisition theory and automatization*).

Returning to the 'KE family' (as discussed in section: *1.3.2 Behavioural genetic research*), procedural learning deficits have also been said to underlie the families' language deficits, linked to the *FOXP2* gene. Watkins et al. (2002) suggest that verbal and non-verbal deficits observed in the affected members of the KE family could be linked to difficulties in sequencing movement, or procedural learning impairment. Animal research provides an example of this hypothesis. Schreiweis et al. (2014) compared the behaviour of two groups of mice, 1) with a humanised version of the *FOXP2* gene, and 2) wild type littermates. In navigational maze experiments, mice with the *FOXP2* gene favoured procedural learning strategies when both memory systems were available. The authors suggest that their experimental findings can be interpreted in terms of the relationship between declarative and procedural learning in human language learning. One suggestion is that, in humans, proceduralization could potentially enable probabilistic language learning through the process of chunking speech and language units into sequences. This could hypothetically free declarative resources at an earlier stage (Schreiweis et al., 2014).

Some researchers have contextualised language according to dual-route paradigms. For example, Paradis (2004, 2009), proposes that the declarative-procedural distinction is explained by the declarative memory system underlying explicit, metalinguistic knowledge, which is learned consciously. On the other hand, procedural memory underlies linguistic competence, where learning is unconscious, implicit and automatic. Paradis (2009) states that the two systems operate alongside each other; as learners become more proficient, they become less dependent on metalinguistic knowledge and more reliant on implicit processing.

Ullman and colleagues' declarative/procedural (DP) model of language learning (Ullman, 2001, 2004, 2016), proposes that declarative memory subserves the mental lexicon,

and the procedural memory system supports grammar learning. Later, Ullman and Pierpont (2005), suggested that individuals with DLD have neural deficits in the basal ganglia, which is central to procedural memory. This theory, known as the Procedural Deficit Hypothesis (PDH), has been reviewed to incorporate the anatomical basis of the PDH, and, in a recent paper, Ullman, Earle, Walenski, and Janacsek (2020) revise their terminology to reflect this, thus the PDH is now termed: 'Procedural (circuit) Deficit Hypothesis'. The PDH represents a unified theoretical account, incorporating neurobiological underpinnings of developmental disorders such as DLD, as well as associated language disorders such as motor-speech disorders and dyslexia (Ullman, et al. 2020). More recently, results from a systematic review and meta-analysis provide evidence of anomalies in the basal ganglia in DLD (Ullman et al., 2024). According to the PDH, people with DLD have deficits with grammatical and phonological skills, but not with vocabulary which is dependent upon the declarative system. The theory suggests that declarative memory is intact and can compensate for procedural deficits which could involve memorizing complex forms as chunks in declarative memory, as well as the explicit learning of grammar rules (Ullman & Pierpont, 2005; Ullman & Pullman, 2015). Ullman et al. (2024), state that there are few structural abnormalities in the medial temporal lobe, adding further evidence to the fact that declarative memory is unimpaired in DLD.

The PDH has motivated a wide body of research in DLD, evidence of which will be discussed later in this thesis (section: *2.6.1 Implicit learning: procedural learning deficits*). Additionally, many studies examine intervention support as motivated by the PDH, with the aim of exploiting strengths in declarative memory to support language skills in people with DLD (Balthazar, Ebbels, & Zwitzerlood, 2020; Calder, Claessen, Ebbels, & Leitão, 2021; Ebbels, 2007; Ebbels, Maric, Murphy, & Turner, 2014).

2.5.3 Limitations regarding dual-processing models

The previous section provides evidence of the proposed relationship between language and dual-based paradigms. Although this thesis does not argue that these systems are not relevant to language, theories delineating the lexicon and grammar do not align well with nonmodular approaches to language acquisition. According to usage-based models, language does not comprise of different units, for example, a lexical module that operates separately from syntax (Behrens, 2009; Bybee, 2010). Instead, usage based theories support a lexicon-grammar continuum (Fillmore, 1988, 1993; Goldberg, 1995). Empirical evidence supports the relationship between grammar and the lexicon in the early stages of language development, with significant correlations observed between the two. In other words, the emergence of grammar depends upon having a sizeable vocabulary (Fenson et al., 1994; Bates & Goodman, 1997). Behavioural genetic studies further substantiate the link between vocabulary and grammar, as demonstrated in large cohorts of same-sex twins of two and three years (Dionne, Dale, Boivin, & Plomin, 2003), and four and a half years of age (Hayiou-Thomas, Kovas, Harlaar, Plomin, Bishop, & Dale, 2006).

Dual based theories such as Ullman's Declarative/Procedural model (Ullman, 2001, 2004, 2016), and the PDH have put forward explicit hypotheses for conditions such as DLD and other neurodevelopmental conditions (Ullman et al., 2020; Ullman & Pierpont, 2005). Whilst many studies have assessed grammatical deficits in DLD from the PDH perspective, Dąbrowska (2010a) suggests that the picture is not quite so clear cut. Dąbrowska does not dispute the link between the PDH and procedural memory but raises the issue that the PDH theory does not correspond well to other languages other than English, particularly in languages that have more complex morphological mechanisms. A similar suggestion has been made by Divjak, Milin, Medimorec, and Borowski (2022), who point out that the grammar/lexicon

divide has ultimately gained traction due to generative based language research that has predominantly focused on the English language, which makes it difficult to see how the two are interconnected. This is shown in a study by Kidd and Kirjavainen (2011) who examine the role of procedural and declarative memory in past tense morphology in Finnish-speaking children (mean age 5;2). The results favour a single-route approach with a strong correlation between past tense morphology and vocabulary development, supported by declarative memory (and no evidence of the relationship between morphology and procedural memory). Ultimately, the authors do not dispute the relevance of the PDH but suggest that the theory is not well aligned to languages with complex inflectional morphology, such as Finnish.

According to Dąbrowska (2010a), children's language acquisition does not align specifically with declarative memory subserving lexical knowledge and procedural memory supporting grammar. Alternatively, Dąbrowska proposes that both explicit and implicit processes are involved in language learning and grammatical development, although they interact at different stages and in differing degrees. This viewpoint is substantiated by Krishnan et al. (2016) who state that language learning is not a unitary process that occurs in either declarative or procedural memory. The authors portray a language learning network where different brain regions are represented as 'key hubs' which interact in language learning (for a detailed illustration, see Krishnan et al., 2016: figure 1., p.704).

Llompart and Dąbrowska (2020) discuss implicit and explicit memory in their work with LT adults. This study incorporated two groups of adults with varying academic attainment, to ascertain how implicit and explicit memory correlate with linguistic outcomes. Their results show a robust relationship between language (i.e., measures of grammar, vocabulary and collocational knowledge), and explicit memory (phonological short-term memory and cross-modal associations, measured by a paired-associates task). However, no

relationship was found between implicit learning (using the SRT task) and linguistic performance. The relationship between explicit memory and both grammar and vocabulary aligns with usage-based models and offers contrasting evidence to the assumption that L1 grammar develops solely through implicit processes.

More recently, Divjak et al. (2022) examined declarative and procedural memory by measuring speed, accuracy and consistency of judgement across different aspects of linguistic knowledge in Polish: morphology (case and aspect), syntax (subordination), and lexical semantics (collocations). Participants were assessed in two conditions; one where only grammaticality of stimuli was measured, and a second condition where participants completed a digit span task at the same time as responding to grammaticality judgements of stimuli, to inhibit access to declarative memory. The results revealed that collocations were mostly supported by declarative memory, whereas aspect, case and subordination rely on procedural memory. The authors suggest that traces of declarative memory and procedural memory are employed across different types of linguistic knowledge. Ultimately the results support a divide between lexicon and grammar, but the authors suggest that two systems collaborate and operate according to a continuum, and in a 'graded fashion', which is in line with usage-based principles (Divjak et al., 2022).

Similarly, McCauley (2020) proposes a single-system theory of language, as opposed to a dual-system model. In line with usage-based accounts, McCauley suggests that 'lexical and grammatical knowledge are taken to form a continuum of learned units varying in abstractness and complexity' (McCauley, 2020, p.1). In an earlier paper, McCauley and Christiansen (2019) present a computational model which processes language in an incremental fashion, centred on chunking and statistical learning. The authors are currently

working towards further development of this model relating to developmental language disorders (McCauley, 2020; McCauley & Christiansen, 2019).

This section provides evidence from studies which propose that language does not operate according to a sharply defined split between lexicon and grammar. Instead, usage-based approaches offer a contrasting theory in that they do not delineate between implicit/explicit processes and syntax and grammar. This is highlighted by Behrens (2009):

> '…usage-based approaches do not make a distinction between (universal) core syntax and other properties of language. Rather, all linguistic phenomena are considered to be form-function correspondences. They may differ in size (from morphemes to idioms) and in degrees of abstractness or schematicity and productivity' (Behrens, 2009, p. 389).

2.6 Implicit learning

The ability to track distributional patterns in the environment has been studied in humans and other species since the 1900s (for a review, see Christiansen, 2019). To date, research has largely been published under two separate literatures: implicit learning and statistical learning, but without much interaction between the two (see Christiansen, 2019; Perruchet & Pacton, 2006). Early studies by Reber (1967), referred to implicit learning as learning that occurs without awareness (Reber, 1996). Implicit learning is usually measured by performance on artificial grammar learning tasks or probability learning procedures (Reber, 1967, 1996), serial reaction time tasks (Nissen & Bullemer, 1987), Hebb repetition tasks (Hebb, 1961), contextual cueing (Chun & Jiang, 1998), or the Weather Prediction Task (Knowlton,

Squire, & Gluck, 1994). Statistical learning is a domain general capacity in which humans (and other species) extract distributional patterns and statistical regularities from the environment (Isbilen & Christiansen, 2022). Therefore, statistical learning explores the way in which individuals learn and extract structures from the input (Aslin, 2017), usually by exposure to patterns of stimuli which often involve visual or auditory paradigms. Statistical learning can be purely implicit and/or involve explicit knowledge (Batterink, Reber, Neville, & Paller, 2015).

The umbrella term 'implicit statistical learning' (ISL) has been suggested as a way of unifying the literatures of implicit learning and statistical learning (Conway & Christiansen, 2006). Generally, most research in DLD tends to examine ISL from the perspective of a statistical learning deficit, or implicit learning focused on a procedural learning deficit (typically centred on brain systems), or both. As per the recommendation by Conway and Christian (2006), this thesis adopts the term 'ISL' to unify the two, although for the purpose of understanding, this section discusses the literature relating to procedural memory and statistical learning separately, where possible. Therefore, this section begins with an overview of the literature pertaining to procedural learning and the PDH (Ullman, 2004; Ullman et al., 2020; Ullman & Pierpont, 2005), followed by research regarding statistical learning. Then, ISL is discussed from results reported by several meta-analyses, before considering current issues regarding the relationship between ISL and language.

2.6.1 Implicit learning: procedural learning deficits

A large body of literature has investigated the possibility of impaired procedural memory in individuals with DLD, according to the PDH (Ullman et al., 2020; Ullman & Pierpont, 2005). One of the most commonly used measurements of procedural learning is the

serial reaction time (SRT) task (Nissen & Bullemer, 1987). In this task, a sequence of stimuli are presented in one of four locations on a computer screen and participants are required to respond to the stimulus as quickly as possible, by pressing corresponding keys on a computer keyboard. The stimuli follow a repeated sequence, although participants are not consciously aware of this. Performance is measured by improvement in accuracy or reaction times. Learning reflects faster response times in sequenced trials than random trials, whereas a smaller (or absence of) difference between sequence and control blocks would indicate that little or no learning had occurred. However, prior research has yielded differing results. Some studies have found an association between poor performance on the SRT task and procedural learning in line with the PDH, and yet others do not find any differences at all. Evidence is discussed accordingly.

West, Vadillo, Shanks, and Hulme (2018) investigated several measures of implicit learning in a large sample of 101 children (mean age: 8;1 years), in relation to the PDH, namely, the SRT task, Hebb task, and a contextual cueing task. Their results show that measures of procedural memory correlated weakly (or non-significantly) with language, literacy and arithmetic, and with each other, raising issues regarding the reliability of these measures. In particular, the test-retest reliability for the SRT task was extremely low (r = -0.001), suggesting no consistency in participant performance across testing session, indicating that the SRT task may not be a sensitive enough measure to capture individual performance consistently over time. In response to this study, Conway, Arciuli, Lum, and Ullman (2019) raised several queries. The first pertains to the sample used by West et al., which comprised a representative sample of children (which they interpret as LT children). They suggest that this is problematic because the PDH implies deficits in children with atypical language development. Furthermore, Conway et al. suggest that measurements of procedural learning were not entirely

valid for evaluating the PDH hypothesis. They raise the fact that tasks such as the Hebb repetition task or the contextual cueing task do not specifically tap into the basal ganglia region. Additionally, Conway et al. (2019) state that West et al. measured the difference between higher and lower probability sequences throughout the SRT task, including the early stage of the task where learning may not yet be stable. Krishnan and Watkins (2019) agree with reliability issues raised by West et al. (2018). They also suggest that there is not a clear-cut delineation between the declarative/procedural system (as discussed in the previous section) and propose that paradigms such as the SRT task often tap into both systems.

Evidence of procedural learning deficits in DLD are shown by Tomblin, Mainela-Arnold, & Zhang (2007), who assessed visuo-motor sequence learning using a SRT task in adolescents with and without DLD. Their data show that the DLD group were slower to learn sequences than age-matched peers (performance in the later pattern blocks were similar in both groups). Further investigation established that learning in the SRT task was related to grammatical impairments, but not vocabulary, which the authors interpret in support of the PDH. Using the same SRT task as Tomblin and colleagues, Hsu and Bishop (2014) investigated motor skill learning in three groups of children aged between seven and eleven years (children with DLD, age-matched LT children, and a group of children matched on grammatical ability). Tasks included an implicit verbal sequence learning task (measuring the Hebb effect[5]), and two motor tasks (deterministic SRT task, to measure sequential learning, and the pursuit rotor task to measure sensory motor learning). Overall, findings varied - the DLD children exhibited slower learning rates in the SRT task in comparison to age-matched controls, and similar levels to younger, grammar matched children. The DLD group also

[5] According to The Hebb effect (Hebb, 1961), serial recall is improved when a sequence is repeated several times in conjunction with a random sequence.

demonstrated poor implicit learning in the verbal sequence learning task, measuring the Hebb effect. However, in the pursuit rotor task, children with DLD performed as per the group of age-matched children, and better than children matched for grammar ability. The authors suggest that poor sequence learning, rather than overall weak procedural learning, could potentially underlie the difficulties in DLD (Hsu & Bishop, 2014).

Further studies supporting the PDH hypothesis include Lum, Gelgec, and Conti-Ramsden (2010), and Lum, Conti-Ramsden, Page, and Ullman (2012). Lukács and Kemény (2014) demonstrate poor sequential learning in children with DLD, in a deterministic SRT task and artificial grammar learning task, but similar performance to age-matched controls in a weather prediction task (nonsequential probabilistic categorization task). More recently, Jackson, Leitão, Claessen and Boyes (2020) explored memory system profiles in children (5 to 8 years of age) with and without DLD – namely, working memory, declarative and procedural memory (using an audiovisual SRT task). Overall findings suggest that procedural learning deficits (measured by an SRT task), were accounted for by deficits in working memory. On the other hand, children with DLD performed similarly to the LT group in visual declarative memory (i.e., information reliant on the visuospatial sketchpad), but not in verbal declarative memory. The authors conclude that although their results offer support for the PDH, they do not account for a core procedural deficit in DLD.

In contrast to earlier findings, some studies do not report an ISL deficit in DLD. For example, Lum and Bleses (2012) used a deterministic SRT task to explore the relationship between grammar and procedural memory, showing similar performance in LT and DLD children. It should be pointed out that the sample size in this study was relatively small, i.e., 13 children with DLD and 20 LT children. In view of mixed results relating to the PDH, Clark

and Lum (2017) investigated the relationship between procedural learning (using a deterministic SRT task), and measures of grammar, reading and motor functioning. The sample was a group of 20 children with DLD, and 20 age-matched LT children, seven to 13 years of age. They hypothesised that the relationship between procedural memory and grammar may be more closely aligned to processing speed rather than accuracy. This hypothesis was found to be true in LT children. However, in the DLD group, no association was found between procedural learning and grammatical processing speed, reading or motor skills.

Inconsistent findings are reported in a number of studies conducted by Gabriel and colleagues. For example, Gabriel, Maillart, Guillaume, Stefaniak, and Meulemans (2011), found that both LT and DLD children showed similar levels of sequence learning in a probabilistic SRT task. Similarly, Gabriel, Stefaniak, Maillart, Schmitz, and Meulemans (2012) found that children with DLD were able to learn a deterministic sequential task, even though their performance was slower and less accurate than their LT peers. Children with DLD were more successful in sequential learning when they received more exposure to the SRT task. In a second experiment, a touch screen was introduced in place of a keyboard to eliminate potential confounding results of motor/cognitive deficits, but this did not change performance. In a successive study, Gabriel, Maillart, Stefaniak, Lejeune, Desmottes, and Meulemans (2013) suggest that variability in findings using the SRT task could be due to the structure of sequences used. For example, the authors note that an 8-element second-order conditional sequence was used by Gabriel et al. (2012), while Tomblin et al. (2007), and Lum et al. (2010, 2011) used a first-order conditional (FOC) 10-element sequence. As explained by Gabriel et al. (2013), a 10-element sequence with four locations would mean that some locations would occur more frequently and therefore sequences could potentially be learned on a probability basis. On the other hand, in a second-order conditional (SOC) sequence, predictive first-order information is

not available. Therefore, in a FOC sequence, learning is dependent on prediction of the previous location, whereas SOC sequence learning relies on knowledge of the previous two positions to determine the forthcoming position. Based upon this information, Gabriel and colleagues (2013) altered the complexity of the sequence used in their study, to a 12-element SOC sequence. This change of sequence showed results in accordance with the PDH; deficits in procedural learning were evident in that children with DLD were slower than LT children. Lastly, Gabriel, Meulemans, Parisse, and Maillart (2015) used a visual and an auditory deterministic SRT task with a group of DLD and LT children. Performance was the same in both groups for the visual SRT task although they made more errors in the auditory SRT task, possibly linked to perception limitations.

A comprehensive study by Lammertink, Boersma, Wijnen and Rispens (2020b) entailed a two-pronged approach. This study comprised: 1) an experimental replication study with children (aged between seven and twelve years), and 2) a meta-analysis. Both explored the relationship between ISL as measured by an SRT task, and expressive grammar (recalling sentences subtest). In the experimental study, both the DLD and LT group showed sensitivity to the SRT task and there were no between group differences. In addition, there was no association between performance on the SRT task and expressive grammar. Results from the meta-analysis (139 children with DLD and 573 LT children; 19 effect sizes), did not find evidence (either in favour or against), of a relationship between ISL (SRT task), and grammatical proficiency.

Several studies show that consolidation of sequence learning is a weakness in individuals with DLD. Hedenius et al. (2011) investigated consolidation and longer-term

learning in DLD using an alternating SRT task (ASRT[6]) to measure probabilistic sequence learning. Participants completed an initial learning session followed up by a further session three days later. The overall results revealed no deficits in initial sequential learning in the DLD, but when participants were grouped based on grammatical impairment, only the LT group showed effects of consolidation and longer-term learning. These findings suggest that poor long-term retention is possibly linked to grammatical deficits typically seen in DLD. Desmottes, Meulemans, and Maillart (2016) also explored consolidation effects in children with and without DLD (aged between 7 and 12 years). Participants completed a deterministic SRT task followed by two further sessions, 24 hours later, and one week later. In the first session, there was similar performance in both groups in the early part of the task but no progression in the DLD group in the second part of the task. In the two follow on sessions the performance in the LT group showed increased sequence knowledge but there was no improvement in the DLD group. In a follow on study, Desmottes, Maillart, and Meulemans (2017) results showed similar sequence learning in children with and without DLD, but the DLD group had deficits in longer term consolidation of learning.

Most research investigating the PDH in relation to DLD focuses on children, yet a recent study by Earle and Ullman (2021) examined procedural memory in adults using an SRT task, and declarative memory using a recognition memory task (100 adults, 21 with DLD, aged 18-24 years). Memory was assessed after the first session and then again 12 hours later to assess consolidation effects. Results showed impaired procedural memory in the DLD group compared to LT adults, and when the groups were tested 12 hours later, there was no improvement in scores. In declarative learning, consolidation effects were evident in the LT

[6] The ASRT task (Howard & Howard, 1997) differs from the deterministic SRT task where learning is based on a patterns which are interspersed with random blocks. In the ASRT task, random stimuli are presented with predetermined stimuli throughout the task making it more difficult to detect regularities through explicit processes.

group, with performance improving following overnight sleep, yet this was not the case for adults with DLD.

The above body of literature highlights the variability in research findings to date regarding the relationship between procedural learning and language in DLD. Furthermore, it raises fundamental issues regarding the validity of the SRT task as a reliable measure of procedural learning. (Issues regarding the measurement of ISL will be discussed further in sections: *2.6.3 Evidence regarding implicit learning from meta-analyses,* and *2.6.3 Evidence regarding implicit learning from meta-analyses*).

2.6.2 Implicit learning: statistical learning

As discussed earlier, statistical learning refers to the ability in which humans determine patterns from the input (Romberg & Saffran, 2010). ISL is evident early on in development, demonstrated by work of Saffran, Aslin, and Newport (1996). This seminal study demonstrated that infants as young as 8-months-of age can extract patterns from a continuous speech stream, by using transitional probabilities to compute distributions and combinations of syllables. Since then, a substantial body of research has explored the relevance of implicit statistical learning in language learning, in both typical and atypical populations. In typical language development, the relationship between ISL and language is established in studies investigating grammatical ability (Kidd, 2012; Kidd & Arciuli, 2016; Lany, 2014), early literacy skills (Spencer, Kaschak, Jones, & Lonigal, 2015), and reading (Arciuli & Simpson, 2012; Qi, Sanchez, Georgan, Gabrieli, & Arciuli, 2019; Ren, Wang, & Arciuli, 2023).

A recent longitudinal study by Kidd, Arciuli, Christiansen, & Smithson (2023) assessed visual and auditory ISL in a cohort of 121 LT children in their first two years of primary school

(mean age: 6;1 years), with assessments at six-month intervals over four time points. In addition, at time point one, the authors assessed covariates of age, verbal working memory, nonverbal IQ, and language proficiency (vocabulary and grammar, assessed by the Australian version of the Clinical Evaluation of Language Fundamentals language assessment battery; Semel, Wiig, & Secord, 2006). The overall results show that both visual and auditory ISL provide independent contributions to language proficiency. Reliability of the ISL tasks increased over time, especially in the visual ISL task. As nonverbal IQ and working memory also related to performance on the visual ISL task at time point three and four, the authors suggest that individual performance could be related to maturational factors which enable sustained attention. One caveat is that children in the study were tested on the same ISL tasks at each time point and this may have resulted in them becoming over-familiar with the task, especially regarding the trained triplets - a point that has also been acknowledged by the authors.

A large and growing body of literature demonstrates impaired ISL in individuals with DLD. Logically, if ISL is important for language learning, then the relationship between ISL and DLD is not surprising (Bogaerts, Siegelman, & Frost, 2021; Saffran, 2018). Exploring auditory ISL in the form of word segmentation, Evans, Saffran and Robe-Torres (2009) conducted two experiments with 113 children (35 with DLD, mean age: 6;5 to 14;4 years), using a continuous speech stream. In the first experiment, children with DLD were less sensitive to transitional properties in the speech stream. In a second experiment conducted 6 months later, 30 participants from the original sample were tested again (15 children with DLD). In the second assessment, children heard the speech stream twice, with results showing that the DLD group were able to detect transitional probabilities with increased exposure. However, despite the fact that DLD children had double the exposure to the speech stream,

they still had difficulty in differentiating the target words from similar sounding words in the testing phase. Similarly, a study by Mayor-Dubois, Zesiger, Van der Linden, and Roulet-Perez (2014) established that children with DLD were unable to detect and learn regularities in a phonotactic learning task of continuous syllables. On the other hand, DLD children did show evidence of sequence learning in a SRT task, although not as efficiently as LT peers. It should be pointed out however that the sample was small (only 18 children with DLD, compared to a group of 65 LT children).

It is worth considering that most studies of ISL are based on online measures and there are few studies exploring outcomes using offline tasks (Lukács, Lukics and Dobó, 2021). Mixed findings are reported by Lammertink, Boersma, Wijnen, and Rispens (2020a) using a nonadjacent learning[7] (NAD) task in an artificial language. Both online and offline measures of learning were assessed in children (mean age: 9;1 years), with and without DLD. In the online measure (using response times), children with DLD showed reduced sensitivity to NADs in comparison to the control group. The offline measure involved the participants listening to utterances and deciding whether they had heard the utterance before. No learning was evident (although the authors suggest that the fact that they asked participants whether they had heard the utterance before may have altered the task from a grammaticality judgement task to a recognition-based task). Similarly, both online and offline measures were assessed in a study by Lukács, Lukics, and Dobó (2021). Children with and without DLD were assessed using both online and offline segmentation tasks (acoustic and visual presentation). In the online phases of the acoustic and visual nonverbal tasks, performance (measuring reaction times and accuracy) in the DLD and LT group were the same. However, the offline measures

[7] The NAD learning task in this study comprised a three-element utterance of an artificial language. Participants were asked to press a button if they thought the third element was a target, or a red button if it was not the target. Element 1 was a monosyllabic Dutch nonword, element 2 was a bisyllabic Dutch non word, and element 3 was a monosyllabic Dutch non word (Lammertink et al., 2020a).

(two-alternative forced choice and production tasks), indicated deficits in the DLD group. The authors suggest that online measures are more sensitive than offline measures in detecting learning, in line with findings from Lammertink et al. (2020a).

Contrary evidence is reported by Lammertink, Boersma, Rispens, and Wijnen (2020), in their study which analysed visual statistical learning in children with and without DLD (36 children, aged 7;8 to 10;4 years), using an embedded triplet design. They found no difference in learning between LT or DLD groups in an online or offline measure, and no association with literacy (reading, spelling, and pseudoword reading).

2.6.3 Evidence regarding implicit learning from meta-analyses

This section discusses research findings from several meta-analyses which have examined the evidence regarding ISL, both from a procedural learning and statistical learning perspective. The studies outlined in this section relate to LT and DLD populations.

Lum, Conti-Ramsden, Morgan, and Ullman (2014) conducted a meta-analysis with the aim of consolidating evidence regarding procedural memory deficits in DLD. This analysis compared results from eight studies using the SRT task, totalling 186 participants with DLD, and 203 LT controls. The average effect size was small, but significant (r = .328), therefore confirming a procedural learning deficit in DLD. Where SRT tasks included more exposure to sequences, effect sizes in the DLD group were smaller showing that their implicit learning is reliant on more experience than LT individuals. Participant age also predicted differences, with smaller effect sizes found in older participants. The authors interpret this finding in relation to the PDH, suggesting potential immature development of neural networks in younger children, or that declarative memory acts in a compensatory fashion, as stipulated by the PDH

(Ullman & Pierpont, 2005). In line with this hypothesis, it is predicted that differences between children with and without DLD should decrease as participants mature (Lum et al., 2014).

The involvement of both declarative and procedural memory in language learning is supported in a multiple meta-analyses study conducted by Hamrick, Lum, and Ullman (2018), which explored the predictions of the DP model based on children's first language (L1) (n = 470) and adult's second language (L2) learning abilities. The adult (n = 254) group was divided into lower and upper proficiency groups. Results showed a relationship between procedural learning and grammar in children (r =.27), and in the higher proficiency adult group (r =.55). There was no relationship between procedural learning and grammar in the lower proficiency adult group (r = -0.09). Declarative memory and lexical abilities were associated in children (r =.41), but a weaker relationship with grammar (r = .16). In adults, declarative memory correlated with grammar in the low proficiency group (r =.45), but not with the higher proficiency group (r = -.07). These results are in line with the PDH in demonstrating that adults with poor language potentially rely on declarative memory for grammatical processing.

A meta-analysis by West, Melby-Lervåg, and Hulme (2021) investigated six procedural learning tasks that are typically used to measure the PDH, namely: the SRT task, Hebb task, artificial grammar learning, weather prediction, contextual cueing, and statistical learning. Analysis was conducted at the group level, i.e., comparing language impaired groups with LT controls on various measures in relation to language ability (language and/or literacy). The results reveal deficits in language impaired groups in comparison to the LT groups, for example, an effect size of g = -.30 on the SRT task (although insignificant when groups were not matched for language ability: g = .03). Other reported relationships include: the Hebb task (g = -.32), the artificial grammar and statistical learning tasks (g = -.48), and the weather

prediction task (g = -.63), although the authors raise questions as to whether this is a pure measure of implicit learning). West et al. point out that, as effect sizes are relatively small, current evidence is insufficient to conclude that procedural deficits are a causal factor in DLD and dyslexia.

A large-scale meta-analysis by Oliveira, Henderson, and Hayiou-Thomas (2023) examined the relationship between procedural memory, measured by the SRT task, and continuous measures of language and literacy in children and adults (LT, DLD and dyslexia groups). Their results, based on 39 studies and 500 effect sizes, demonstrated a negligible relationship between procedural learning and language and literacy, of r = .06. This association did not retain significance at the group level, i.e., when separate analyses were conducted for each group (LT, DLD or dyslexia). Overall, the results suggest that there is limited evidence for a relationship between procedural learning and language and literacy, and for the declarative/procedural model and PDH hypothesis. The authors suggest that these findings could be linked to poor psychometric properties of the SRT task.

A meta-analysis by Lammertink, Boersma, Wijnen, and Rispens (2017) examined auditory verbal statistical learning in children and adolescents with and without DLD (eight studies: 213 participants with DLD and 363 LT children, aged six to 19 years). They found that individuals with DLD were less effective in detecting statistical regularities, when compared to LT children (mean effect size of r = .54). The authors also investigated whether age affected learning and whether the type of linguistic processing (i.e., word segmentation or grammar) had an impact on the results. Neither of these factors proved relevant to ISL learning.

Similarly, a meta-analysis by Obeid, Brooks, Powers, Gillespie-Lynch, and Lum (2016) examined ISL in participants with DLD (14 studies) and ASD (13 studies), reporting impaired ISL learning in DLD in comparison to LT controls: Hedges g = .46. A variety of paradigms were explored, namely: the SRT task, contextual cueing, artificial grammar learning, transitional probabilities from a speech stream, observational learning based on complex visual scenes, and probabilistic classification learning. In comparison, the ASD group did not show impaired ISL. A further analysis was conducted to compare task modality (visual vs. auditory), and age, but this did not influence effect sizes.

Isbilen and Christiansen (2022) investigated auditory linguistic statistical learning and language in their meta-analysis encompassing 25 years of research, and spanning a wide age range (from 5 months to just over 78 years). The authors report a mean effect size of d = 1.22 across populations (in adults: d = 1.50, children: d = .99, and infants: d = .63). However, they acknowledge publication bias, and additionally they raise several issues such as the effect of task type, and other methodological limitations (the latter will be discussed in the following section: *2.6.4 Limitations regarding implicit learning*).

In a recent meta-analysis, Ren, Wang and Arciuli (2023) studied the relationship between ISL and language, and ISL and reading[8]. The correlations were similar for both outcomes of interest: language related (r = .236) and reading related (r = .239). The only significant moderator of language was age, and it was found that associations with age were stronger for adults than children, despite there being more children in the overall sample. The

[8] The language measures included syntax, semantics, vocabulary, grammar-related skills, verbal working memory, receptive and expressive language, speech perception, lexical processing proficiency, and metaphor comprehension. Reading-related outcomes were assessments of word reading, non-word reading, sentence reading and passage comprehension, rapid automatized naming, spelling and verbal Scholastic Achievement Test results, metalinguistic awareness (comprising phonological, morphological, and orthographical awareness).

authors suggest that this potentially highlights more ISL stability in adult populations, or perhaps limited ISL capacity in children. Overall, the analyses comprised 53 studies (201 effect sizes). Of these, there were three correlations pertaining to DLD participants, with a reported effect size for language: r = -.034. There were no relevant correlations relating to DLD in the reading measures.

Lastly, Boeve, Zhou, and Bogaerts (2023) explore ISL and language skills incorporating a range of ISL paradigms, language abilities and participants. This meta-analysis incorporated the results of Ren et al. (2023) and Oliveira et al. (2023), and incorporated 97 papers (710 effect sizes). Consistent with the study by Oliveira et al. (2023), this meta-analysis investigated ISL and language using continuous variables (correlations between ISL and language). The authors report an effect size of r = .16 between ISL and language (syntax, comprehension and vocabulary). In contrast to Ren et al. (2023) they found no effect of age – the only variation across studies was found to relate to features of ISL tasks, such as ISL paradigm, stimulus modality and domain.

Overall, the aforementioned evidence provides variable findings regarding the relationship between ISL and language. The general inconsistency of results has been raised by several researchers and is a growing concern in the field of ISL, as discussed in the next section.

2.6.4 Limitations regarding implicit learning

As established at the beginning of this section, evidence regarding the relationship between ISL and language is varied. To date, a labyrinth of theories propose differing accounts regarding the relationship between ISL and language, often using different terminology to refer

to similar concepts and using the same experimental tasks to measure different constructs (Bogaerts, Siegelman, & Frost, 2021). Several researchers have raised concerns regarding terminology and the reliability of tasks designed to measure implicit learning (Arciuli & Conway, 2018, 2018; Arnon, 2020; Bogaerts et al., 2021; Isbilen & Christiansen, 2022; Siegelman & Frost, 2015). This is succinctly put by Bogaerts, Siegelman, & Frost (2021, p. 322):

> "…present studies tying SL to reading disabilities or SLI are (a) vague regarding the demarcation lines between the different theoretical constructs in the domain of learning (e.g., "motor procedural learning," "nonmotor procedural learning," "implicit learning," "sequence learning," "statistical learning"), (b) vague about the mapping of the experimental tasks they use to the theoretical constructs that these tasks are supposed to tap, and (c) vague about how these theoretical constructs are linked to language difficulties."

If ISL is not a well-defined construct, as Bogaerts et al. (2021) suggest, this may be why there are issues regarding terminology, and indistinct mapping of tasks to theoretical constructs. Kidd et al. (2023) suggest that researchers do not give enough consideration to the type of ISL used and tend to treat all ISL tasks as analogous to each other. Additionally, prior studies have largely measured ISL at the group level (i.e., comparison of impaired versus unimpaired groups), and only recently has the field begun to look at individual differences in ISL (Arciuli & Conway, 2018). Another potential issue, raised by Isbilen and Christiansen (2022), is that ISL tasks frequently use two-alternative forced-choice (2AFC) paradigms. These reflection-based measures require the participant reflecting upon prior learning and make an explicit choice between the target and a foil (Isbilen et al., 2020). This method potentially adds

noise to the data and results in weaker effect sizes (Bogaerts, Siegelman, Christiansen, & Frost, 2021; Isbilen & Christiansen, 2022). Isbilen and Christian (2022) suggest that processing-based tasks could be a more effective method to assess learning. For example, the statistically induced chunking recall (SICR) task exposes participants to an artificial language which includes tri-syllabic nonce words. Participants are then asked to recall strings of six-syllables. Results show that the SICR task has acceptable test-retest reliability and is more sensitive to individual differences than traditional 2AFC tasks (Isbilen, McCauley, Kidd, & Christiansen, 2020, Isbilen, Frost, Monaghan, & Christiansen, 2022; Kidd, Arciuli, Christiansen, Isbilen, Revius, & Smithson, 2020). This could be a promising avenue for ISL research, but there is no evidence to date as to whether it is relevant to ascertain ISL capabilities in people with DLD.

The above section highlights that there is much work to be done to ascertain whether ISL plays a role in language learning. This is an area of current investigation in the field and research regarding ISL is moving away from group level study designs to measuring ISL at an individual level. Continued investigation explores ways in which to reliably test ISL to capture the variability that typically occurs in individuals with DLD and ensure that we accurately measure the role of ISL in language.

3.0 Cognitive skill learning

3.1 Skill acquisition theory and automatization

This section discusses the theory of skill acquisition (also discussed as cognitive skill learning). Different theories pertaining to skill learning will be introduced, before discussing

their relevance to language learning, and lastly, automaticity, which is key to research in this thesis.

The process of skill learning involves learning and practicing a series of consecutive actions until they can be executed effortlessly and without controlled attention (Graybiel, 2008). The psychological theory of skill acquisition was originally proposed by Fitts (1964) and Fitts and Posner (1967), whereby it was proposed that learning a new skill comprised three main phases of development; 1) the cognitive phase, 2) the associative phase, and 3) autonomous phase. This theory served as the foundation for Anderson's Adaptive Control of Thought theory (ACT, ACT*: Anderson, 1982; 1983), which illustrates the relationship between cognition and performance, or more specifically, the idea that a cognitive skill is comprised of production rules.

Anderson and colleagues' theory of cognitive skill acquisition has been adapted and refined over the years (Anderson, 1982, 1993, 2007; Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Taatgen, Huss, Dickison, & Anderson, 2008). For example, ACT-R is founded on declarative and procedural memory as active knowledge repositories. Knowledge is represented as a 'chunk' in declarative memory and as a 'production' in procedural memory (Anderson, 1993). All knowledge is initially expressed in a declarative manner, and ACT* and ACT-R characterise the three basic stages of skill acquisition as follows: (1) the declarative stage, during which a new procedure is created using declarative memory. Working memory is essential in this phase because the learner must memorise subroutines to perform the skill; (2) this is followed by the knowledge gathering phase, which involves a gradual improvement in performance and reduction in cognitive input as the learner moves from the declarative to procedural juncture; and (3) the procedural stage, where procedural memory becomes

functional and automated. As a result, in the cognitive stage, learning a new skill necessitates problem-solving and planning techniques. The next stage involves considerable practice, where development in skill learning is measured by accuracy and speed. Finally, the skill is entirely automated and is no longer reliant on cognitive or declarative processes.

DeKeyser's (2015) model of skill acquisition theory identifies three distinct stages leading to automatization. DeKeyser does not discuss memory systems as such, but instead refers to declarative and procedural knowledge. DeKeyser's model has been widely applied to second language learners, with three stages as follows: (1) the declarative stage, where declarative knowledge is used to create explicit, linguistic rules, (2) the proceduralization stage involves 'acting on' knowledge to transform it into a behaviour. Once proceduralized, the information is available as a 'chunk' in procedural memory, (3) the final stage is automatization. These stages involve practice, to reduce reaction times and error rate, and finally the learner reaches automatization where performance is implicit and occurs without conscious effort. A contrary opinion is put forward by Paradis (2004), who suggests that metalinguistic declarative/explicit knowledge is not 'converted' into procedural/implicit knowledge. Instead, Paradis suggests that both metalinguistic and linguistic knowledge develop separately. As learners become more competent, there is a move from controlled (declarative) to automatic (procedural) processes: "When a skill is automatized, the performance is carried out rapidly, accurately and stably" (Paradis, 2004, p.40).

Performance is usually modelled as a power function, based on the 'power law of practice' (Newell & Rosenbloom, 1981), where, through practice, there is a reduction in reaction times and error rates over time. The process involves an initial steep learning curve as the learner moves from the declarative to procedural stage, but thereafter, reaction times and

error rates stabilize as the learner reaches the stage of proceduralization. DeKeyser (2015) states that individual differences can occur in the process leading to automatization, for example, individual aptitude, learning preferences, speed of processing, and teaching environment. Some learners may prefer to observe and listen prior to 'acting on' information, which converts the knowledge to a behaviour (DeKeyser, 2015).

The processes involved in cognitive skill learning can be applied to areas involving learning a new skill, such as driving a car, playing a musical instrument, or touch typing. An example is given by Johnson (1996), who identifies key components of skill learning involved in learning how to play tennis. Initially the learner pays conscious attention to acquiring the specific skills to play the game. The first step involves the cognitive phase, where one learns how to hold the tennis racket correctly, and how to execute individual strokes such as a forehand, backhand, and how to serve a ball. These steps involve mastering a new technique and a great deal of practice is involved before a player can competently return the ball to another person. Here, Johnson (1996) considers the hierarchical and goal-orientated structure of tennis, as well as the fact that the player continually evaluates their behaviour and performance to consider the speed and angle of the ball, the position of the ball on the tennis racket, and so on. The player also needs to decide upon the correct stroke and appropriate technique in response to their opponent and this involves combining many different skills at once. Next, during the associative phase, one practises the newly learned steps in chunks of movements, consistently and slowly, until they can be executed smoothly with fewer errors. Thereafter the player advances to more difficult shots such as volley, slice, or top spin. Finally, in the autonomous phase, the player can perform all steps automatically and play a steady game of tennis.

It could be questioned as to how skill learning is relevant to language. Behaviours that are executed automatically are beneficial as they can be produced efficiently and reduce cognitive load. Many of the aforementioned steps in skill learning involve components that are relevant to language, as summarized succinctly by Johnson (1996, p. 44):

> "Language is goal-directed hierarchically organized, non-stereotyped behaviour. From the environment, the performer receives information along various parameters. The performer's response is selected from a large repertoire of possible response. It must be appropriate along all the relevant parameters (hence exhibiting considerable combinatorial skill'), and in many cases must be executed speedily)."

Skill acquisition theory is usually applied to language in the context of L2 learning (Johnson, 1996; DeKeyser, 2015), although the notion of language as a complex, cognitive skill has also been applied to L1 acquisition by many researchers (Chater, McCauley, & Christiansen, 2016; Chater & Christiansen, 2018; Kamhi, 2019; McCauley, 2020). Chater and Christiansen (2018) discuss the concept of 'language as a skill', rather than 'language as knowledge'. They suggest that this view aligns with usage-based theories of language acquisition where children learn language according to daily exposure and experience and are therefore able to generalise constructions from the input that they receive. Specific to DLD, Kamhi (2019, 2020) considers language development from the perspective of skill learning and dual process models. He suggests that clinicians could support language development in DLD by focusing on certain areas of speech and language which can be practiced and eventually routinized. Kamhi's approach to skill learning emphasises the end goal of automatization which he considers integral to effortless communication.

Johnson (1996) states that a newly learned skill involves a great deal of attention, which he refers to as 'channel capacity'. Johnson proposes that in order to achieve automatization, a learner requires channel capacity for higher-order skills. For instance, in a two way conversation, automatic processing of tense would free channel capacity for higher-order cognitive processes such as understanding an utterance and conveying meaning (Johnson, 1996). Once performance is automatized, it is less likely to be affected by disruption and so, not only is language processed more efficiently, but it is also speedy, precise, and stable – factors that contribute to fluency (Segalowitz, 2003). Language is usually processed rapidly and efficiently in humans. This is evident in conversation where conversational gaps between two speakers are as little as 0.5 seconds or less (Bull & Aylett, 1988). This is not surprising in itself - a task that is performed automatically is usually processed without conscious effort and operates efficiently, requiring little attention (Reber, 1996). However, if you consider the processes involved in a two-way conversation (semantics, decoding, retrieving lexical units, advance planning, grammar, etc.), there are remarkable processes occurring underground, so to speak, and rapid exchange of speech would not be possible if linguistic knowledge were not automatized.

In language learning, automatization has been explored in the context of L2 learning, with automatization as a hallmark of proficiency (Hulstijn, Van Gelderen, & Schoonen, 2009; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998). Automatization is measured by response times, where RTs typically decrease as performance becomes automatized. Yet Segalowitz and Segalowitz (1993) draw a distinction between general speed-up, where performance improves *without* automaticity, and qualitative restructuring which is improvement *with* automaticity. In order to provide a pure measure of variability, the authors propose the coefficient of variation (CV), derived from dividing an individual's mean standard deviation of RT by their mean RT (CV_{RT}) (Segalowitz & Segalowitz, 1993; Segalowitz et al.,

1998). Ultimately, automatized behaviour results in lower CVs as performance improves and stabilises. Several studies have measured automatization using the coefficient in variation in L2 language learning (see Hulstijn et al., 2009). In this thesis, the CV is used as a measure of the speed of automatization.

In L1 language acquisition, automaticity has been explored by several researchers in the context of dyslexia. For example, dyslexia and skill learning (Vakil, Lowe, & Goldfus, 2015), the relationship between automaticity and executive functions in dyslexia (Smith-Spark & Gordon, 2022). Additionally, the Delayed Neural Commitment framework proposes that comorbidities commonly seen in neurodevelopmental disorders such as dyslexia are due to underlying deficits regarding motor skills, executive functions and language – leading to a delay in the development of automaticity (Nicolson & Fawcett, 1990, 2007, 2019). However, there is little research regarding automatization in DLD, aside from Kamhi's (2019, 2020) review, as detailed earlier in this section.

Three research studies in this thesis (a pilot study, study one and study two), assess the speed of automatization as measured by multiple trials of the Tower of Hanoi puzzle (MToH), as detailed in the next chapter.

PART TWO: EMPIRICAL STUDIES

4.0 Introduction to empirical studies and overall research questions

This thesis examines cognitive predictors of linguistic outcomes in children and adults, with and without DLD, to explore the following questions:

1. How does the speed of automatization predict linguistic outcomes in children and adults, and does this differ between LT and DLD participants? This question is investigated using a novel paradigm, the Multiple-trial Tower of Hanoi task (study 1a, 1b, and study 2).
2. What role does language aptitude/language analytic ability (metalinguistic knowledge) play in first language acquisition, in LT and DLD populations? (study 2 and 3).
3. Are there between group differences in language and cognition, in children and adults with and without DLD? (study 1b, 2, and study 3)
4. How do cognitive abilities predict linguistic outcomes in **children** with and without DLD? This thesis examines specific cognitive abilities in children: namely: nonverbal intelligence, working memory, implicit statistical learning, and speed of automatization (study 1), and analogical reasoning, language aptitude/analytic ability, and implicit statistical learning (study 3).
5. How do cognitive abilities predict linguistic outcomes in **adults** with and without DLD? Study 2 assessed cognitive predictors of nonverbal intelligence, language aptitude/analytic ability, pseudoword learning, implicit statistical learning, and speed of automatization.

This thesis comprises a pilot study and three empirical studies, involving the following samples:

Study	**Sample**	**Mean age (years)**	**Description of participants**
Pilot study:	36 L2 adolescents	15,5	Students learning English as a second language.
Study 1a:	78 LT children	8;1	Language typical children (study 1a and 1b involve the same sample of LT children).
Study 1b:	54 LT children 23 children with DLD	8;3	Children with and without DLD (study 1a and 1b involve the same sample of LT children).
Study 2:	30 LT adults 30 adults with DLD	38	Adults with and without DLD.
Study 3:	39 LT children 34 children with DLD	10;8	Children with and without DLD.

LT = language typical, DLD = Developmental Language Disorder

4.1 Pilot study: Dąbrowska and Blake (2022)

This section presents the accepted manuscript version of a book chapter entitled:

Speed of automatization predicts performance on "decorative" grammar in second language learning

This chapter was published in a volume edited by Piske, T., & Steinlen, A.K. (2022). *Cognition and second language acquisition: Studies on pre-school, primary school and secondary school children*, published by Narr Francke Attempto. pp. 285-309[9].

This pilot study examined the role of automatization and linguistic outcomes in 36 adolescents learning English as a second language. The results of this study determined the use of the MToH in later studies exploring the speed of automatization in children and adults with and without DLD (studies 1a, 1b, and 2).

[9] The formatting of this manuscript (i.e., figure and table numbers) have been altered to ensure consistency with the overall thesis document.

Published manuscript

SPEED OF AUTOMATIZATION PREDICTS PERFORMANCE ON 'DECORATIVE GRAMMAR' IN SECOND LANGUAGE LEARNING

Ewa Dąbrowska[1] Ashley Blake[2],

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, [2]University of Birmingham

Abstract

A large body of research shows that adult second language learners have particular difficulty with what we call 'decorative' grammar (aspects of grammar which have abstract and largely redundant meanings), while they often do well on tests of 'functional' grammar (those aspects of grammar which make clear contributions to meaning). We suggest that 'decorative' grammar is challenging for L2 learners because when grammatical routines are not automatized, speakers will tend to focus their processing resources on those aspects of language that are most relevant to meaning. We examine the relationship between 'functional' and 'decorative' grammar, assessed using a grammaticality judgement task (GJT) and a picture selection task (PST), respectively, and two cognitive abilities which are hypothesised to support language learning, namely the ability to automatize a complex cognitive procedure (assessed using the Multiple-trial Tower of Hanoi task, MToH) and explicit language aptitude (assessed using the Language Analysis test from the PLAB). As predicted, we found a positive correlation between performance on the MToH and the GJT, while Language Analysis predicted performance on both the PST and the GJT. We did not find a relationship between the MToH and the PST, suggesting that 'functional' grammar is less reliant on automatized routines.

Introduction

Language as a skill

Since the cognitive revolution, linguists have tended to think of language, and in particular, grammar, as a body of knowledge – either a set of rules or principles (e.g. Chomsky, 1965, 1981) or an inventory of conventional linguistic units (e.g. Langacker, 1987). It is undeniably true that knowing a language involves possessing a large body of knowledge, including a considerable amount of knowledge about the idiosyncratic properties of individual lexemes or classes of lexemes. However, knowledge alone is not sufficient. In order to be able to make a contingent contribution to a conversation, a speaker must decode the previous speaker's utterance, decide on their own response, retrieve the relevant lexical items, combine them in a grammatically appropriate way, and launch the appropriate articulatory sequence. This procedure is executed very quickly: The gaps between individual speakers' turns in conversational interchange are on average only 0.5 seconds long, and often they are even shorter. It is not unusual for conversational turns to overlap by about a syllable, which means that the second speaker has to do all of the above while the previous speaker is still executing their turn (Dąbrowska, 2004).

This kind of performance requires an enormous amount of skill, and a number of researchers have drawn parallels between language use and skilled behaviours such as driving a car, touch-typing or playing a musical instrument. Johnson (1996), for example, points out

both are goal-directed behaviours that are hierarchically organised and non-stereotyped, and both involve a selection from a large repertoire of options and the ability to combine units, often under time pressure. He explains how insights from research on skill acquisition can be applied to second language learning and teaching. A broadly similar view of L2 learning is advocated by DeKeyser (2015), while other researchers (e.g. Chater, McCauley, & Christiansen, 2016; Chater & Christiansen, 2018; Kamhi, 2019) have argued that language is a complex cognitive skill that develops through repeated use and practice before becoming effortless and automatic. Thus, research on skill learning is also relevant to first language development and processing.

Psychologists working on skill learning have identified three stages in the acquisition of skills, namely the cognitive, associative and autonomous phase (Anderson, 1982, 1993; Anderson, Greeno, Reder & Simon, 2000; Anderson, Bothell, Byrne, Douglass, Lebiere & Qin, 2004; Fitts, & Posner, 1967; Taatgen, Huss, Dickison & Anderson, 2008). In the cognitive stage, a new procedure is generated using declarative knowledge and working memory. During this stage, performance is slow and effortful, and error rates are relatively high. During the associative stage, declarative knowledge is incorporated into the procedures necessary to perform the skill. This stage sees a reduction in cognitive involvement and a gradual increase in the speed of processing. Finally, in the autonomous stage, procedural knowledge becomes increasingly automatic. This is evidenced by a further decrease in reaction times and error rates. In this stage, performance depends entirely on procedural memory and makes minimal demands on attentional resources, enabling skilled performers to direct their attention to other matters.

The behavioural changes observable during skill acquisition (i.e. reduction in error rates and speed) are very rapid in the early stages of learning (i.e. in the cognitive phase), then gradually level off. This decrease can be modelled by a power function (the 'power law of practice': Newell & Rosenbloom, 1981). Beyond the cognitive stage, the improvement in performance is very slow; however, psychological and neuroscientific research suggests that important qualitative changes are still happening "underground", so to speak. This is evidenced by the fact that different cognitive abilities are associated with performance at different phases of skill learning (Ackerman, 1992; Beaunieux, Hubert, Witkowski, Pitel, Rossi, Danion, Desgranges & Eustache, 2006; Hubert, Beaunieux, Chetelat, Platel, Landeau, Danion, Viader, Desgranges & Eustache, 2007). In the cognitive phase, individual differences in performance are predicted by differences in (non-verbal) intelligence, working memory and episodic memory. During the associative phase, episodic memory is no longer relevant, but intelligence and working memory continue to play a role (albeit not as pronounced as in the cognitive phase). In addition, psychomotor abilities become significant predictors (although their role is less prominent than in the autonomous phase). The associative phase can thus be considered a 'mixed phase', as only parts of the skill are proceduralised and there is continued involvement of working memory and intelligence. Finally, in the autonomous phase, differences in performance depend primarily on psychomotor abilities.

There is also evidence that different areas of the brain are involved during different phases of skill learning. For example, Hubert et al. (2007) found that the regions activated during the cognitive phase included the prefrontal cortex, the cerebellum, and the parietal regions. During the associative phase, higher levels of activation were observed in the occipital regions, right thalamus, and caudate nucleus, while the final, autonomous phase activated the left thalamus and caudate nucleus. These findings dovetail with the behavioural changes

described earlier: The involvement of the frontoparietal network in the cognitive stage indexes the use of problem-solving strategies; the involvement of the occipital regions during the associative and the autonomous phase suggests the use of mental imagery, while the involvement of the cerebellum in the autonomous phase fits in well with the fact that performance during this phase depends primarily on psychomotor abilities.

Language acquisition from a usage-based perspective

According to usage-based models of language development (Behrens, 2009; Bybee, 2010; Ellis, Römer & O'Donnell, 2016; Goldberg, 2006; Tomasello, 2003), grammar learning involves learning constructions or form-meaning pairings of varying size and degree of specificity. Learning a construction involves at least three stages. In the first stage, the learner acquires a database of analysed exemplars. This involves segmenting complex forms into smaller chunks and matching these chunks of form with salient semantic substructures. For example, a learner who hears the utterance where's the cat? in a situational context where it is clear that the speaker is asking about the location of a particular member of the species felis catus might be able to link part of the phonological form, namely /ˈkæt/ to the semantic substructure corresponding to the animal, and the form /ˈwɛər/ to the semantic substructure corresponding to a request for information about where something is located, as in the left-hand part of Figure 4. (Note that the segmentation and matching process could be partial, in the sense that some parts of the phonological form may initially not be attached to any semantic substructure. In the hypothetical example represented in the figure, the phonological forms corresponding to the definite article and the copula are not yet associated with any specific meaning).

The second stage involves generalisation. This is a crucial stage since this is what allows learners to produce novel utterances and can be thought of as involving a proportional analogy, where a previously learnt exemplar (or several exemplars) are used as a model in order to infer a missing term in a target. For example, imagine a learner who has learnt the phrase where's the cat? But does not initially know how to ask about the location of a pteranodon, although s/he knows the word pteranodon. Our learner can construct the novel form where's the pteranodon? on analogy to where's the cat? by retaining the matching elements (where's the …) and replacing the non-matching element (cat) with new content (in this case, pteranodon), as illustrated in Figure 4.

Analogical generalisation is a process that involves retrieving an appropriate analogical base, establishing correspondences between the component units of the base and target at both the semantic and the phonological level and projecting inferences about the missing term. This process is quite slow and effortful the first time it is performed. However, each subsequent analogical mapping is facilitated by previous experience, so that over time, the procedure for performing the mapping gradually becomes routinised and optimised. Thus, the final stage of the acquisition of a construction involves the entrenchment, or automatization, of the operations needed to generate novel forms.

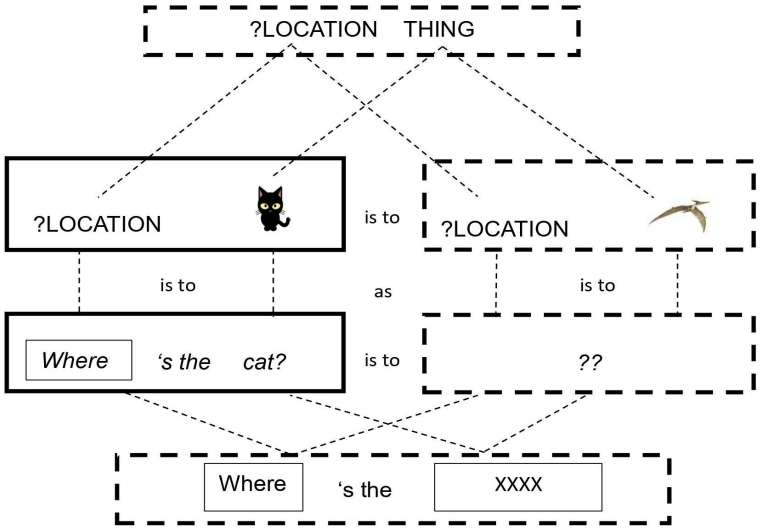


Figure 2. Generalisation as a proportional analogy problem

The first two stages of the acquisition of constructions (accumulation of analysed exemplars and generalisation) correspond to the cognitive phase of skill acquisition, the stage that relies on declarative knowledge and slow, effortful processing. The final stage of entrenchment corresponds to the associative and autonomous phases. This is when performance gradually becomes more fluent and more reliable as a result of practice. Importantly, a construction that is highly entrenched can be processed quickly and efficiently, with minimal effort; as a result, the speakers can concentrate their attention on the content of the message and the meaning they wish to convey.

Automatization is a slow process and requires vast amounts of practice. As an example, consider the acquisition of regular past tense marking in English. The relevant rule is relatively simple and can be acquired after just a few exposures to relevant exemplars. However, it takes much longer for learners to reach a stage where they supply the correct markers consistently whenever required. Many adult L2 learners never reach this stage: Instead, they show residual

optionality (i.e. supply the correct form most of the time but not all of the time). This is sometimes attributed to less effective procedural memory, although there is little evidence that older children or adults are less effective procedural learners than young children[10]. It is important to note that children acquiring their first language also go through a (sometimes lengthy) period of inconsistent performance. For example, English-speaking children start using the regular past tense inflexion at about age 2;4, and they begin to use it productively (as evidenced e. g. by overgeneralisation errors) a few months later. They reach Brown's (1973) criterion for mastery (i.e. supply the correct form 90 % of the time in obligatory contexts) by about age 4 and supply the target form at adult-like levels (i.e. close to 100 % correct) after 5;6 (Rice, Wexler, & Hershberger, 1998). This means that there is a period of inconsistent use spanning at least three years. Considering the frequency of the regular past tense marker, this is a very long time. Note, too, that if we assume that language acquisition begins in earnest at age 1 and that children are exposed to the language for eight hours every day (a rather conservative estimate!), by the time they are 5;6 they will have had 13140 hours of exposure (4.5 years * 365 days * 8 hours). Many late L2 learners are not exposed to this quantity of input; so even if they learn at the same rate as children, they may not have had enough input to fully automatize even very frequent inflexions.

Whatever the reason, it is well-known that many L2 learners do not fully automatize the grammatical routines of the second language: L2 processing relies to a greater extent on

[10] Most studies which report an age-related decline in procedural/implicit learning compared young and elderly adults (see Howard & Howard, 2013), and hence their results are not relevant to determining whether or not children are better at procedural learning than adults. Most studies that directly compared young children and young adults either found no difference or better learning in older children and adults compared to younger children (Lukács & Kemény, 2015; Thomas, Hunt, Vizueta, Sommer, Durston, Yang & Worden, 2004). To our knowledge, the only study that found an advantage in younger learners was Janacsek, Fiser & Nemeth (2012). However, the advantage only showed up when comparing raw reaction times, but not on accuracy measures or on Z-transformed RT measures.

controlled processes, hence processing is slower, more effortful and more prone to error (Ellis, Loewen, Elder, Erlam, Philp, & Reinders, 2009). Another important consideration is that controlled processes place heavy demands on working memory. Working memory capacity is strictly limited; that is to say, we can only consciously attend to a small number of elements at a time (Baars, 1997; Just & Carpenter, 1992). The implications for linguistic processing are clear: When processing is not fully automatized, speakers will attend to those aspects of an utterance which are most relevant to getting the message across, such as content words and aspects of grammar which make a clear contribution to meaning ('functional' grammar) at the expense of what Dąbrowska, Becker, and Miorelli (2020) call 'decorative' grammar, or those aspects of grammar whose contribution to meaning is relatively abstract and often redundant, such as tense and agreement marking, articles and subcategorization restrictions (Goldschneider & DeKeyser, 2001; VanPatten, 1996, 2004).[11]

Tower of Hanoi as a model of skill acquisition

In this study, we used a non-linguistic task (which we call 'Multiple-trial Tower of Hanoi', or MToH) to assess participants' skill learning abilities, and in particular, the speed of automatization. The Tower of Hanoi (ToH) (see Figure 5) is a puzzle consisting of three vertical rods mounted on a base and a number of discs that differ in size (in this study, we used a four-disc version). In the initial configuration, all four discs are stacked on the leftmost rod in ascending order of size, with the smallest one at the top. The participant's task is to move

[11] It should be emphasised that the 'functional'/'decorative' distinction is a continuum rather than dichotomy: Grammatical morphemes such as past tense, agreement markers and articles do add to the meaning, at least in some contexts. See Dąbrowska, Becker & Miorelli (2020) for further discussion.

the discs to the rightmost rod following certain rules, namely: (i) only one disc can be moved at a time, and (ii) one cannot place a larger disc on top of a smaller one. The solution to the ToH puzzle has a recursive structure in the sense that goals have subgoals embedded in them. For example, in order to reach the target configuration, one must first move the largest disc to the rightmost rod; in order to be able to do that, one must first remove the three discs that are on top of it, and so on. Thus, participants must hold goals and subgoals in memory while planning their moves.

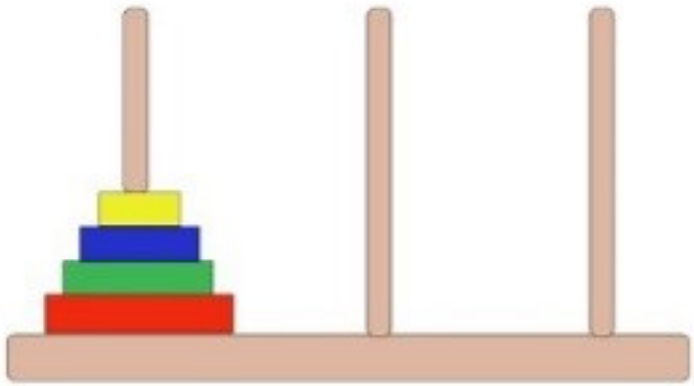


Figure 3. The Tower Of Hanoi puzzle

The ToH puzzle is often used in psychological assessment to measure executive functioning (Lezak, Howieson, Loring, Hannay, & Fischer, 2004) and problem-solving abilities (Donnarumma, Maisto, & Pezzulo, 2016; Marengo, 2015). In such studies, participants solve the puzzle just once, or at most two or three times. The ToH puzzle can also be used to measure cognitive procedural learning (Beaunieux et al., 2006; Hubert et al., 2007). In this case, participants solve the puzzle many times and the measure of interest is how quickly performance improves, as assessed by the number of moves per trial and/or time per trial. We adopt a similar experimental paradigm in our study.

Beaunieux et al. (2006) and Hubert et al. (2007) examined correlations between time per trial and performance on various cognitive measures in order to establish when a particular phase begins and ends. Given that in the cognitive phase the time taken to complete the puzzle depends primarily on IQ, while in the autonomous phase psychomotor abilities are the best predictor, they stipulated that the cognitive stage comprises those trials in which IQ is a significantly better predictor of completion time than psychomotor abilities, while in the autonomous phase the reverse would be true. The intervening trials, where there is no significant difference between the effect of IQ and psychomotor abilities, correspond to the associative phase. Using these anchoring points, they established behavioural criteria for delimiting the phases as follows:

- The cognitive phase ends when a participant finds the optimal solution to the puzzle (i.e. one which involves 15 moves).
- The associative phase begins immediately after the cognitive phase and ends when a participant applies the optimum solution in at least four out of five consecutive trials.
- The autonomous phase begins immediately after the end of the associative phase.

The current study

The current study has two main goals. Our first goal is to explore the predictive validity of the MToH task. To this end, we examine the extent to which measures of proceduralization and automatization derived from the task predict performance on measures of foreign language achievement. As a measure of proceduralization, we use the length of the associative phase, as defined by Hubert et al. (2007). Our measure of the degree of automatization is based on the fact that automatised skills require less conscious attention and hence are less susceptible to

interference from other tasks (cf. Anderson, 1982; DeKeyser, 2015). To assess the degree to which a skill is automatised, we compare performance on the ToH when the participant's attention is partly occupied by a concurrent task with baseline performance (i.e. without the concurrent task).

We measure grammatical achievement in L2 English using two different tasks. The first task assessed participants' grammatical comprehension using picture selection. In this task, participants were presented with a sentence and two pictures and were asked to choose the picture that went with the sentence. To be able to do this, participants must process the sentences for meaning. Therefore, the picture selection task is a measure of 'functional' grammar. The second task was a grammaticality judgment task in which participants were presented with sentences such as those in (1) and (2) and had to decide whether or not they were well-formed.

(1) (a) *Last night the old lady die in her sleep.

(b) Last night the old lady died in her sleep.

(2) (a) *The man lets his son to watch TV.

(b) The man lets his son watch TV.

In order for participants to be able to reject an ungrammatical sentence without any additional information about the intended meaning, the sentence itself must contain some clue about what the correct form should have been. For example, sentence (1a) is ungrammatical because the adverbial last night indicates that the action occurred in the past, but the verb is in the present tense form; and (2a) is ungrammatical because the verb let takes bare infinitives rather than to infinitives. In other words, for the task to work, the information provided by the

element which makes the sentence ungrammatical must be redundant and/or incompatible with some other element of the sentence. Therefore, grammaticality judgment tasks assess participants' knowledge of 'decorative' grammar.

The second goal of our study is to test a specific hypothesis about the role of two distinct aspects of language aptitude, namely inductive learning ability, or the ability to infer the rules governing language structure, and the speed of proceduralization/automatization in L2 learners' acquisition of 'decorative' and 'functional' grammar. We hypothesise that both aspects of aptitude will be positively correlated with performance on both tasks; however, for the reasons explained earlier, we expect inductive language learning ability to correlate more strongly with functional grammar and speed of proceduralization and automatization with decorative grammar.

Method

Participants

We recruited 36 students (24 males and 12 females) from a Grammar School in the Erlangen area of Germany who were learning English as a foreign language. The students were in the tenth grade and aged from 14 to 18 years (mean 15.5). All were fluent in German. 32 of the participants reported speaking German as their first language; of these, six indicated an additional first language to German (Vietnamese, Thai, Russian, Polish, Romanian and Turkish). The remaining four participants reported a first language other than German (two Greek, one Turkish, one Bosnian/Croatian). All participants had had at least five years of regular English instruction (estimated 722 classroom hours); six had technically had an additional year since they had repeated a school year.

Materials

Background questionnaire

The participants completed a background questionnaire that elicited information about their age, gender, native language(s) and any foreign languages they were learning in addition to English. They also chose an alias that was used to match their records for sessions 1 and 2.

Grammaticality Judgment Task (GJT)

The grammaticality judgment task (GJT) used the same stimuli as Dąbrowska et al. (2020) who in turn used a subset of the stimuli created by DeKeyser (2000) and Johnson & Newport (1989). The stimuli assessed basic aspects of grammar, including past tense, plural and third-person singular inflexions, determiners, particle movement, subcategorization, word order, the progressive, and pronominalisation. 80 of the 200 items used by DeKeyser (2000) were selected, making sure that each of the eleven categories from the original study was represented in proportions similar to those in the original test. The sentences were recorded by a female native speaker of English. Half of the stimuli were grammatical, and the other half were ungrammatical (see Table 1 for examples).

Table 1. Example of items used in the Grammaticality Judgment task

Category	**Grammatical variant**	**Ungrammatical variant**
Past tense	Last night the old lady died in her sleep.	Last night the old lady die in her sleep.
Plural	Three boys played on the swings in the park.	Three boy played on the swings in the park.

Third person singular	John's dog always waits for him at the corner.	John's dog always wait for him at the corner.
Present progressive	Tom is working in his office right now.	Tom working in his office right now.
Determiners	Tom is reading a book in the bathtub.	Tom is reading book in the bathtub.
Pronominalisation	The girl cut herself on a piece of glass.	The girl cut himself on a piece of glass.
Particle movement	The man climbed up the ladder carefully.	The man climbed the ladder up carefully.
Subcategorization	The man lets his son watch TV.	The man lets his son to watch TV.
Yes-No questions	Will Harry be blamed for the accident?	Will be Harry blamed for the accident?
WH questions	What is Martha bringing to the party?	What Martha is bringing to the party?
Word order	The woman asked the policeman a question.	The woman the policeman asked a question.

Following earlier studies, participants heard both the grammatical and ungrammatical version of each sentence. The items were divided into two blocks so that if the grammatical version of an item occurred in block 1, the ungrammatical version occurred in block 2 and vice versa. Within each block, the sentences were presented in the same random order to all participants.

Participants were instructed to type "g" if they thought the sentence was grammatical or "u" if they thought it was ungrammatical. They were asked to keep their index fingers on the relevant keys throughout the task. Each sentence was played twice, with a two second pause between repetitions. The next trial began as soon as the participant responded. If the participant did not respond within ten seconds from the offset of the first presentation, the trial timed out and the next trial began. There was a short break after the first 40 items.

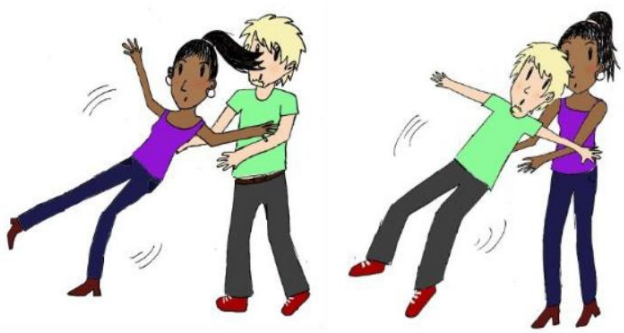
The Picture Selection Task (PST)

The picture selection task is an online version of Dąbrowska's (2018) Pictures and Sentences Test. The PST tests the comprehension of ten different structures (actives, passives, subject clefts, object clefts, subject relatives, object relatives, simple locatives, locatives with quantifiers, possessive sentences with quantifiers, post-modifying prepositional phrases). Each item in the test consists of a sentence (e.g. *The woman was the one that the man caught*) and two pictures, e. g. a woman catching a man and a man catching a woman (see Figure 6). There are eight items per structure (hence 80 items in total), presented in a semi-random order, so that no two sentences of the same structure were presented immediately next to each other (see Table 2 for examples of all sentence types).

Table 2. Example of items used in the Picture Selection task

Construction	**Example**	**Pictures**
Active	The man fed the girl.	Target: man feeding girl Distractor: girl feeding man
Passive	The girl was fed by the man.	
Subject cleft	It was the man who fed the girl.	
Object cleft	It was the girl that the man fed.	
Subject relative	The man was the one who fed the girl.	
Object relative	The girl was the one that the man fed.	
Locative w/quantifier	Every pencil is in a box.	Target: three boxes, each containing a pencil plus an extra box Distractor: three boxes, each containing a pencil plus an extra pencil
Possessive w/quantifier	Every box has a pencil in it.	Target: three boxes, each containing a pencil plus an extra pencil Distractor: three boxes, each containing a pencil plus an extra box
Simple locative	The spoon is in the cup.	Target: spoon in cup Distractor: spoon next to cup
Postmodifying PP	The spoon in the cup is red.	Target: red spoon in green cup Distractor: yellow spoon in red cup

The presentation order was the same for all participants. The sentences were spoken by a female native speaker of English. Participants responded using arrow keys (the left arrow to choose the picture on the left and the right arrow to choose the picture on the right). They were asked to keep their index fingers on the keys throughout the task. The response deadline was ten seconds per item. If the participant did not respond within this time, the trial timed out and the next trial began.



The woman was the one that the man caught.

Figure 4. An example of an item from the Picture Selection task

The Language Analysis Task

The Language Analysis task was adapted from the Language Analysis subtest of the Pimsleur Language Aptitude Battery (Pimsleur, Reed & Stansfield, 2004). Participants were presented with words and simple sentences in an unfamiliar language and their English translations. Their task was to translate 15 novel sentences into the language. Participants selected the correct answer from an array of four by clicking an on-screen button. All stimuli were presented in writing and the vocabulary and model sentences were visible on the screen at all times. There was no time limit. The maximum possible score in this task was 15.

Multiple-trial Tower of Hanoi (MToH)

The four-disc version of the puzzle was used. Participants were asked to solve the puzzle 40 times, with a secondary task introduced in the last five trials. The secondary task involved tapping in response to a recording consisting of a random sequence of the words one and two, with four second pauses between the words. Participants were instructed to tap with their non-dominant hand in response to the numbers (one tap in response to word one and two taps in response to the word two) while solving the puzzle with their dominant hand.

Procedure

Participants' parents were contacted by letter and asked to sign a consent form. The experiment consisted of two sessions. Session 1 was administered via an online platform (Pavlovia). Participants accessed the experiment by following a link given in the letter to their parents. The link took them to a web page that explained what the experiment would involve. They were then asked to click a button to indicate their consent to take part. They were advised that they could withdraw from the study at any time simply by closing the browser. The tasks were administered in the following order:

1. Background questionnaire
2. Spoken Grammaticality Judgement Task (Block 1)
3. Language Analysis Task
4. Picture Selection Task
5. Spoken Grammaticality Judgement Task (Block 2)

Session 2 comprised the MToH task. Participants were tested individually in a quiet room at their school during the school day. The experimenter explained the rules, showed

examples of legal and illegal moves and offered further clarifications if necessary. Following Beaunieux et al. (2006), the experimenter also demonstrated the first move (placing the smallest disc on the middle rod); this was done in order to prevent a possibly random search through the possibility space. Participants completed the puzzle 40 times, with the secondary task introduced after the 35th trial. The experimenter counted the number of moves per trial using a hand-held counter and recorded the length of each trial using a stopwatch on a smartphone. In addition, the participants' hand movements were video recorded for later checking.

Results

In the GJT, data for two participants were excluded because they had either more than 20 consecutive timeouts or more than 20 consecutive RTs shorter than 1000ms, suggesting they had not engaged with the task. In the Language Analysis Task, two participants were excluded as they completed the task in 22.5 seconds and 90 seconds respectively, which also suggests a lack of engagement. (The task normally takes 10-12 minutes to complete.) Two participants were excluded from the MToH as they indicated that they were already familiar with the puzzle.

Performance on the MToH task

Figures 7 and 8 show how performance on the MToH improved as a function of practice. Figure 7 shows the decrease in the average number of moves per trial, and Figure 8 the decrease in the average time per trial. Both curves show a typical power function shape. At the beginning of the MToH task, participants needed on average 21 moves per trial and took about 70 seconds to complete the puzzle. Both the mean number of moves per trial and the

mean time per trial decreased sharply over the first six trials and then started to level off. After 15 trials, there is very little improvement: that is to say, the participants' performance begins to stabilise at about this time. Most participants stabilised at the optimal (i.e. 15-move) solution; three participants stabilised at a 17-move solution and at a 19-move solution.

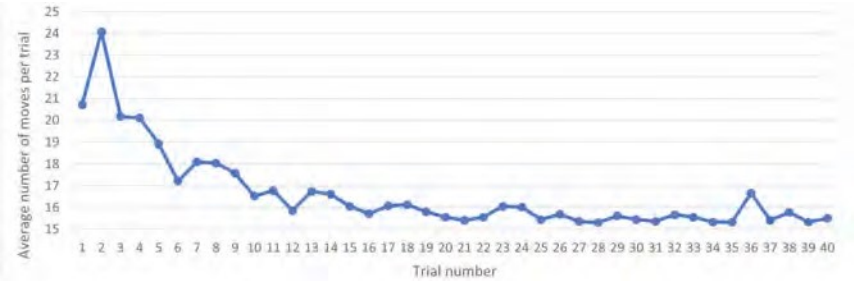


Figure 5. Average no. of moves per trial

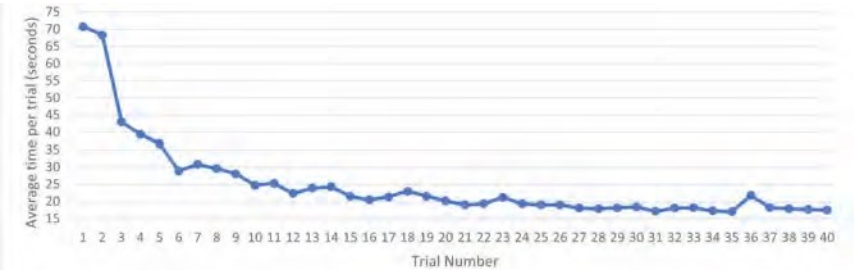


Figure 6. Average time per trial

In the last five 'normal' trials (trials 31-35) participants averaged about 15.5 moves and 17.5 seconds per trial – i.e. they needed just over one second per move. There was a slight increase in both the number of moves and completion time on trial 36 (when the secondary task was introduced), but most participants returned to previous levels of performance very quickly, indicating a relatively high degree of automatization.

Descriptive statistics and correlational analysis

For the Picture Selection Task and the Language Analysis Task, participants were given one point for each correct answer (therefore maximum possible scores are 80 and 15 respectively). For the Grammaticality Judgement Task, performance was assessed using the sensitivity index (d'). The sensitivity index is a measure derived from signal detection theory. Unlike a simple proportion of correct responses, d' takes a participant's response bias into account, thus providing a more sensitive measure of how well a participant is able to discriminate between grammatical and ungrammatical items (Huang & Ferreira, 2020).

The remaining four measures are derived from performance on the MToH task. LAssoc1 and LAssoc2 are measures of the length of the associative phase. According to the criteria introduced by Hubert et al. (2007), the cognitive stage ends when a participant achieves the optimal solution for the first time, and the associative phase ends when he/she achieves the optimal solution in four out of five consecutive trials. LAssoc1 is based on this criterion, with one difference. As indicated earlier, some of our participants never discovered the optimal solution and stabilised on a solution with 17, 18 or 19 moves (see below). For this reason, we define the associative stage with respect to the stable solution achieved by a particular participant (i.e., the solution which they replicate on most subsequent trials). The associative stage, therefore, begins on the trial immediately following the trial in which a participant achieved the stable solution for the first time, and ends on the trial at which the participants achieved a stable solution on at least four out of five consecutive trials. Thus, LAssoc1 corresponds to the number of trials between these two points in time, including the last trial of the associative stage. LAssoc2 is based on a stricter criterion for the end of the associative phase, whereby a participant is required to achieve the stable solution in five out of five consecutive trials and corresponds to the number of trials between the trial on which the participant achieved the stable solution for the first time and the trial on which the participant

achieved a stable solution on five consecutive trials (including the fifth consecutive successful trial).

Two additional measures (difference in the number of moves: DiffMoves and difference in time: DiffTime) were motivated by the fact that when a procedure is automatized, its execution does not require much conscious attention and therefore performance should not be disrupted if a secondary (tapping) task is introduced (cf. Anderson, 1982; DeKeyser, 2015). These two measures were computed by subtracting the average number of moves (DiffMoves) or the average time per move (DiffTime) in the trials with the secondary task (i.e. trials 36-40) from the average number of moves or the average time per move in the last five 'normal' trials (i.e. trials 31-35). A negative figure for these two measures indicates that performance was affected by the secondary task, and the lower the figure, the more it was affected.

The data were analysed in the R software (R Development Core Team, 2020); d'values were computed using the `psycho` package (Makowski, 2018). The descriptive statistics for performance on the four experimental tasks are given in Table 3.[12] As can be seen from the table, there are considerable individual differences in performance on all tasks.

Table 3. Descriptive statistics

Measure	Mean	SD	Range
PST Accuracy	69.44	6.18	58-79
GJT Accuracy	53.97	9.99	26-71
GJT d'	0.90	0.72	-0.92-2.33
GJT RT	4812.28	1033.11	3046-6707

[12] The statistics for reaction time reported in Tables 1 and 2 are based on raw reaction times. We also computed RT values using the standard filtering procedure (i.e. considering only RTs for correct items and excluding values that were more than two standard deviations above or below the mean for each participant). However, since the two measures were virtually perfectly correlated (r = .99), we chose to report the unfiltered values.

PST RT	2995.32	578.58	2323-4815
Language Analysis	10.91	3.49	4-15
LAssoc1	6.59	3.73	4-16
LAssoc2	12.21	10.37	4-45
DiffMoves	-0.27	1.20	-5.60-2.20
DiffTime	-1.15	2.74	-13.27-3.51

Table 4 presents the pairwise Pearson product-moment correlations between all variables. As expected, there was a positive relationship between performance on the two language tasks for both accuracy (r = .29) and reaction time (r = .60), although only the latter was statistically significant. This is not surprising as both tasks measure different aspects of grammatical proficiency. Correlations between accuracy and reaction times on the two tasks were non-significant (and negative), which indicates that there was no speed-accuracy trade-off. As predicted, we found a significant correlation between the Language Analysis task and accuracy on the picture selection task (r = .45). The correlation between Language Analysis and performance on the GJT was weaker but also significant
(r = .35).

Table 4. Pairwise correlations between all variables

	GJT d'	**GJT RT**	**PST Acc**	**PST RT**	**LgAnal**	**LAssoc1**	**LAssoc2**	**DiffMoves**	**DiffTime**
GJT d'	1	-0.15	0.29†	0.07	0.36*	0.01	0.01	0.47**	0.49**
GJT RT	-0.15	-0.15	-0.43**	0.60***	-0.11	-0.09	-0.14	0.23	0.19
PST Acc	0.29†	-0.43**	1	-0.27	0.45**	-0.14	-0.19	0.09	-0.08
PST RT	0.07	0.6***	0.27	1	0.09	0.1	-0.1	0.18	0.17
LgAnal	0.36*	-0.11	0.45**	0.09	1	-0.49**	-0.24	-0.08	-0.07
LAssoc1	0.01	-0.09	-0.14	0.1	-0.49**	1	0.27	0.07	0.01
LAssoc2	0.01	-0.14	-0.19	-0.1	-0.24	0.27	1	0.3†	0.18
DiffMoves	0.47**	0.23	0.09	0.18	-0.08	0.07	0.3†	1	0.86***
DiffTime	0.49**	0.19	-0.08	0.17	-0.07	0.01	0.18	0.86***	1

**** Correlation is significant at .001 level*
*** Correlation is significant at .01 level*

*** Correlation is significant at .05 level*
† Correlation approaches significance (p < 0.10).

The two measures of the length of the associative phase based on Hubert et al. (2007). LAssoc1 and LAssoc2, do not correlate with performance on any of the language tasks. They are also not correlated with the interference-based measures and only weakly correlated with each other. DiffMoves and DiffTime, in contrast, are strongly correlated with each other (r = .86). There is a significant relationship between DiffMoves and performance on the GJT (r = .47) and DiffTime and performance on the GJT (r = .49). There were no significant correlations between any of the interference measures and PST or between the interference measures and reaction times on either task.

Discussion

The main objectives of our study were to investigate the predictive validity of different measures of proceduralization and automatization and to examine the effects of various aspects of language aptitude on the acquisition of 'functional' grammar (aspects of grammar that provide a clear contribution to meaning) and 'decorative' grammar (aspects of grammar where the contribution is abstract and often redundant). In the following section, we discuss the validity of our measures. The next two sections are devoted to a discussion of our results in the context of the predictions outlined in the introductory section.

Measures of proceduralization and automatization

The interference-based measures introduced here (DiffMoves and DiffTime) show moderately strong correlations with performance on the GJT, and they are also strongly correlated with each other, suggesting that they are both reliable and valid measures of

automatization. In contrast, the two measures based on Hubert et al.'s (2007) criteria (LAssoc1 and LAssoc2) did not correlate with performance measures on any of the language tasks. They are also not correlated with the interference-based measures (DiffMoves and DiffTime). It could be argued that this is because they measure something different, namely the speed of proceduralization (as opposed to automatization); however, the fact that they are only weakly correlated with each other raises doubts about their reliability.

It is worth noting that the effects of automatization may be even stronger than our results suggest. A closer inspection of the participants' performance on the MToH by trial shows that performance stabilises around trial 15. This means that participants had about 20 trials left to automatise the procedure before the secondary task was introduced. This might have been too much so that virtually all participants reached high levels of automaticity. The average DiffMoves score was -.27, meaning that participants made, on average, only .27 more moves per trial when the secondary task was introduced. The average DiffTime was -1.15 seconds, which means that they needed just over one extra second per trial. Over the five trials with the secondary task, this equates to 1.35 moves and approximately six extra seconds. Furthermore, only 15 participants out of 34 had negative difference scores, and for twelve of them, the effect was minimal (less than one move per trial). In other words, the disruption in performance produced by the secondary task was very small and affected less than half of the participants. This suggests that we may have introduced the secondary task too late in the experiment: in other words, if the secondary task were introduced earlier (e.g. after 20 or 25 trials), we are likely to have observed more interference and larger individual differences – and possibly even stronger effects of speed of automatization.

Inductive language learning ability and speed of automatization as predictors of performance on 'functional' and 'decorative' grammar

The Language Analysis subtest of the PLAB is an established test of explicit language aptitude which is particularly relevant to L2 grammar, particularly in earlier stages of acquisition. Since fluent language processing relies on automatized knowledge, we hypothesised that the speed of automatization (as measured by the MToH) should also make an important contribution to explaining grammatical attainment. Given that explicit attention is a limited commodity, language learners tend to attend primarily to those aspects of form which are most relevant to meaning. Therefore, we predicted that Language Analysis scores would correlate more highly with performance on 'functional' grammar (assessed using a picture selection task), while MToH should be more strongly associated with performance on 'decorative' grammar (assessed with a traditional grammaticality judgement task). These predictions were mostly met. Language Analysis predicted performance on both the PST and the GJT, with the former relationship being stronger than the latter (.45 vs. .35). There was also a robust relationship between both interference measures (DiffMoves and DiffTime) and performance on the GJT (.47 and .49 respectively). In contrast, our data show no significant relationship between the interference measures and performance on the PST. These results suggest that explicit language aptitude (and in particular, inductive learning ability as measured by the Language Analysis test) play a role in the acquisition of both 'decorative' and 'functional' grammar, while speed of automatization appears to be relevant for 'decorative' grammar only.

Reaction time measures

Our primary measure of proficiency was accuracy. However, one would also expect to find a relationship between the reaction time measures and aptitude, and particularly speed of automatization. This was not the case, which is somewhat surprising, since RT measures are typically more sensitive than accuracy measures. Furthermore, as noted earlier, our results show no evidence of a speed-accuracy trade-off that might have masked the effects of one or both of these measures. It is possible that all our participants concentrated more on accuracy than on speed. Whatever the reason, our findings suggest that reaction-time based measures are not always more sensitive than accuracy measures.

Conclusion

Our findings suggest that interference from the secondary task on the MToH is a promising measure of the speed of automatization. Measures based on the length of the associative phase, on the other hand, had no predictive validity. This could indicate that the benefits of proceduralization only become evident once the performance is highly automatised. On the other hand, the fact that the two measures of the speed of proceduralization do not correlate with each other suggests that they are simply not very reliable. As discussed earlier, our interference measure may have been introduced too late, this should be taken into account in future research. Our results also support the hypothesis that speed of automatization is relevant for mastery of 'decorative' grammar in L2 learning in instructional settings. Further research will be needed to determine if it is also relevant for more functional aspects of grammar and whether it also holds for other age groups (e.g. younger learners) and in other settings (e.g. naturalistic exposure).

Acknowledgements: We would like to thank Anna Schreiber for collecting and coding the data and Laura Becker for programming the experimental tasks. This project was funded by an Alexander von Humboldt Professorship (ID-1195918) awarded to the first author.

4.2 Study one (a): Blake, Dąbrowska, and Riches (under review).

This section presents a manuscript entitled:

Cognitive predictors of children's linguistic abilities

This paper refers to study one of this thesis and explores the relationship between cognitive abilities and language in LT children. The primary aim of this study was to examine the cognitive processes underlying language ability within a usage-based framework of language acquisition. Specifically, the study explored how nonverbal IQ, working memory, implicit statistical learning, and speed of automatization contribute to vocabulary and grammar outcomes (examined using standardized assessments and elicited speech samples), in a sample of 78 LT children (mean age: 8;1 years).

NB. Study one (a) and study one (b) incorporate the same sample of children.

COGNITIVE PREDICTORS OF CHILDREN'S LINGUISTIC ABILITIES

Ashley Blake[1], Ewa Dąbrowska[1,2], Nick Riches[3]

[1]University of Birmingham, [2]Friedrich-Alexander University, Erlangen-Nuremburg, Germany

[3]Newcastle University

Abstract

Differing theoretical approaches to language acquisition offer contrasting explanations to the cognitive processes underlying linguistic acquisition and knowledge. Nativist theories suggest that language is domain specific and independent from cognition whereas usage-based models of language propose that language is dependent upon general cognition. In the present study, we explore cognitive predictors of children's linguistic attainment in a group of 78 children. Overall, our results show that nonverbal IQ and working memory are the most significant predictor of language outcomes. Interestingly, the relationship between nonverbal IQ and grammar is stronger than between nonverbal IQ and vocabulary. This finding is problematic for modular approaches to language and suggests that explicit learning and problem solving play a significant role in children's native language acquisition.

Keywords: nonverbal intelligence, working memory, implicit statistical learning, speed of automatization, usage-based models of children's language acquisition

Introduction

Background

The relationship between language and cognition has been widely debated in psycholinguistic research. The notion of a module for language became popular in the 1980s following Fodor's book 'The Modularity of Mind' (1983), and led to the emergence of the modularity hypothesis: the view of language as a domain specific system which develops

independently from general cognition (Bellugi, Lichtenberger, Jones, Lai, & St George; Curtiss, 2012; Smith & Tsimpli, 1995). Other researchers have proposed that language itself is composed of smaller modules such as grammar, lexicon, and phonology (Chomsky, 1981). In contrast, usage-based models abandon such distinctions as their central assumption is that linguistic knowledge is captured in terms of symbolic units, i.e., form meaning pairings of varying sizes and degrees of generality. In other words, grammar and vocabulary form a continuum (Fillmore, 1993; Fillmore, 1988; Goldberg, 1995).

Evidence for the domain specificity of language has been proposed through double dissociations in neurodevelopmental disorders such as Williams syndrome (WS) and specific language impairment (SLI). For example, children with SLI are said to have impaired language ability yet they perform within the normal range on IQ tests (Bishop, 2006; Leonard, 2014a), whereas children with WS have an uneven cognitive profile yet relative strengths in language ability (Bellugi et al., 2000). Pinker (1999) suggested that this double dissociation demonstrated that language was domain specific. Following decades of research, we now know that the examples provided by Pinker are not quite so straightforward. Children with SLI often have non-linguistic deficits, suggesting that SLI is not specific to language. This, and issues with terminology and identification led to the new recommended term for SLI: developmental language disorder (DLD: Bishop, Snowling, Thompson, Greenhalgh, and the CATALISE-2 consortium, 2017). According to revised diagnostic criteria for DLD, inconsistency between language ability and cognition is no longer required - DLD is a heterogenous disorder and language proficiency, and cognition differ considerably between individuals. Similarly, children with WS have uneven cognitive profiles and their language development trajectory is complex and multifaceted (Van Herwegen, 2015).

Children differ in the way they acquire language, subject to various external factors such as quality and quantity of input, socioeconomic status, parental education and knowledge of child development (Fernald, Marchman, & Weisleder, 2013; Hirsh-Pasek et al., 2015; Hoff, 2003; Huttenlocher et al., 1991, Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2008; Rowe & Snow, 2020; Rowe, 2012). Furthermore, language acquisition is dependent upon several different cognitive abilities which interact in complex ways. As a result, there are vast individual differences both in the paths that individual learners follow and in the end states that they achieve (Dąbrowska, 2015; Kidd & Donnelly, 2020). In our research, we investigate language acquisition in a group of primary-school children, but we move away from modular claims and explore the relationship between cognitive processes and linguistic ability from a usage-based perspective.

The overall purpose of this study is to investigate the role of cognitive processes in language development in primary-school children with typically developing language. In this paper, we begin by providing a brief outline of usage-based theories of language acquisition followed by a general overview of four specific cognitive abilities explored in our study: nonverbal intelligence, working memory, implicit statistical learning and the speed of automatization.

Usage-based models of language acquisition

In contrast to modular accounts, usage-based theories propose that children's language acquisition is not due to innate domain-specific abilities or knowledge, but involves the interaction of cognitive processes and language use (Behrens, 2009; Bybee, 2010; Ellis et al., 2016; Goldberg, 2006; Tomasello, 2003). Children learn their mother tongue through

experience, by listening to those around them, inferring meaning from contextual information, and extracting regularities from what they hear. This process starts in infancy and in their first year of life, children are already using specific skills such as intention-reading, pattern-finding, and analogical reasoning (Tomasello, 2003). Broadly speaking, learning the grammar of one's language involves three stages:

1) The first stage involves the child learning lexically specific constructions, defined as form-meaning pairings. These constructions differ in form and schematicity (Langacker, 2005), and include basic forms which are low in complexity such as *deer* or *banana*, as well as constructions that are more complex (see Hilpert, 2021), for example, *(it's) all gone!* Children extract forms (words, phrases and morphemes) to form a database of exemplars which can be matched with corresponding chunks and salient semantic substructures.
2) During the second stage, the child uses processes of analogical mapping and generalisation which form 'slot and frame' patterns. For example, a learned exemplar of *it's a lorry* could be used as a model to create a new form such as *it's a big, red tractor.*
3) Finally, through repeated usage events, linguistic knowledge becomes entrenched (Schmid, 2020). This process requires considerable practice and exposure until constructions are automatized and language becomes fluent and effortless (see Dąbrowska, 2004, 2010a; Dąbrowska & Blake, 2022).

Whilst usage-based researchers maintain that language is dependent upon general cognition, it is not clear which cognitive abilities play a role in this process, and at what point in development (Dąbrowska, 2019). Consequently, in this study we evaluate the relationship between language development and cognition, specifically nonverbal intelligence, working

memory, implicit statistical learning, and the speed of automatization. These abilities will be reviewed in the following section.

Nonverbal IQ

Nonverbal or 'fluid' intelligence involves reasoning and problem-solving abilities, and is often thought to be largely independent of experience (Cattell, 1943, 1963; Cattell et al., 1941). It is considered a purer measure of ability than verbal or 'crystallised' intelligence which involves recalling and applying acquired knowledge (e.g., vocabulary). Nonverbal intelligence (IQ) tests are widely used to predict performance in a range of areas, e.g., education, employment and well-being (see Sternberg, Grigorenko, & Bundy, 2001). For instance, IQ is considered a predictor of academic achievement. This is demonstrated in a large, longitudinal study which found a robust association (r = 0.81) between latent intelligence traits measured at the age of 11 years (Spearman's *g* from the Cognitive Abilities Test: Thorndike, Hagen, & France, 1968), and academic achievement at 16 years of age (Deary, Strand, Smith, & Fernades, 2007).

However, the relationship between nonverbal IQ and grammar is less straightforward. For example, it is widely accepted that children's grammar relies on implicit processes (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001), and is therefore independent of IQ (Feldman, Kerr & Streissguth, 1995; Reber, Walkenfeld, & Hernstadt, 1991). Modular theories generally assume that all children acquiring the same language converge on the same grammar, as illustrated by the following quotations:

> "... the grammars that are in fact constructed vary only slightly among speakers of the same language, despite wide variations not only in intelligence but also in the conditions under which language is acquired." (Chomsky 1972, p.69)

> "… every child … acquires knowledge of his language, and the knowledge acquired is, to a very good approximation, identical to that acquired by others on the basis of their equally limited … experience (Chomsky 1975, p.30)

Long (2013) has a similar perspective:

> "…all but the most severely intellectually impaired children achieve native-like abilities in the L1 provided simply that they are exposed to the language early enough." (Long, 2013, p.17), and likewise:

> "…all children exposed to a language before closure of hypothesized SPs succeed in achieving native abilities, despite sizable differences in IQ and exposure." (Long, 2013, p.23).

We know from previous research that there are correlations between IQ and language. For example, the relationship between intelligence and vocabulary is well-established (Raven, 2000), and there are strong correlations between vocabulary and grammatical ability in development (Bates, Bretherton, & Snyder, 1988; Dixon & Marchman, 2007). However, in order to learn grammar (i.e., generalizations about how words and morphemes are combined), one needs to have a critical mass of items that can be combined, so the relationship between IQ and grammar could be simply due to the fact that vocabulary learning paces grammar

acquisition in the early stages of development. Therefore, in this study, we examine nonverbal IQ on the one hand, and grammar and vocabulary on the other in school-aged children, in order to demonstrate whether this relationship also holds in later stages of language acquisition.

Working memory

The importance of working memory in language acquisition has been emphasised in a wide body of research spanning over fifty years. As a limited-capacity system, working memory enables the temporary storage and processing of information used in complex cognitive tasks (Baddeley, 2000; Baddeley & Hitch, 1974). Word learning, language processing, comprehension, and production all place demands on working memory capacity in that they require manipulation of information. For example, word learning involves learning the phonological and semantic properties of novel words, linking the two, and then storing the newly learned word in memory (Jackson et al., 2021). Thus, the relationship between phonological working memory and vocabulary retrieval is well documented (Gathercole, 2006; Gathercole et al., 1992), particularly in children with DLD who typically have poor phonological working memory (Gathercole & Baddeley, 1990; Henry & Botting, 2017; Montgomery et al., 2010). However, working memory deficits are not limited to DLD and individual differences in working memory capacity occur in the general population. For instance, children with low phonological memory (as assessed by non-word repetition), produce shorter and less diverse sentences than children with better phonological memory (Adams & Gathercole, 2000; Adams & Gathercole, 1995).

Given that usage-based models assume that the abstraction of schemas relies on analogy (Tomasello, 2003), in order to extract a schema, a child has to establish a proportional analogy

of form, i.e., how form meaning one is to form one, as form meaning two is to form two (see Dąbrowska & Blake, 2022; Winckel & Dąbrowska, 2024). This process of analogical mapping places heavy demands on working memory capacity (Simms et al., 2018).

Crucially, working memory is key to academic achievement and plays an important role in reading attainment (Peng et al., 2018), comprehension (Daneman & Carpenter, 1980), and mathematics ability (Bull & Scerif, 2001). Although some researchers maintain that working memory is closely associated with IQ, the two are not synonymous. This has been shown in a longitudinal study which investigated the role of working memory and IQ in children and found that, at the age of five years, working memory is more significant than IQ in predicting literacy and numeracy outcomes six years later (Alloway & Alloway, 2010).

The relationship between working memory and language is usually assessed using variations of complex span tasks. In the present study, we use a non-verbal variation of the backwards digit span task to explore working memory as a predictor of linguistic attainment.

Implicit statistical learning

Another cognitive mechanism contributing to language acquisition is implicit statistical learning (ISL), the ability to detect regularities in the environment (Isbilen & Christiansen, 2022). ISL is considered a stable characteristic and is not linked to other high-level cognitive processes such as working memory or intelligence (Kaufman, DeYoung, Gray, Jimenez, Brown, & Mackintosh, 2010; Siegelman & Frost, 2015 - although see Arciuli, 2017). In language learning, ISL is evident early on in development, demonstrated in the well-known study by Saffran, Newport and Aslin (1996), which showed that babies of 8-months of age can

segment words from a continuous pseudo speech stream. Since then, numerous studies have provided evidence of the relationship between ISL and linguistic skills, for example, grammatical ability (Kidd, 2012; Kidd & Arciuli, 2016; Lany, 2014), early oral language, vocabulary knowledge and phonological processing (Spencer et al., 2015), and reading (Arciuli & Simpson, 2012; Qi et al., 2019). Further evidence is summarized in recent meta-analyses: for example, Ren, Wang and Arciuli (2023) examined the association between statistical learning and a) language-related skills, and b) reading-related skills, reporting a similar correlation for both outcomes of interest (r = .24). A slightly larger meta-analysis of 97 studies used meta-regression to investigate the association between individual differences on statistical learning and language ability, with a small effect size of r = .16 (Boeve et al., 2023).

However, while it is generally believed that ISL plays a role in language, evidence is not robust due to a lack of effect found in many studies, possibly linked to task reliability (Arciuli, 2017; Arciuli & Conway, 2018; Arnon, 2019, 2020; Siegelman et al., 2017). A meta-analysis by Oliveira et al. (2023) reveals a weak relationship between the SRT task and procedural learning of r = .06, and West et al. (2021) provide mixed results regarding the relationship between a variety of ISL tasks and language, in individuals with typically developing language, DLD and dyslexia.

The speed of automatization

Linguistic knowledge is highly automatized - language processing occurs swiftly and is usually effortless in skilled speakers. For example, consider a two-way conversation exchange where there is rapid turn-taking between speakers. Gaps between conversational turns are only 0.5 seconds long on average, and sometimes even less (Bull & Aylett, 1988).

This is remarkably fast considering that during this time an interlocuter is required to understand and decode a spoken utterance, access the meaning and relate it to context, and then retrieve the appropriate lexical units in which to form a grammatically correct reply. Although we take this process for granted, the ease and speed of language comprehension and production is made possible by the automatization of linguistic knowledge. Automatization has been widely explored as a predictor of performance in second language learning (DeKeyser, 1997; DeKeyser, 2001; DeKeyser, 2015), or investigated from the perspective of an automaticity deficit in neurodevelopmental disorders such as dyslexia (Nicolson & Fawcett, 1990, 2007, 2019). However, it is not generally studied in first language acquisition.

In the present study, we use the Tower of Hanoi (ToH) puzzle to explore the speed of automatization. Previous research has used tower tasks (ToH and Tower of London) as a measure of executive functioning (Lezak et al., 2004), to explore problem-solving (Donnarumma, Maisto, & Pezzulo, 2016; Marengo, 2015), and cognitive skill learning (Beaunieux et al., 2006; Hubert et al., 2007). Vakil, Lowe, & Goldfus (2015) used the ToH puzzle to investigate the relationship between skill learning and reading skills in children with dyslexia. In the present study, we use 30 trials of the ToH to investigate how the speed of automatization predicts linguistic performance.

While modular theories postulate that grammar is independent of cognition, usage-based models predict that there is a relationship. In the present study, the relationship of interest is between domain-specific abilities. We explore how cognitive abilities relate to general linguistic outcomes and also between different aspects of language.

Method

Participants

Our analysis is based on a group of 78 children aged between 7;0 and 10;8 years (mean age: 8;1 years). Participants were recruited through schools and social media platforms, e.g., Facebook and X (formerly Twitter). The methodology described here is detailed in Blake and Dąbrowska (2024), and involves the same sample of children as described in Blake, Dąbrowska, & Riches (under review).

Materials

Standardized language assessments

We used three standardized language assessments:

1. **Test for Reception of Grammar.** (TROG-2; Bishop, 2003). This receptive grammar test is presented as a picture selection task and assesses 20 different aspects of grammatical knowledge (e.g., relative clauses, word order, the *neither… nor…* construction, pronoun gender, postmodifiers, passives, and more). There are 80 items in blocks of four and the participant must answer all four items correctly to pass a block.

2. **The British Picture Vocabulary Scale** (BPVS3; Dunn, Dunn, Styles, & Sewell, 2009) is a receptive vocabulary assessment task. Participants are presented with 4 pictures and are asked to select the number of the picture that matches the word read aloud by the experimenter. Basal and ceiling sets are established so that the child starts and

finishes at the correct stage for their ability. There are 14 sets of 12 test items, which increase in difficulty until the ceiling set is reached (eight or more errors in a set).

3. **Recalling sentences sub-test of Clinical Evaluation of Language Fundamentals** (CELF-5; Wiig, Semel, & Secord, 2013). This is a sentence repetition task where the child is asked to listen to spoken sentences and repeat the sentence back to the researcher without changing the content. The sentences increase in length and complexity as the task progresses. The assessment is discontinued after five consecutive sentences with four or more errors in each sentence.

Language measures based on narrative samples

A narrative task and an expository discourse task were used to elicit speech samples. We describe these below, as well as the measures used for analyses.

1. 'The Frog Story' is a picture story without words based on 'Frog, Where are You' (Mayer, 1969). We used an adapted version of the story with colour illustrations by Mònica Pascual Holder (made available by Esther Pascual), to make it more visually appealing for children. The experimenter advises the child that they will look at a story without any words. The child sees the story twice, the first time to look at the pictures, and the second time, they narrate the story in their own words. The narrative is recorded for later transcription.

2. Expository discourse tasks (Nippold, Mansfield, Billow & Tomblin, 2008) are recommended as a method to capture linguistically complex sentences (see Nippold et al., 2005, 2008). To do this, the child is asked to talk about their favourite game, sport,

or hobby. The experimenter asks the child to explain in more detail by imagining the experimenter is an alien who knows nothing about the planet earth. The experimenter uses specific prompts, for example: what are the rules of the game, sport, or hobby?, recommended strategies, and why the game, sport or hobby is their favourite. The child's response is recorded for transcription purposes.

Specific measures were derived from the narrative and expository discourse tasks, detailed as follows:

1. Mean length of T-units in morphemes. When using data from older children, it is difficult to determine what counts as an utterance, especially when they are producing longer texts. We therefore used mean length of T-unit rather than the mean length of utterance, as a measure of general syntactic complexity. A T-unit is defined as a 'main clause plus whatever subordinate clauses happen to be attached to or embedded within it' (Hunt, 1965, p. 301).
2. Subordination index: this measure is calculated by counting the total number of clauses (including main and subordinate clauses) and dividing this by the total number of T-units.
3. Productive vocabulary: vocabulary diversity (VOCD). This is a vocabulary diversity measure computed by the KIDEVAL program, in CLAN (Computerized Language ANalysis) (MacWhinney, 2000; MacWhinney & Fromm, 2016).
4. Speech rate: automatically computed in Praat (Boersma & Weenink, 2022) using the Syllable Nuclei v2 script (De Jong & Wempe, 2009). This script provides a computation of the number of syllables per second, divided by the total time of the speech sample.

5. Disfluencies (per minute): this measure is computed by using three types of disfluencies, these are: filled pauses (e.g., ‘um’, ‘er’), retractions (words or phrases that are withdrawn and corrected), and repetitions (where a word or phrase is repeated). The total count of these disfluencies is divided by the duration of the speech sample in minutes, which provides an average rate of disfluencies per minute of speech.

Cognitive tasks

1. **The Multiple-trial Tower of Hanoi** (MToH) task (Blake & Dąbrowska, 2024; Dąbrowska & Blake, 2022). Here, we use the Tower of Hanoi puzzle which comprises three vertical rods mounted on a wooden base. We use a four-disc version of the puzzle for our study. At the start of the task, the discs are arranged on the leftmost rod, in ascending order, with the largest disc on the bottom. To solve the puzzle, the child is advised to move the discs from the leftmost rod to the rightmost rod, so that when they have solved the puzzle, all four discs will be stacked on the rightmost rod, in ascending order. There are two rules: 1) to move one disc at a time, and 2) a larger disc cannot be placed on a smaller disc. After explaining the rules, the experimenter demonstrates examples of legal and illegal moves. The experimenter also shows the child the first move by placing the first disc on the second rod (following Beaunieux et al., 2006), to make the puzzle easier to solve and reduce possible frustration. Children repeat the puzzle 25 times followed by five trials of a secondary task which involves using their dominant hand to solve the puzzle and at the same time, tap in response to a recording (with their non-dominant hand). The recording plays a sequence of the words ‘one’ and ‘two’, with a four-second pause between them. The child is asked to tap in response to the numbers (two taps when they hear the word ‘one’ and one tap in response to the

word ‘two’). The reason for including the secondary task is to measure automatization: if the procedure for solving the puzzle is not fully automatized, then we expect the secondary task to interfere with performance. The number of moves and the time for each trial is recorded.

2. **Raven’s Coloured Progressive Matrices** (Raven’s CPM: Raven, Raven, & Court, 1998) is a nonverbal IQ test designed to measure general cognitive ability. Each diagrammatical problem has a part missing and the participant is required to select the missing piece from one of six options. The child selects their answer by naming the number of the option corresponding to the missing piece of the puzzle. There are 36 items in 3 sets, each containing 12 problems which increase in difficulty as the task progresses.

3. **Backwards Colour Span task** (Riches, 2012; adapted from Zoelch, Seitz, & Schumann-Hengsteler, 2005). This is a variation on the backwards digit span task. The experimenter places different coloured balls in an opaque tube and the child’s task is to tell the experimenter in what order the balls will come out of the tube. Each block has six trials, with two balls per in block one, three balls in block two, and so on. In order to pass a block, the participant must achieve four correct responses, and after three failures in a block, the task is discontinued. There are nine different colours used which are equivalent to digit recall stimuli. These are black, white, red, green, pink, grey, and brown, which are equivalent to eight one-syllable stimuli, and one two-syllable stimulus: yellow, corresponding to the number seven. Before starting the task, the experimenter asks the child to name each of the colours.

4. **Embedded triplet task** (adapted from Arciuli & Simpson, 2011). This is a child-friendly, implicit statistical learning task based on an embedded triplet paradigm and developed on the Gorilla experiment builder platform (https://gorilla.sc: Anwyl-Irvine et al., 2020). The task comprises visual stimuli of colourful cartoon aliens drawn by the first author. In the familiarisation stage, the child is told that they will meet friendly aliens who are lining up to go into their spaceship. The children are then presented with a continuous stream of 12 cartoon aliens which are shown on the screen one at a time. Unbeknown to the participants, stimuli are presented in triplets with three aliens in each triplet (ABC, DEF, GHI, JKL) appearing in the same order. To monitor attention, a cover task is included, and the children are instructed to look out for an alien that may appear twice, and when this happens, they should press the space bar. Each base triplet appears in the familiarisation stream 18 times and in 6 instances, one alien is presented twice in a row, e.g., AABC.

   After the familiarisation stage, the test phase comprised the original triplets together with 'impossible triplets', which were created by combining one alien from each of the original triplets, i.e., AEI, DHL, GKC and JBF. Participants then complete a forced-choice task. They are asked to help the aliens line up so that they are in the correct order to return to their spaceship. The child's task is to choose which group of aliens line up in the correct order from the long line of aliens that they saw at the beginning of the task. On each trial, the child is presented with a base triplet and an impossible triplet. There are 64 trials, with each base triplet and impossible triplet appearing 16 times.

Procedure

All research sessions were conducted online due to the COVID-19 pandemic. Prior to the research sessions, participant's parents and guardians were asked to complete an online consent form and background questionnaire for their child. After this, a Tower of Hanoi puzzle was posted to the family with the request that they should not open the parcel until the first session. This was to ensure that they did not familiarise themselves with the puzzle before starting the online sessions. Parents or guardians and their child met with the experimenter on the videoconferencing platform, Zoom (https://zoom.us). The experimenter used a digital visualiser to present the tasks to the child, using the screen share option on Zoom. In the case of picture selection tasks, the child was asked to select their answer by saying the number of the picture that matched the word or sentence read aloud by the experimenter. There were three sessions, each taking between 45 minutes and 1 hour. The tasks were administered in the following order:

Session One:

Speed of automatization (Multiple-trial Tower of Hanoi task)

Session Two:

Narrative task: The Frog Story

Receptive grammar (TROG-2)

Productive grammar (Recalling Sentences sub-test of CELF-4)

Expository discourse

Session Three:

Receptive vocabulary (BPVS3)

Working memory (Backwards Colour Span)

Nonverbal IQ (Raven's Coloured Progressive Matrices)

Implicit statistical learning (Embedded triplet task)

Results

Data processing

Linguistic measures based on narrative samples

The CLAN (Computerized Language ANalysis: see MacWhinney, 2000) software program was used to transcribe and code the speech samples from the narrative and expository discourse tasks. Data were coded for pauses, filled pauses, repetitions, retractions, elision, T-units (see below), and subordinate clauses. Two researchers verified coding of approximately 10% of the complete transcriptions to ensure that they were consistent and accurate. Once the samples were transcribed, an automatic analysis was conducted using the KIDEVAL program, a component of CLAN (see MacWhinney & Fromm, 2016), to provide a language profile for each child. The audio files were firstly trimmed to ensure that they only contained child speech and any utterances from the experimenter were removed. As the speech samples were relatively small, a composite measure was formed by combining narrative and spontaneous speech data. Speech rate was measured using the Syllable Nuclei script v2 (de Jong & Wempe, 2009), in Praat (Boersma & Weenink, 2022). The script detects syllables in running speech by identifying peaks and dips in intensity. This automatic method of detecting syllable counts has

good reliability when compared to manual methods of syllable counts (r = .88, see de Jong & Wempe, 2009).

Measures of the speed of automatization

To assess the speed of automatization using the MToH, we used two different measures. The first measure used is difference in moves: “AutoDiff”, which is obtained by subtracting the average number of moves in the secondary task (trials 26-30) from the average number of moves in the last five standard trials (trials 21-25). The assumption underlying this operationalization is that the more automatized a procedure is, the less interference there will be from the secondary task. Negative scores for this task indicate the presence of interference, the lower the number, the greater the interference.

Our second measure, "the coefficient of variation, “AutoCV”, is used to measure performance on the MToH task by dividing the standard deviation of a participant’s number of moves by their mean number of moves (as response times and number of moves were strongly correlated, we decided to focus our analysis on moves). This measure is derived from previous research conducted by Segalowitz and colleagues. By dividing the mean RT by the standard deviation of RT, the coefficient of variation controls for pure speed up and reflects a pure measure of variability (see Hulstijn et al., 2009; Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998). Therefore, we would expect automatized behaviour to reflect a decrease in the coefficient of variation over trials.

All data were analysed with R Statistical Software (ver. 4.2.1, R Core Team, 2022). Missing data from the implicit statistical learning task (four participants, equating to 5%) were

imputed using the `mice` package (van Buuren & Groothuis-Oudshoorn, 2011). This task was problematic as some children found the task difficult and did not know the answers so they did not want to continue. Due to poor internet connection, speech samples for two participants were not clear and these data were imputed.

The descriptive statistics are provided in Table 5. Both standardized and raw scores are provided for norm-referenced language assessments of receptive and productive grammar, and receptive vocabulary, whereas raw scores are provided for elicited speech measures and cognitive variables. Participants' raw scores on the nonverbal IQ assessment correspond to standardized scores above 85, placing them within the average range of 85-115 (raw scores are reported for nonverbal IQ, this is because Raven's CPM standardized scores are given in five-point intervals, which limits precision when converting from raw to standardized scores).

There are potential ceiling effects for some participants on the receptive grammar and receptive vocabulary assessments, especially at the highest performance levels. In contrast, the cognitive measures indicate a narrower range of performance, indicating less variability in comparison to the standardized language assessments.

Reliability estimates (Spearman-Brown corrected) were computed for the two novel tasks, the ISL task and the MToH task, using the `splithalf` package (version 0.8.2; Parsons, 2021), with 5000 random splits. For the ISL task, the reliability was 0.52 (95% CI [0.30, 0.68). For the MToH task, once participants' performance stabilises (usually at around the midpoint of the task), their scores show little variability. Therefore, to account for this, the split-half reliability was conducted by dividing the data into odd and even trials. For odd trails, the reliability was 0.64 (95% CI [0.49, 0.76), and for the even trials, 0.62 (95% CI [0.43, 0.76]).

Descriptive statistics

Table 5. Descriptive statistics

Variables	Mean	SD	Median	Range
Language measures				
Receptive grammar (standardized)	106.81	14.23	109.00	55.00 – 125.00
Receptive grammar (raw, blocks)	15.90	3.10	17.00	5.00 – 20.00
Productive grammar (standardized)	11.59	2.97	12.00	4.00 – 17.00
Productive grammar (raw)	48.28	12.11	51.50	42.25 – 69.00
Receptive vocabulary (standardized)	108.32	13.87	108.00	70.00 - 140.00
Receptive vocabulary (raw)	122.56	19.17	123.00	71.00 – 173.00
Productive vocabulary	34.81	8.05	35.37	12.19 - 51.28
Subordination index	1.23	0.11	1.22	1.02 – 1.52
Mean length of T-unit	8.46	1.81	8.62	5.00 – 17.32
Speech rate	2.71	0.46	2.68	1.30 - 3.50
Disfluencies	13.93	6.06	12.89	4.69 - 44.18
Cognitive measures				
Nonverbal intelligence	28.32	5.10	30.00	14.00 - 36.00
Working memory	14.41	2.94	14.50	8.00 - 21.00
Implicit statistical learning	50.86	9.23	49.00	36.00 - 83.00
AutoCV	0.04	0.04	0.03	0.00 - 0.17
AutoD	-1.29	1.53	-0.80	-7.60 – 2.40

Language measures: receptive grammar (TROG), productive grammar (CELF: sentence repetition), receptive vocabulary (BPVS3), productive vocabulary (vocabulary diversity: VOCD); subordination index, MLTU: mean length of T-units, speech rate, disfluencies per minute. Cognitive measures: Nonverbal IQ (Raven's CPM: raw score), working memory (backwards colour span; total correct), implicit statistical learning (percent correct), AutoCV: coefficient of variation for the MToH task (standard deviation divided by mean number of moves: trials 1-25), AutoDiff: difference in moves: MToH task (average number of moves in trials 26-30 minus average number of moves in trials 21-25). All scores are raw unless stated otherwise.

Correlations

Inferential statistics are reported on raw scores to facilitate easier interpretation of the results. Table 6 shows correlation coefficients between cognitive and linguistic variables. As shown; age, working memory, and nonverbal intelligence are correlated, and they correlate positively with most language measures. It is worth nothing that the relationship between nonverbal intelligence and receptive grammar (r = .69) is higher than that between intelligence and receptive vocabulary (r = .53). A Steiger's Z-test revealed that the difference was statistically significant, Z = 2.20, p < 0.05. The correlation between our measures of automatization, AutoDiff and AutoCV is very weak (r = -.14). Whilst we expect a negative relationship (a high degree of automatization is associated with a low coefficient of variation and a higher value for difference in moves), the relationship between these two variables should be stronger if they measure the same thing. So, it is possible that they are measuring different aspects of automatization.

Most language measures are positively correlated with each other, although in a few cases the correlation coefficient is close to zero. This is also true of disfluencies which is surprising as we would expect more advanced learners to be more fluent as well. However, disfluencies are also weakly correlated with age, nonverbal IQ and working memory: this could be because older children, and those with higher intelligence and a larger working memory span, attempt to produce more complex language. It is not surprising that age and nonverbal IQ scores are correlated, and both of these correlate with language measures. However, apart from vocabulary, the correlations with IQ tend to be stronger than with age.

Table 6. Pearson's correlations

	Age	IQ	WM	ISL	AutoDiff	AutoCV	RGr	PGr	RVoc	PVoc	SI	MLTU	Disfl	SR
Age	-													
IQ	**0.37**													
WM	**0.40**	**0.40**												
ISL	0.16	0.16	0.11											
AutoD	0.04	-0.05	-0.06	0.22										
AutoCV	-0.07	**-0.34**	0.04	0.11	-0.14									
RGr	0.21	0.69	**0.43**	0.06	-0.04	**-0.30**								
PGr	**0.36**	**0.49**	**0.44**	0.00	-0.09	-0.18	**0.74**							
RVoc	**0.54**	**0.53**	**0.45**	0.11	0.03	-0.11	**0.62**	**0.68**						
PVoc	0.23	0.06	0.02	0.06	0.21	-0.08	0.16	0.24	0.15					
SI	**0.36**	**0.42**	0.27	0.11	0.08	-0.08	**0.35**	**0.37**	**0.42**	0.28				
MLTU	0.20	**0.37**	0.16	0.22	-0.06	-0.06	**0.34**	**0.32**	0.29	0.08	0.70			
Disfl	**0.31**	0.24	0.21	0.17	-0.03	0.06	0.26	0.24	0.22	0.01	0.25	0.28		
SR	0.18	**0.41**	0.05	0.27	0.17	-0.17	**0.38**	**0.38**	**0.31**	0.22	**0.45**	**0.54**	**0.37**	-

IQ: nonverbal IQ; WM: working memory, ISL: implicit statistical learning task; RGr: receptive grammar, PGr: productive grammar; RVoc: receptive vocabulary; PVoc: productive vocabulary; AutoDiff: difference in moves as a measure of automatization (MToH task), AutoCV: co-efficient of variation as a measure of automatization (MToH task); SI: subordination index; MLTU: mean length of T-units in morphemes; Disfl: disfluencies per minute; SR: speech rate.
** Values of .30 or above (in boldface) are significant at p <.05.*

Regression analysis

To examine the effects of our cognitive predictors (age, nonverbal IQ, working memory, implicit statistical learning and two measures of automatization: AutoCV and AutoDiff), and linguistic outcomes, we used ordinary least squares linear regression. We initially entered all predictors into our models and removed non-significant predictors one at a time. The 'relative importance' of each predictor was assessed using the `'lmg'` metric from the `relaimpo` package in R (Grömping, 2006). The calc.relimp function provides an estimate of the proportion of variance of each regressor's influence to the model.

The `performance` package in R (Lüdecke, Ben-Schachar, Patil Waggoner, & Makowski, 2021), was used to establish model diagnostics. Where assumptions for heteroscedasticity were not met, the values for standard errors, *t*-values and *p*-values were adjusted using the Eicker-Huber-White method (White, 1980). This method provides robust standard errors to ensure that the estimated coefficients remain valid despite potential violations of the homoscedasticity assumption. The 'lmg' metric only works with the standard model (without the correction for heteroscedasticity), but as the models are virtually identical, the 'lmg' metric is presented below (see Table 7). We now discuss the results of each model.

Grammar

Table 7 refers to the linear regression model for receptive grammar. The overall model explained 50% of variance in receptive grammar scores, indicating a moderate to strong fit to the data. Nonverbal IQ accounted for 39% of the explained variance, and working memory 11%. Similarly, for productive grammar, 31% of the variance in scores is explained by nonverbal IQ which accounts for 18% of variance, and working memory accounting for 13%. These results indicate that that nonverbal IQ and working memory contribute to performance in both models of grammar, with nonverbal IQ playing a larger role in explaining the variance.

Vocabulary

The overall model for receptive vocabulary explains 44% of the variance in scores. Age contributes to most of the variance in the model (18%), followed closely by nonverbal IQ (16%), and working memory, which accounts for 10% of variance. For productive vocabulary, age is the sole predictor explaining 5% of variance.

Measures of grammatical and lexical complexity, and fluency

As can be seen, the model of subordination index explains 23% of variance overall, with both age and nonverbal IQ emerging as significant predictors. Nonverbal IQ accounts for most of the variance (14%), while age accounts for 9%.

Nonverbal IQ is the sole predictor of mean length of T-units, with the overall model accounting for 14% of variance.

Age predicts disfluencies, explaining 10% of the variance.

Lastly, the regression model for speech rate accounts for 21% of variance. Nonverbal IQ is a significant predictor, explaining most of the variance (15%), while implicit statistical learning approaches significance and accounts for 6% of variance in the model.

It is worth noting that these models, while informative, do not explain as much variance as the previously discussed models for standardized measures of language.

Table 7. Regression analyses: predictors of linguistic measures

Variable	Estimate (β)	Std.error	*t*-value	*p*-value	95% CI	lmg
Receptive grammar ⋆						
(Intercept)	2.47	2.47	0.99	0.32	-0.72, 5.66	
Nonverbal IQ	0.37	0.08	4.42	***	0.26, 0.48	.39
Working memory	0.20	0.08	2.37	*	0.01, 0.38	.11
Model R^2						**.50**
Adjusted R^2 = .49						
$F(2,75) = 38$, $p = < 0.001$						
Productive grammar ⋆						
(Intercept)	5.87	9.39	0.62	0.53	-8.81, 20.56	
Nonverbal IQ	0.89	0.28	3.10	**	0.39, 1.38	.18
Working memory	1.19	0.42	2.87	**	0.33 – 2.05	.13
Model R^2						**.31**
Adjusted R^2 = .29						
$F(2,75) = 16.96$, $p = < 0.001$						
Receptive vocabulary ⋆						
(Intercept)	2.41	16.40	0.15	0.88	-30.26 – 35.08	
Age	8.44	2.22	3.80	***	4.00 – 12.87	.18
Nonverbal IQ	1.21	0.39	3.10	**	0.43 – 1.99	.16
Working memory	1.19	0.71	1.69	0.09	-0.21 – 2.60	.10
Model R^2						**.44**
Adjusted R^2 = .41						
$F(3,74) = 19.19$, $p = < 0.001$						

Variable	Estimate (β)	Std.error	*t*-value	*p*-value	95% CI	lmg
Productive vocabulary						
(Intercept)	15.46	9.37	1.65	0.10	-3.20 – 34.11	
Age	2.38	1.15	2.08	*	0.09 – 4.67	
Model R^2						**.05**
Adjusted R^2 = .04						
$F(1,76) = 4.3$, $p = 0.04$						
Subordination Index						
(Intercept)	0.74	0.12	6.14	***	0.50 – 0.98	
Age	0.03	0.02	2.19	*	0.00 – 0.06	.09
Nonverbal IQ	0.00	0.00	3.01	**	0.00 – 0.01	.14
Model R^2						**.23**
Adjusted R^2 = .20						
$F(2,75) = 10.9$, $p = 0.04$						
Mean length of T-units						
(Intercept)	4.75	1.09	4.35	***	2.57 – 6.92	
Nonverbal IQ	0.13	0.03	3.45	***	0.05 – 0.21	
Model R^2						**.14**
Adjusted R^2 = .12						
$F(1,76) = 11.9$, $p = < 0.001$						
Disfluencies						
(Intercept)	-5.47	6.80	-0.80	0.424	-19.02 – 8.08	
Age	2.39	0.83	2.86	**	0.72 – 4.05	
Model R^2						**.10**
Adjusted R^2 = .08						
$F(1,76) = 8.2$, $p = < 0.01$						
Speech rate						
(Intercept)	1.25	0.34	3.63	***	0.56 – 1.92	
Nonverbal IQ	0.03	0.009	3.60	***	001 – 0.05	.15
Implicit statistical learning	0.01	0.005	1.98	0.052	-0.00 – 0.02	.06
Model R^2						**.21**
Adjusted R^2 = .18						
$F(2,75) = 9.81$, $p = < 0.001$						

★ *denotes Eicker-Huber-White correction for heteroscedasticity.*
**** p < 0.001; ** p < 0.01; * p < 0.05*
Observations = 78

Discussion

The present study was designed to explore cognitive predictors of linguistic attainment in children. Whilst modular theories propose a distinction between grammar, vocabulary and lexicon, usage-based theories of language acquisition specify that language structure emerges from a combination of language use and general cognitive abilities. We selected four cognitive variables of interest: nonverbal IQ, working memory, implicit statistical learning and automatization. Firstly, we discuss our findings relating to standardized language assessments in comparison to spontaneous speech measures. Then, we discuss the cognitive predictors in our investigation.

Overall, our cognitive predictors account for more variance in performance on the standardized language assessments than spontaneous speech measures. This could be because standardized tests are more reliable (in this study, our spontaneous speech samples were relatively small). Alternatively, performance on the spontaneous speech tasks may depend more strongly on other factors: for example, creativity and motivation, whereas there is not much scope for creativity on standard tasks.

The most striking finding is the robust relationship between nonverbal IQ and receptive grammar. The simple correlation between these two measures (r = .69) is stronger than the correlation between nonverbal IQ and receptive vocabulary (r = .53) (the difference between the two correlations is statistically significant, Steiger's $Z = 2.20$, $p < 0.05$). Furthermore, in the regression analysis, IQ accounted for 39% of the variance for receptive grammar and only 16% of the variance in receptive vocabulary. As discussed in the introduction, the relationship between nonverbal IQ and vocabulary is well-established (Raven, 2000), to the extent that vocabulary tests are often part of verbal IQ measures. In fact, the Peabody Picture Vocabulary

Test (PPVT: Dunn & Dunn, 2012), an American test on which the British Picture Vocabulary Scale (BPVS: Dunn et al., 2009) is based on, was originally marketed as a verbal intelligence measure (Bell, Lassiter, Matthews, & Hutchinson, 2001). The BPVS is described as an achievement test of receptive vocabulary and a screening test of verbal ability (i.e., verbal intelligence). More importantly, our results demonstrate that nonverbal IQ is a more significant predictor of grammatical development than vocabulary, and therefore contradict the widely accepted assumption that grammar is independent of IQ. This pattern is consistent with previous research in adults which demonstrate a robust relationship between nonverbal IQ and grammar (Dąbrowska, 2018, 2019).

Age correlates with fluid intelligence (r =.37) in our sample. Age also correlates with all our language measures (with correlations from r = .16 for implicit statistical learning to r =.54 for receptive vocabulary). However, when both age and raw nonverbal IQ are entered into the regression, nonverbal IQ emerges as the only significant predictor for receptive grammar, productive grammar and mean length of T-units. Furthermore, for receptive vocabulary and subordination index, age and IQ appear to make independent contributions. These relationships suggest that the correlations between age and linguistic ability are largely driven by age-related increases in IQ. Our results show that age was the only predictor of productive vocabulary - this could be because it takes time to entrench newly acquired words until they can be accessed in real-time production. Age is also the only predictor of disfluencies which is possibly a developmental effect, as older children attempt more difficult structures and therefore potentially produce more disfluencies.

Our results show that working memory is the second most important predictor of language, relevant for receptive grammar, productive grammar and receptive vocabulary -

although the effects are smaller than for nonverbal IQ. The relationship between working memory and grammar could be due to the processing demands of acquisition, as constructing grammatical representations places considerable demands on processing resources (Daneman & Case, 1981). For receptive vocabulary, working memory is the third most important predictor, after age and IQ. Since the demands of the task used in our study (BPVS) are minimal, the effect is most likely due to the difficulty of inferring meaning during acquisition.

Our measure of implicit statistical learning is significant to our model of fluency (speech rate), but the effect is small (only 6% of the variance). We did not find a relationship between ISL and any other linguistic outcomes, nor did our measure correlate with any of our cognitive predictors. There is universal agreement that ISL is relevant for language. However, the evidence to date is varied, as shown by recent meta-analyses (Boeve et al., 2023, Ren et al., 2023; Oliveira et al., 2023; West et al., 2021). Our results demonstrate group performance at chance which indicates that our measure could be problematic, with relatively low reliability of .52 (95% CI [0.30, 0.68]). However, this is a problem for all measures of ISL, especially for children. Over the past few years, concerns have been raised regarding the psychometric properties of many of the tasks used to measure ISL (Arnon, 2020; Siegelman & Frost, 2015). Furthermore, it has been suggested that forced-choice tasks may underestimate children's implicit learning abilities (Isbilen & Christiansen, 2022). Reflecting upon the performance of participants in our study, they found the task difficult and struggled to maintain attention. We are not alone in reporting difficulty of ISL tasks in children. Arnon (2020) raises similar concerns, stipulating that simplifying statistical learning tasks is not an easy thing to do. There is ongoing work in the field, and we welcome initiatives such as that of Boeve et al. (2023), who invite analysis and additional studies based on their data to improve theory and methodology of ISL.

We used two measures of the speed of automatization (AutoDiff and AutoCV). AutoCV correlated with nonverbal IQ (r = -.34), and also receptive grammar (r = -.30) but these relationships were not relevant in regression analyses. AutoDiff, on the other hand, was not related to any of our language measures. As discussed in the introduction, automatization has been widely explored in second language learning, and it may be that it is not relevant to first language acquisition in typical populations. Conversely, in a parallel study, we find that AutoCV is a significant predictor of language in children with DLD (Blake, Dąbrowska, & Riches, in preparation), and so the current results are difficult to interpret. However, the fact that AutoCV and AutoDiff do not correlate suggests that they may be measuring different aspect. Split-half reliability, based on odd/even trials was .64 for odd trials (95% CI [0.49, 0.76], and .62 for even trials (95% CI [0.43, 0.76], indicating acceptable reliability for this exploratory measure.

There are potential limitations that should be considered in this study. While a clear pattern emerged from our data, with age, nonverbal IQ, and working memory significantly predicting linguistic outcomes, the regression models were not corrected for multiple comparisons, which poses the risk of Type I errors. The sample size of 78 participants is slightly below the recommended 82 participants for optimal statistical power (based on the general rule of thumb of $N \geq 50 + 8\ m$, for multiple correlations, where m is the number of independent variables: see Green, 1991; VanVoorhis & Morgan, 2007). The observed results should be interpreted in light of these caveats.

Conclusion

The present study was designed to determine the effect of cognitive predictors on linguistic outcomes in language-typical children. Out of the four cognitive predictors that were measured, our study revealed that nonverbal IQ and working memory had the highest predictive capacity of linguistic outcomes. Furthermore, we report a strong relationship between nonverbal IQ and grammar. Overall, our results support the idea that grammar does not develop purely implicitly and that both explicit and implicit processes are involved in children's language acquisition. These findings are interesting from a theoretical perspective in that they do not align with modular theories and are fully consistent with usage-based models, which assume a grammar-lexicon continuum.

Acknowledgements: Thank you to the children who participated in our study and to their families for facilitating the online format of this study during COVID-19. This project was funded by an Alexander von Humboldt Professorship (ID-1195918) awarded to the second author.

4.3 Study one (b): Blake, Dąbrowska, and Riches (under review).

This section presents a manuscript entitled:

Exploring cognitive predictors of language in children with developmental language disorder: the case for nonverbal IQ, working memory, implicit statistical learning, and speed of automatization

This paper refers to study one of this thesis and involves an analyses of cognition and language in LT children and children with DLD.

This study investigated how various aspects of nonverbal cognition relate to linguistic outcomes in children with and without DLD. Firstly, we examined cognitive predictors of group membership (LT or DLD), by investigating specific cognitive abilities of nonverbal IQ, working memory, implicit statistical learning, and speed of automatization. Secondly, we explored the role of these predictors in relation to grammar, vocabulary, and various linguistic measures derived from elicited speech samples (narrative and expository discourse). This study involves 54 LT children and 23 children with DLD (mean age; 8.3 years) and refers to the same sample of children as described in study one (a).

EXPLORING COGNITIVE PREDICTORS OF LANGUAGE IN CHILDREN WITH DEVELOPMENTAL LANGUAGE DISORDER: THE CASE FOR NONVERBAL IQ, WORKING MEMORY, IMPLICIT STATISTICAL LEARNING, AND SPEED OF AUTOMATIZATON

Ashley Blake[1], Ewa Dąbrowska[2], Nick Riches[3]

[1]University of Birmingham, [2]Friedrich-Alexander-Universität Erlangen-Nürnberg

[3]Newcastle University

Abstract

Children with developmental language disorder (DLD) have varied strengths and difficulties in both language and cognition, but we do not yet have a comprehensive understanding as to how these abilities are interrelated. In this study, we explore performance in language typical children and children with DLD to evaluate how different areas of nonverbal cognition predict linguistic outcomes. We investigate nonverbal IQ, working memory, implicit statistical learning, and the speed of automatization. 77 children (54 language typical children and 23 children with DLD), aged between 6;9 and 10;8 years, completed a battery of cognitive and language tasks. Our results show between group differences in both language and cognitive abilities. We propose a cumulative risk model, suggesting that most significant predictors of language in children with DLD are a combination of nonverbal IQ, working memory, and the speed of automatization.

Keywords: language acquisition, cognitive abilities, developmental language disorder, speed of automatization

Introduction

Background

Language acquisition is a remarkable feat that occurs seemingly with little difficulty for most children in their first few years of development. However, children do not always

converge on the same grammar (Dąbrowska, 2015; Dąbrowska & Street, 2006; Street & Dąbrowska, 2010, 2014), and some children have language difficulties due to neurodevelopmental differences. There have been many terms used to refer to this group of children in the past, the most common being specific language impairment (SLI). Diagnosis of SLI is determined by low language ability, yet normal nonverbal intelligence and cognitive abilities (Reilly et al., 2014). However, there are numerous reasons why this terminology is problematic. Firstly, the term SLI implies that difficulties are language specific and yet SLI does not only affect language: children with SLI also perform poorly in standardized assessments across multiple domains, e.g., nonverbal IQ tasks (Gallinat & Spaulding, 2014), temporal processing (Tallal & Piercy, 1973a, 1973b, 1974, 1975), and other paradigms (see below). Secondly, historical lack of agreement in diagnostic criteria relating to SLI, particularly in the context of comorbid conditions, meant that many children with language difficulties were not able to access specialist services (Bishop et al., 2016, 2017). Finally, variability in terminology has hindered research development in the past (Bishop, 2010; McGregor, 2020). In response to these issues, a multinational and multidisciplinary consortium was formed in 2016 with the aim of reaching consensus on criteria for identification of language impairments in children. As a result, the term Developmental Language Disorder (DLD) was recommended for language difficulties not explained by biomedical conditions such as Down Syndrome, Autism Spectrum Disorder, or hearing loss, and inclusive of children with low IQs (Bishop et al., 2016, 2017). In this paper we use the term DLD to refer to children with language difficulties.

DLD occurs in approximately 7.5% of the population (Norbury et al., 2016; Tomblin et al., 1997), and is characterized by difficulties with learning, understanding and producing language (Bishop, 2017; Leonard, 2014a). Although language skills usually progress with

development, the language gap between children with DLD and language typical peers is maintained (Norbury et al., 2017), and many have argued that language difficulties may lead to poor cognitive, social, and educational outcomes (Conti-Ramsden et al., 2018; Johnson et al., 2010; Ziegenfusz et al., 2022).

Linguistic difficulties include learning new words (Alt et al., 2004; Alt & Plante, 2006; Gray, 2003; Kan & Windsor, 2010), word naming or word-retrieval difficulties (Messer & Dockrell, 2006), poor phonological awareness (Bishop et al., 2017; Leonard, 2014a), semantics and pragmatics (Adams, 2002), and difficulties with complex sentences and some aspects of inflectional morphology (van der Lely, 2005). Children with DLD show slow development of tense and agreement marking such as the English past tense (*-ed*) and third person plural (*-s*) inflections (Rice et al., 1995, 1998; Rice & Wexler, 1996). Language difficulties frequently co-occur with weakness in cognitive processes such as inhibitory control and cognitive flexibility (Pauls & Archibald, 2016), motor skills (Bishop, 2002; Hill, 2001), working memory (for reviews see: Archibald, 2017; Henry & Botting, 2017; Montgomery Magimairaj, & Finney, 2010), speed of processing (Zapparrata et al., 2023), and implicit statistical learning (see Lammertink et al., 2017; Lum et al., 2014; Obeid et al., 2016). Overall, the picture is complex and characterized by comorbidity of multiple language difficulties of which some combinations are more frequent than others, e.g., expressive and receptive difficulties (Conti-Ramsden, Crutchley, & Botting, 1997; Conti-Ramsden & Botting, 1999; Lancaster & Camarata, 2019). Cognitive difficulties are relatively subtle and are often not assessed in therapeutic contexts as the focus is usually on supporting language development as the primary area of need. Because of the variability in the way DLD presents between individuals, DLD is best seen as a continuum, similar to the approach adopted with other neurodevelopmental

disorders such as autism or attention deficit hyperactivity disorder (Lancaster & Camarata, 2019).

There is strong consensus that DLD is a complex condition with a polygenic basis and consequently there are many interactive risk factors (Mountford et al., 2022). However, whilst children with DLD have strengths and difficulties in both language and cognition, we do not yet have a full understanding as to how these abilities are interconnected or how specific underlying cognitive factors work together to support language development in DLD. Furthermore, DLD is typically heterogeneous in nature which adds to the complexity of understanding the needs of people with DLD and how best to support them. A cumulative risk approach has been examined by Hayiou-Thomas, Smith-Woolley, and Dale (2021), using data from a sample of 210 children in the Twins Early Development Study (a longitudinal twin study in the mid-1990s). Overall results from this study suggest that the presence of multiple-risk factors at four years of age (such as family history of language or reading difficulty, and poor language, speech, and emerging literacy and nonverbal skills), increase the likelihood of poor oral language, reading fluency and reading comprehension at the age of 12. The authors propose that a cumulative risk model increases the probability of difficulties far more than overall severity of deficits. Furthermore, even if risks are not detected in early childhood, language difficulties may still emerge later in development. Astle, Holmes, Kievit, and Gathercole (2022) challenge diagnostic thresholds that dominate research regarding neurodevelopmental disorders (such as DLD), and suggest that these methods do not capture the wide variability typically seen in neurodevelopmental conditions. Instead, they advocate for a more inclusive transdiagnostic approach, using dimensional or clustering methods to provide an in-depth assessment of cognition, behavioral and neural features. A deeper

understanding of non-linguistic abilities is important, both in terms of adding to our theoretical knowledge, and to inform therapeutic interventions in the future.

In the present study, we explore the possible effects of four cognitive traits: nonverbal IQ, working memory, implicit learning, and the speed of automatization, and examine how these processes predict linguistic abilities in children with and without DLD. Our motivation for selecting these cognitive predictors is discussed below.

Nonverbal IQ

General intelligence has traditionally been subdivided into two categories: crystallised intelligence (Gc), the ability to use acquired knowledge and information (such as vocabulary), and fluid, or nonverbal, intelligence (Gf), the ability to solve problems through reasoning and abstract thinking, independent of acquired knowledge (Cattell, 1943, 1963; Cattell et al., 1941). An individual's intelligence quotient (IQ) is typically measured by psychometric tests which usually assess several different abilities. However, standardized tests only capture certain elements of intelligence and do not measure criteria such as creativity, common sense, cultural differences, and many other factors (Neisser et al., 1996). It is generally accepted that there is a robust relationship between nonverbal IQ and vocabulary (Raven, 2000). Conversely, IQ is not considered relevant to grammar (Feldman, Kerr, & Streissguth, 1995; Reber, Walkenfeld, & Hernstadt, 1991), due to the consensus that grammar relies on implicit processes (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). However, some studies indicate that this may not be the case – for example, Dąbrowska (2018) observed a significant relationship between nonverbal IQ and grammar in adult native speakers (r = .46). Additionally, in a population-based sample of children with DLD, Lancaster and Camarata's (2019) results show that

receptive grammar is more significantly correlated with IQ than other linguistic variables, e.g., vocabulary. It is also worth noting that a diagnosis of SLI previously required a mismatch between linguistic abilities and IQ. However, several studies have found that children with SLI frequently score below than language-typical children in nonverbal IQ assessments (Gallinat & Spaulding, 2014, although see Lancaster et al., 2024 for possible reasons for this), and diagnostic criteria for DLD does not stipulate that there should be a difference between verbal and nonverbal ability (Bishop, 2017; Bishop et al., 2017). It remains to be seen to what extent the nonverbal abilities are predictive of language status (language typical vs. DLD); to investigate this issue, we include measures of nonverbal abilities in our study.

Working memory

Defined as a limited capacity cognitive system, working memory enables the temporary storage and processing of information (Baddeley, Hitch, & Allen, 2021). Although numerous models exist, here we refer to the Baddeley and Hitch model which has been widely used to explore working memory in DLD (Baddeley, 2000; Baddeley & Hitch, 1974). This model consists of four components, a central executive (CE) responsible for attentional control, which is supported by sub-systems for phonological and visuospatial information, and the episodic buffer which forms an interface to long-term memory.

Research suggests that both language processing and language acquisition make demands on working memory in typical populations (for a review, see Baddeley, 2003). Therefore, it is not surprising that different aspects of working memory capacity affect language functioning in DLD (for reviews, see 2017; Henry & Botting, 2017; Montgomery et al., 2010). A key area of difficulty is verbal short-term memory (Archibald & Joanisse, 2009;

Bishop et al., 1996; Botting & Conti-Ramsden, 2001; Gathercole & Baddeley, 1990). Phonological working memory is required to store a word for a short period of time to extract phonological and semantic characteristics – therefore, if children have difficulties in holding verbal information in mind, then this will have an influence upon their vocabulary acquisition. Poor phonological working memory is demonstrated by weak performance in non-word repetition tasks and is suggested to be a reliable clinical marker of DLD (Archibald & Joanisse, 2009; Bishop et al., 1996; Conti-Ramsden, 2003; Conti-Ramsden et al., 2001).

Children with DLD also show poor performance in sentence recall tasks, typically used to measure functioning of the episodic buffer (Alloway, Gathercole, Willis, & Adams, 2004; and Archibald & Joanisse, 2009, although sentence recall potentially relies on a range of cognitive processes, see Riches, 2012). Another area of difficulty for children with DLD are tasks that involve CE functioning, however, research in this area is limited. A strong relationship has been found between verbal CE and language, which pose difficulty not only for language development, but also for understanding and following instructions and other classroom-based tasks (Henry & Botting, 2017). Gray et al. (2019) used a comprehensive working memory battery to explore working memory profiles in a large group of children classified as language typical, DLD, dyslexia and concomitant DLD and dyslexia. Their results established that the group with combined DLD/dyslexia had difficulty with a CE task involving number updating. Interestingly, they also found that working memory deficits were not closely aligned with diagnostic criteria, and do not consistently occur with DLD or dyslexia. It seems possible that these results could be due to the heterogeneity of DLD or the sample size in their study (302 children overall, but only nine with DLD and 44 with combined DLD/dyslexia). In the present study, we are interested in the role of the CE alongside other cognitive predictors, using a non-verbal variation of the backwards digit span task.

Implicit statistical learning

Implicit statistical learning (ISL) refers to the ability to detect and extract regularities in the environment. In language learning, this process begins early in development and infants as young as eight months of age can detect sequences of syllables from a continuous speech stream (Saffran et al., 1996). ISL has been explored in many different domains and in a variety of linguistic abilities. Theoretically, if people with DLD have difficulty with detecting statistical regularities, e.g., frequently co-occurring segments in artificial languages, then it is reasonable to suggest they may have atypical language learning (Bogaerts et al., 2021; Saffran, 2018). Previous research provides evidence of impaired ISL in DLD. For example, it has been found that children with DLD have difficulty in detecting transitional probabilities in a speech stream (Evans et al., 2009; Haebig et al., 2019; Mayor-Dubois et al., 2014). Further evidence is highlighted in several meta-analyses: for DLD: (Lammertink et al., 2017; Lum et al., 2014; Obeid et al., 2016), and in both DLD and wider populations: (Boeve, Zhou & Bogaerts, 2023; Ren, Wang, & Arciuli, 2023). However, contrary results are reported by Lammertink et al. (2020), showing similar performance on visual statistical learning in children with and without DLD and no relationship between ISL and literacy measures.

Research regarding ISL is mixed and there are several reasons for this. As illustrated by Bogaerts et al. (2021), many studies investigating ISL are ambiguous about how different theoretical processes are linked to language learning (e.g., procedural learning, implicit learning, sequential learning), or how they account for cognitive processes in learning. As a result, they suggest that it is difficult to distinguish between theories, their explanations of ISL, and how they are linked to language learning. Another issue is that statistical learning tasks used in research with children indicate low reliability and typically use group designs.

Statistically significant differences at the group level do not indicate that all children with DLD have difficulties with ISL as there is substantial variation in each group (see Arnon, 2020). Lastly, prior research has followed two separate literatures, namely, implicit learning and statistical learning (Arnon, 2019; Christiansen, 2019; Perruchet & Pacton, 2006). There has been little collaboration between the two and as a way of unifying these concepts, Conway and Christiansen (2006) use the term 'implicit statistical learning', a definition we adopt in this paper.

A wide body of empirical research has explored ISL deficits in DLD according to the Procedural circuit Deficit Hypothesis (PDH: Ullman, 2004; Ullman et al., 2020; Ullman & Pierpont, 2005), which proposes to account for both linguistic and non-linguistic deficits found in DLD. According to the PDH, brain structures that support procedural memory are impaired in children with DLD. Procedural deficits account for grammatical processing difficulties in DLD as well as deficits in non-procedural functions such as working memory which are dependent on the same brain structures (Ullman et al., 2020). On the other hand, the PDH suggests that declarative memory is largely unimpaired and people with DLD can compensate for procedural deficits by explicitly memorizing complex forms in chunks, which are stored in declarative memory. Where individuals have superior lexical/declarative memory, greater compensation takes place and possibly results in their grammatical deficits improving over time (Ullman et al., 2020; Ullman & Pullman, 2015).

A common method of assessment of implicit learning (widely used to measure procedural learning), is the serial reaction time (SRT) task (Nissen & Bullemer, 1987). The SRT task involves participants responding to a sequence of stimuli that appear in one of four locations on a computer screen. Participants are asked to react to the stimulus as quickly as

possible by pressing corresponding keys on a computer keyboard or button box, and unbeknown to them, the stimuli follow a deterministic or probabilistic sequence. Performance is measured by improvement in accuracy or reaction times.

There have been many studies assessing procedural learning in DLD. Several studies demonstrate impaired procedural learning in DLD (Desmottes et al., 2016; Gabriel et al., 2013; Hsu & Bishop, 2014; Lum et al., 2010, 2012, 2014; Tomblin et al., 2007), however contrary evidence is reported by Clark and Lum (2017), Lum and Bleses (2012), and Gabriel and colleagues (Gabriel et al., 2011, 2015). West, Vadillo, Shanks, & Hulme (2018) investigated the PDH in association with language, literacy, and arithmetic attainment, using several measures of implicit learning: the SRT, Hebb serial order-learning, and contextual cueing tasks. A relationship was found between declarative learning and language attainment, but implicit learning tasks demonstrated poor reliabilities and weak correlations with attainment measures. Although this study focused on typically developing children only and did not consider performance of children with DLD, the findings raise concern regarding the reliability of implicit learning measures (although see discussion points raised by Conway, Arciuli, Lum & Ullman, 2019, and Krishnan & Watkins, 2019). Another issue is that because the SRT task involves a visuo-motor sequence, it is not clear whether poor performance is confounded by problems with motor skills found in some children with DLD. A recent study by Lum and Clark (2022) used the SRT task to assess manual sequence learning together with eye tracking. The DLD group showed impaired sequence learning in the manual domain but there were no differences between typically developing children and the DLD group in oculomotor sequence learning.

Two recent meta-analyses have explored the relationship between the PDH and DLD and dyslexia. West et al. (2021) examined several tasks widely used to measure procedural learning, reporting group level deficits in children with language difficulties based on the SRT task: g = -.30, and the Hebb task: g = -.32. On the other hand, Oliveira, Henderson, and Hayiou-Thomas (2023) report a negligible relationship (r = .06) between procedural memory (as assessed by the SRT task), and language and literacy overall, and no association when examining the language-typical, dyslexia and DLD groups separately.

Overall, these findings demonstrate variability between research findings and necessitate the need to improve reliability and develop more fine-grained measures of implicit statistical learning (see Boeve et al., 2023).

The speed of automatization

Fluent language comprehension and production rely on being able to process linguistic information quickly and automatically, without much conscious effort. The ultimate goal is automaticity, yet the processes leading to fluent, automatic language processing occurs gradually, through experience, and by acquiring a considerable amount of knowledge (e.g., vocabulary). These processes have led some researchers to draw parallels between cognitive skill learning and the notion of language as a complex, cognitive skill (Chater, McCauley, & Christiansen, 2016; Chater & Christiansen, 2018; Johnson, 1996; Kamhi, 2019).

Most research regarding automatization tends to be in the domain of second language acquisition (DeKeyser, 1997; DeKeyser, 2001; DeKeyser, 2015). However, automatization is not usually explored in first language acquisition, probably because of the assumption that

one's native language occurs implicitly, without conscious effort (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). Several studies have linked automatization to dyslexia, for example, Nicolson and Fawcett (1990, 2007, 2019) suggest that automaticity deficits underlie reading difficulties in dyslexia. Similarly, Smith-Spark and Gordon (2022) propose that both automaticity and executive abilities are related to literacy problems in dyslexia. Lastly, automaticity has been explored in children and adolescents with and without dyslexia using the SRT task and Tower of Hanoi puzzle. The learning rate was similar in both groups, although the dyslexia group performed both tasks more slowly (Vakil et al., 2015). However, to the best of our knowledge, ours is the first study to use a multiple-trial version of the ToH puzzle to explore automatization in children with DLD.

Aims of the present study

The aim of the current study is firstly to establish cognitive predictors of group membership (i.e., language typical and DLD). As outlined in the introduction, children with DLD have strengths and weaknesses in both language and cognition, however, the relationship between the two is not well-established. We consider a cumulative risk approach in exploring several cognitive abilities to investigate their interaction in children with DLD. These abilities are nonverbal IQ, working memory, implicit statistical learning, and the speed of automatization.

Method

The following methodology is detailed in Blake and Dąbrowska (2024).

Participants

Participants were recruited by sharing information with primary schools and on social media platforms such as Facebook and X (formerly Twitter). Recruitment criteria specified that participants should be monolingual English speakers with: (a) typically developing language, or (b) a primary diagnosis of DLD. Our sample comprised 97 children (58 male, 39 female), with a mean age of 8;1 years, referring to the same sample as in study one (a): Blake, Dąbrowska, & Riches (under review).

Children with DLD were included based on an official diagnosis of DLD which was confirmed by parents and caregivers. In addition, we conducted further analysis to determine the accuracy of our groupings. Nitido and Plante (2020) and Plante and Vance (1994) recommend the validation of diagnostic criteria as vital for both precision and reproducibility of research, and to ensure that children that meet criteria for DLD are identified correctly. In our analysis, children were classified as DLD if they scored more than 1.25 SDs below the mean on three or more language measures, i.e., standardized language assessments of: receptive grammar (Test for the Reception of Grammar; Bishop, 2003), productive grammar (Recalling sentences sub-test of Clinical Evaluation of Language Fundamentals; Wiig, Semel, & Secord, 2013), receptive vocabulary (British Picture Vocabulary Scale; Dunn, Dunn, Styles, & Sewell, 2009), and a speech measure of syntactic complexity: the subordination index. This eliminated 19 LT children and 1 DLD child from our sample. Therefore, the final sample included 77 children: 54 LT children and 23 children with DLD (average age of 8;3 years: range: 6;9 to 10;4 years).

Materials

Language tasks

We used three standardized language assessments to explore receptive and productive grammar, and receptive vocabulary. Expressive language was assessed using language samples from a narrative and expository discourse task. These are described as follows.

Standardized language assessments

1. **Test for Reception of Grammar:** TROG-2 (Bishop, 2003). This is a measure of receptive grammar, consisting of 20 grammatical constructions (80 four-choice items), which increase in grammatical complexity as the child progresses through the task. The test is presented as a picture selection task; the child listens to a sentence which is read aloud by the experimenter and is asked to say the number of the picture that matches the sentence. The participant should answer all four items correctly to pass a block and testing is discontinued after five consecutive failed blocks.

2. **The British Picture Vocabulary Scale:** BPVS3 (3rd Ed). (Dunn, Dunn, Styles, & Sewell, 2009). The BPVS3 is a receptive vocabulary test. Each test item consists of four illustrations and the child is asked to select a picture corresponding to a word read aloud by the experimenter. The test begins at the appropriate level for the child's age and there are 14 sets of 12 items, which increase in difficulty. Testing is discontinued when a child reaches eight or more errors in a set.

3. **Recalling sentences sub-test of Clinical Evaluation of Language Fundamentals**: CELF-5 (Wiig et al., 2013). This task assesses a child's ability to recall and reproduce

sentences which increase in length and complexity. The test is discontinued after five consecutive sentences with four or more errors in each sentence. Sentence repetition tasks are considered as a reliable clinical marker of DLD in childhood (Conti-Ramsden, et al., 2001), and in adulthood (Poll et al., 2010).

Language measures based on narrative samples

A narrative and an expository discourse task were used to elicit speech samples. Linguistic measures from language samples have been widely used as both clinical and diagnostic measures in DLD (see below). Responses were audio recorded for later transcription, to derive measures of grammatical complexity, lexical diversity, and fluency from the samples. Below is a description of the two measures used and our reasoning for selecting them.

1. **Narrative task.** Narrative language assessments provide a wealth of information regarding a child's language development and have been found to be reliable indicators of language impairment (Bishop & Edmundson, 1987). A recent study explored the diagnostic consensus of language profiles of ten children, by a group of 243 speech and language pathologists (SLPs). In addition to standardized tests, 78% of SLPs report using additional measures such as narrative language assessments to support accurate diagnosis of DLD (Kuiack & Archibald, 2023). Here, we use 'The Frog Story', based on 'Frog, Where are You' (Mayer, 1969). Our version includes colour illustrations to make the story more appealing to children and to capture their attention (illustrations by Mònica Pascual Holder, made available by Esther Pascual). The child sees the picture story twice and on the second occasion, they tell the story in their own words.

2. **Expository discourse task**. Bliss (2002) defines expository discourse as language that is used to convey information. Language used in expository discourse is typically more syntactically complex than conversational language, with longer utterances and more use of subordination (Nippold et al., 2005; Scott & Windsor, 2000). As such, it is recommended as an effective measure of complex syntax, which is potentially challenging for children with DLD (Lundine & McCauley, 2016; Nippold et al., 2008)

   The expository discourse task used in our study follows methodology used by Nippold, Mansfield, Billow & Tomblin (2008). In this task, the experimenter asks the child to explain their favourite game, sport, or hobby, using prompts such as: 1) what is your favourite sport, game, or hobby? 2) why is it your favourite? 3) if I were an alien from outer space, how can you explain it to me? 4) what are the rules? and, 5) can you recommend some good strategies?

Five measures were extracted from the speech samples, which are described below.

1. Mean length of T-units in morphemes.
2. Subordination index: calculated by counting total clauses (including main and subordinate clauses) and dividing the number of clauses by the total number of T-units.
3. Productive vocabulary (VOCD): a vocabulary diversity measure which is computed by KIDEVAL in the Computerized Language ANalysis program (CLAN: MacWhinney, 2000).
4. Speech rate: number of syllables per second, divided by total time of the speech sample.
5. Disfluencies: computed by adding the number of filled pauses (e.g., ‘um’, ‘er’), retractions (words that are withdrawn and corrected), and repetitions (where a word or

phrase is repeated). The total count of these disfluencies is divided by the total time of the speech sample to provide an average rate of disfluencies per minute of speech.

Cognitive tasks

We use nonverbal cognitive tasks where possible, to reduce linguistic demands on children with DLD, and furthermore, to obtain measures of cognitive abilities which are unconfounded by the interaction with language. These are described below:

1. **The Multiple-trial Tower of Hanoi** (MtoH) task (Blake & Dąbrowska, 2024; Dąbrowska & Blake, 2022). The MToH is used as a measure of automatization, using the Tower of Hanoi (ToH) puzzle. The ToH is a wooden puzzle with three vertical rods mounted on a wooden base, and several round discs, differing in size (we use a four-disc version of the puzzle). At the beginning of each trial, the discs are arranged in ascending order on the leftmost rod, with the largest disc on the bottom. To solve the puzzle, the child moves the discs from the leftmost rod to the rightmost rod, following two rules: 1) to move one disc at a time, and 2) a larger disc cannot be placed on a smaller disc. The optimal solution is to solve the puzzle in 15 moves.

   The experimenter explains the rules and demonstrates legal and illegal moves. Following Beaunieux et al. (2006), the experimenter demonstrates the first move, by putting the smallest disc on the middle rod of the puzzle. The child is invited to do a practice trial and ask questions before starting the task, and then the child repeats the puzzle 30 times. During the last five trials, the child is asked to solve

the puzzle while performing a secondary task which involves tapping in response to a recording. The recording plays a sequence of the words ‘one’ and ‘two’, with a four second pause between the two words. Participants are asked to do the puzzle with their dominant hand and then tap with their non-dominant hand in response to the numbers (two taps in response to the word ‘one’ and one tap in response to the word ‘two’). The number of moves and time per trial is recorded.

2. **Raven’s Coloured Progressive Matrices** (Raven’s CPM: Raven, Raven, & Court, 1998). Raven’s CPM is a nonverbal intelligence test that has been widely used as a short-form assessment of IQ. The task comprises 36 visual analogy problems, and possibly also involves central executive and working memory components. In each diagrammatical problem, there is a ‘piece’ missing and the child is asked to select the missing piece from a choice of six options. The problems are simple to begin with but increase in difficulty as the task progresses. The child indicates their choice by saying the number that corresponds to the missing piece of the puzzle.

3. **Backwards Colour Span task** (Riches, 2012, adapted from Zoelch, Seitz, & Schumann- Hengsteler, 2005). This is a visual variation of the backwards digit span task, using different coloured balls which are placed in an opaque tube. Nine colours are used, equivalent to digit recall stimuli: e.g., black, white, red, green, pink, grey, and brown (equivalent to eight one-syllable stimuli), and one two-syllable stimulus: yellow (corresponding to the number seven). Before beginning the task, the experimenter shows the child the coloured balls and asks them to name each colour. The experimenter then advises the child that the balls will be placed into a tube, in a specific order. The child’s task is to watch carefully and try to remember the order

that they are placed in the tube, as they will then be asked what order the balls will come out of the tube. Because the balls are placed in the tube at the same end that they come out, the balls come out in reverse order. The task comprises blocks of six items, with block one containing two items per stimulus, three items per stimulus in block two, and thereon increasing in difficulty with each block. Four correct responses are required to pass a block, and the test is discontinued after three errors in a block.

4. **Embedded triplet task.** This is an implicit statistical learning task modelled on an original task by Arciuli and Simpson (2011) and developed for our study on the Gorilla Experiment Builder platform (https://gorilla.sc; Anwyl-Irvine et al., 2020). The task consists of visual stimuli of cartoon aliens, drawn by the first author. The storyline involves friendly aliens who have arrived from outer space, and they appear on the screen in the order that they line up to go into their spaceship. This forms the familiarization stage, and the stimuli are presented in triplets: ABC, DEF, GHI, JKL, appearing in the same order. Each triplet appears in the familiarization stream 18 times and in 6 instances, one alien is presented twice in a row, e.g., AABC. The child is advised to press the space bar when they see an alien appear twice – this is a cover task to ensure that children maintain attention. The test phase involves the original triplets and 'impossible triplets', created by combining one alien from each of the original triplets, i.e., AEI, DHL, GKC and JBF. In this phase, the storyline continues with the aliens returning to their spaceship, and the child's task is to help the aliens line up in the correct order. In each trial, the child sees six stimuli, one original triplet (group A), and one impossible triplet (group B), which are presented consecutively with a short pause after the first triplet. The child's task

is to choose which triplet appeared at the beginning of the task, by clicking group A or group B. There are 64 trials in total, with each base triplet and impossible triplet appearing 16 times.

Procedure

Data was collected during the COVID-19 pandemic, and therefore the research sessions were conducted online using the videoconferencing platform, Zoom (https://zoom.us). Participants' parents were asked to complete a consent form and background questionnaire for their child. This included information regarding the parents' education, employment status, and information regarding their child's language difficulties, if applicable.

The experimenter posted a Tower of Hanoi puzzle to each family, and asked parents to ensure that their child did not open the puzzle prior to the first session, so that they did not practice solving the puzzle in advance. After receiving the puzzle, the parents were invited to schedule the online study sessions with the experimenter. During the first session, the experimenter explained the overall purpose of the study to the child and advised that they were welcome to take breaks or stop at any time. The study was moderated with the experimenter guiding the children through the experimental tasks, explaining each task in turn. The screen sharing option on Zoom was used to share tasks, using a digital visualizer to project images to the child's screen where necessary. Parents were invited to sit with or near their child if they wished. There were three sessions in total, each lasting approximately 45 minutes to one hour. The tasks were administered in the following order:

Session 1 : Automatization (Multiple-trial Tower of Hanoi task)

Session 2 : Narrative: The Frog Story

Receptive grammar (TROG-2)

Productive grammar (Recalling Sentences sub-test of CELF-4)

Expository discourse

Session 3 : Receptive vocabulary (BPVS3)

Working memory (Backwards Colour Span)

Nonverbal IQ (The Raven's Coloured Progressive Matrices)

Implicit statistical learning (Embedded triplet task)

Results

Data processing

Linguistic measures based on narrative samples

Speech samples derived from the narrative and expository discourse tasks were transcribed and coded using the Computerized Language ANalysis program (CLAN: MacWhinney, 2000). Data were coded for pauses, filled pauses, repetitions, retractions, elision, T-units, and subordinate clauses. Approximately 10% of transcripts were double coded by one additional researcher to ensure accuracy. Once speech samples were transcribed, they were analyzed in CLAN using the KIDEVAL program (MacWhinney & Fromm, 2016), to provide a language profile for each child. Audio files were analyzed in Praat (Boersma & Weenink, 2022), using the Syllable Nuclei script v.2 (de Jong & Wempe, 2009) for fluency measures.

Because the spontaneous speech samples were relatively small, we averaged the narrative and expository speech samples to form a composite measure.

Measures of the speed of automatization

We derived two measures of speed of automatization, using data from the MToH task. The first measure is "AutoDiff" (difference in moves), based on the secondary task. AutoDiff is calculated by subtracting the average number of moves in the secondary task trials (26-30), from the average number of moves in the last five standard trials of the MToH (21-25). If the procedure for solving the puzzle is not fully automatized, then we expect the secondary task to affect performance. However, some children with DLD (n = 8) found the secondary task challenging and could not master reverse tapping, or they did not tap at all. Therefore, it was not possible to include these data in our analysis.

The second measure is "AutoCV" (the coefficient of variation), which is motivated by previous research in second language acquisition (see Hulstijn et al. 2009; Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998). The coefficient of variation measures automaticity by dividing the standard deviation of a participant's reaction time by their mean reaction time. Strongly automatized behaviours usually show very little within-participation variation and in the case of high levels of automaticity, often the coefficient of variation is reduced to zero. In our study we measured performance on the MToH task by calculating the standard deviation of a participant's number of moves, divided by their mean number of moves. Therefore, we expect automatized behaviour to reflect a decrease in AutoCV over numerous trails.

Summary statistics

The data were analyzed with R Statistical Software (ver. 4.2.1, R Core Team, 2022). Table 8 presents the descriptive statistics for both the language typical (LT) and DLD groups.

Regarding linguistic measures, in the receptive grammar task, some children in the LT group achieve scores near the upper limits, approaching ceiling effects. In the DLD group, the range of scores demonstrate heterogeneity; particularly in receptive grammar, productive vocabulary, and disfluencies, highlighting the variability commonly observed in children with DLD. In contrast, there is low variability in our measure of subordination index, suggesting more uniform performance across the DLD group.

Regarding cognitive variables, the LT group displays a broader range of scores in working memory, when compared to the DLD group. However, the overall mean score is lower in the DLD children. In both groups, there is less variability in scores in implicit statistical learning and automatization.

Raw scores are reported for nonverbal IQ in the DLD group due to the Raven's CPM standardized scores being provided in intervals of five points, limiting precision when converting from raw to standardized scores. While a diagnosis of SLI stipulated a nonverbal IQ of 85 or higher, it is important to note that a mismatch between verbal and nonverbal ability is no longer a requirement for a diagnosis of DLD (Bishop et al., 2017). However, in this study, all participants' raw scores on Raven's CPM correspond to standardized IQ scores of above 85. Although there is some overlap in the range of nonverbal IQ scores between the LT and DLD group, the mean scores differ significantly.

Reliability estimates (Spearman-Brown corrected) were computed for the two novel tasks, the ISL task and the MToH task, using the `splithalf` package (version 0.8.2; Parsons, 2021), with 5000 random splits. For the ISL task, in the LT group the reliability was 0.60 (95% CI [0.34, 0.76), and in the DLD group, -0.15 (95% CI [-0.60, 0.51). The latter indicates significant issues, which could be accounted for the small sample size and missing data in the DLD group for this task (n = 9).

For the MToH task, the split-half reliability was conducted by dividing the data into odd and even trials. This was done to account for account for stability of scores which usually occurs around the midpoint of this task, when the participant discovers the optimum solution to the puzzle. Subsequent trials usually show little variability after this point, although this is not the case for every participant. Therefore, to account for this, split-half reliability was assessed by dividing the data into odd and even subsets. For the LT group, the reliability for odd trials was 0.67 (95% CI [0.51, 0.80), and for the even trials, 0.66 (95% CI [0.46, 0.82]). For the DLD group, the reliability for odd trials was 0.53 (95% CI [0.19, 0.80), and for the even trials, 0.54 (95% CI [0.28, 0.74]).

Table 9 presents group comparisons and effect sizes. Comparisons were conducted using Welch two-sample *t*-tests, to account for unequal variance between the two groups (Welch, 1947). As can be seen, the differences between the two groups are significant (or approach significance, as is the case with AutoCV), and the LT group perform better in all measures. The only exception to this is implicit statistical learning (ISL) and disfluencies, where there is similar performance between both groups. The between-group differences are larger for linguistic variables, showing that the primary area of difficulty in DLD is language,

consistent with the definition of DLD. The ISL task was difficult for children and especially so for many children in the DLD group. This resulted in missing data of 8.47% overall: two children in the LT group and nine children in the DLD group.

Table 8. Descriptive statistics

Variables	Mean: LT (SD)	Range: (LT)	Mean: DLD (SD)	Range: (DLD)
Language measures				
Receptive grammar (standardized)	111.35 (7.30)	99.00 – 125.00	67.43 (11.46)	55.00 – 95.00
Receptive grammar (raw, blocks)	17.07 (1.39)	15.00 – 20.00	7.26 (3.14)	3.00 – 14.00
Productive grammar (standardized)	12.50 (2.31)	9.00 – 17.00	5.39 (1.53)	3.00 – 9.00
Productive grammar (raw)	52.96 (7.43)	39.00 – 69.00	22.35 (8.97)	8.00 – 41.00
Receptive vocabulary(standardized)	112.70 (11.56)	89.00 – 140.00	81.83 (9.48)	70.00 – 100.00
Receptive vocabulary (raw)	129.94 (14.25)	104.00 – 173.00	92.17 (12.36)	71.00 – 121.00
Productive vocabulary	36.57 (6.73)	19.04 – 51.28	28.94 (8.92)	6.78 – 42.55
Subordination index	1.27 (0.10)	1.11 - 1.52	1.10 (0.06)	1.00 – 1.29
Mean length T-units	8.98 (1.04)	6.76 - 10.98	6.63 (1.18)	3.58 – 8.89
Speech rate	2.82 (0.45)	1.30 - 3.50	2.32 (0.39)	1.68 – 3.13
Disfluencies	15.00 (6.47)	7.18 - 44.18	15.60 (7.80)	5.37 – 42.00
Cognitive measures				
Nonverbal intelligence	29.39 (4.35)	17.00 - 36.00	21.57 (5.33)	14.00 – 31.00
Working memory	15.05 (2.68)	9.00 - 21.00	10.39 (2.69)	4.00 – 14.00
Implicit statistical learning	51.37 (10.41)	36.00 - 83.00	50.43 (5.58)	41.00 – 48.00
AutoCV	0.04 (0.04)	0.00 - 0.17	0.06 (0.05)	0.00 – 0.20

Language measures: receptive grammar (TROG); productive grammar (CELF5: sentence repetition subtest), receptive vocabulary (BPVS3: total correct); productive vocabulary (vocabulary diversity: VOCD); subordination index, MLTU: mean length of T-units, speech rate (syllables per second); disfluencies (per minute). Cognitive measures: nonverbal IQ (Raven's CPM: total correct: raw score); working memory (backwards colour span: total correct); implicit statistical learning: percent correct); AutoCV: automatization (coefficient of variation: standard deviation divided by mean number of moves on the MToH task: trials 1-25). Raw scores are reported except where stipulated otherwise.

Table 9. Group Comparisons

Variables	*t*-test	Cohen's *d*	*p*-value
Language measures			
Receptive grammar	*t*(30.00) = 17.00	5.03	***
Productive grammar	*t*(61.00) = 16.00	3.36	***
Receptive vocabulary	*t*(50.00) = 12.00	2.81	***
Productive vocabulary	*t*(33.14) = 3.70	1.02	***
Subordination index	*t*(61.38) = 9.00	1.91	***
Mean length T-units	*t*(37.30) = 8.30	2.17	***
Speech rate	*t*(48.00) = 4.90	1.16	***
Disfluencies	*t*(36.00) =-0.32	0.09	0.80
Cognitive measures			
Nonverbal intelligence	*t*(35.09) = 6.21	1.68	***
Working memory	*t*(41.49) = 6.90	1.72	***
Implicit statistical learning	*t*(39.84) = 0.45	0.10	0.65
AutoCV	*t*(36.69) = -1.80	0.47	0.08

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Correlations

Pearson correlations are presented in Table 10 for the LT group, and Table 11 for the DLD group. As can be seen, in the LT group, nonverbal IQ correlates with most language measures, and there is a relationship between working memory and receptive vocabulary. By comparison, ISL and AutoCV do not correlate with any of the language measures, apart from a weak correlation between ISL and receptive grammar (r = .23) and speech rate (r = .27).

In the DLD group, IQ is the most significant predictors of language. It should be noted that there is a significant correlation between IQ and receptive grammar (r = .52), this is higher

than in the LT group (r = .37). However, comparison of these correlations using a Fisher's Z test was not statistically significant, Z = 0.712, p = 0.47. Results show negative relationships between ISL and several of the linguistic measures: receptive, and productive grammar, receptive vocabulary, and subordination. However, our ISL results are potentially unreliable; performance is at chance and many of the children did not complete the task because they found it too difficult. AutoCV correlates with receptive grammar (r = .41).

Table 10. Pearson's correlations for the LT group

	Age	***IQ***	***WM***	***ISL***	***AutoCV***	***RGr***	***PGr***	***RVoc***	***PVoc***	***SI***	***MLTU***	***Disf.***	***SR***
Age	-												
IQ	**0.47**												
WM	**0.35**	0.28											
ISL	0.13	0.24	0.21										
AutoCV	-0.09	-0.23	0.16	0.13									
RGr	0.12	**0.37**	0.29	0.23	-0.10								
PGr	**0.33**	0.18	0.25	0.01	-0.13	**0.43**							
RVoc	0.50	**0.34**	**0.34**	0.15	-0.10	0.28	**0.36**						
PVoc	0.09	0.09	-0.19	-0.01	0.03	0.13	0.14	0.04					
SI	0.24	**0.32**	0.16	0.04	-0.04	0.05	0.05	0.20	0.15				
MLTU	0.22	**0.35**	0.23	0.19	-0.10	0.17	0.11	0.13	-0.03	0.72			
Disf.	0.26	0.22	0.17	0.14	0.16	0.18	0.25	0.10	-0.06	0.11	0.15		
SR	0.16	**0.31**	0.00	0.27	-0.05	0.09	**0.31**	0.11	0.25	**0.33**	**0.46**	0.26	-

*IQ, nonverbal IQ; WM, working memory, ISL, implicit statistical learning task; RGr, receptive grammar, PGr, productive grammar; RVoc, receptive vocabulary; PVoc, productive vocabulary; AutoCV (co-efficient of variation as a measure of automatization - MToH task); SI, subordination index; MLTU, mean length of T-units in morphemes; Disfl, disfluencies per minute; SR, speech rate per second. * Values of .30 or above (in boldface) are significant at p <.05.*

Table 11. Pearson's correlations for the DLD group

	Age	*IQ*	*WM*	*ISL*	*AutoCV*	*RGr*	*PGr*	*RVoc*	*PVoc*	*SI*	*MLTU*	*Disf.*	*SR*
Age	-												
IQ	**0.30**												
WM	-0.02	**0.30**											
ISL	-0.25	0.17	-0.02										
AutoCV	0.02	0.22	**0.33**	-0.21									
RGr	**0.35**	**0.52**	**0.41**	**-0.56**	**0.41**								
PGr	**0.48**	0.21	**0.31**	**-0.48**	0.08	**0.54**							
RVoc	**0.38**	0.06	-0.17	**-0.62**	0.06	**0.33**	**0.41**						
PVoc	0.25	**0.31**	0.22	-0.05	0.05	0.18	**0.41**	0.20					
SI	**0.38**	**0.41**	-0.17	**-0.41**	0.21	0.27	0.10	0.01	0.25				
MLTU	**0.50**	**0.36**	-0.15	-0.06	-0.03	**0.36**	**0.58**	0.18	**0.54**	0.28			
Disfl.	**0.45**	**0.43**	-0.01	-0.06	0.29	**0.53**	**0.36**	0.17	0.17	0.15	**0.68**		
SR	-0.01	0.25	-0.03	-0.28	0.18	0.13	0.18	0.11	**0.42**	**0.37**	0.20	0.14	-

*IQ, nonverbal IQ; WM, working memory, ISL, implicit statistical learning task; RGr, receptive grammar, PGr, productive grammar; RVoc, receptive vocabulary; PVoc, productive vocabulary; AutoCV (co-efficient of variation as a measure of automatization - MToH task); SI, subordination index; MLTU, mean length of T-units in morphemes; Disfl, disfluencies per minute; SR, speech rate per second. * Values of .30 or above (in boldface) are significant at p <.05.*

Statistical models

The main aim of this study is to ascertain which cognitive predictors support language acquisition in DLD. However, there is often a high correlation between cognitive variables, which makes it difficult to determine whether certain variables are an important explanatory factor, when analyzed in isolation. To establish which cognitive variables uniquely predict the most variance, we fitted a logistic regression to analyze the relationship between our four cognitive predictors (nonverbal IQ, working memory, implicit statistical learning, and AutoCV), and group (i.e., language typical or DLD). The predictor variables were scaled prior to fitting the model to make the coefficients more interpretable. As implicit statistical learning

was not significant; it was removed from the final model. Model assumptions were tested with the `performance` package in R (version 0.10.1; Lüdecke et al., 2021).

Odds ratios in the model represent the odds of language differences (i.e., LT or DLD), as predicted by cognitive variables (LT children are coded as 0, and children with DLD as 1). Odds ratios are typically used in medical research to assess contributing factors to a relevant outcome of interest, and are increasingly used in social sciences (Osborne, 2006). Our final model shows that there is a higher probability of children with DLD having weaknesses in nonverbal intelligence, working memory and slow speed of automatization. A summary of the model is provided in Table 12.

Table 12. Logistic regression model showing cognitive predictors of group membership

	β	**St Error**	***Z* statistic**	***p*-value**	**OR[1]**	**95% CI[2]**
Intercept	-1.99	0.58	-3.44	***	0.14	0.03, 0.36
Nonverbal IQ	-1.14	0.44	-2.58	**	0.32	0.12, 0.72
Working memory	-2.57	0.80	-3.18	**	0.08	0.01, 0.29
Automatization	0.76	0.38	2.02	*	2.15	1.05, 4.91

OR[1]: Odds Ratio; CI[2]: Confidence interval
*Significance codes: 0 '***' 0.001 '**' 0.01 '*'*
McFadden pseudo-R^2: 0.56

To further investigate cognitive differences between our two groups, we conducted a median split with scores above the median classified as 'high' and below the median classified as 'low'. We found that all (100%: n = 23) children with DLD in our sample have deficits in working memory. Additionally, 61% of the sample have low working memory, low nonverbal IQ, and poor speed of automatization in comparison to the median of the LT group. This relationship is shown in Table 13.

Table 13. The number of DLD children falling above or below the median values for the LT group

Automatization	Low NVIQ	High NVIQ
High speed of automatization	n = 8	n = 0
Low speed of automatization	n = 14	n = 1

**All children in the DLD group have low working memory*

To evaluate the diagnostic potential of our logistic regression model, we extracted model metrics by conducting a Receiver Operating Characteristics (ROC) curve, using the `caret` package (version: 6.0.93; Kuhn, 2022), see Figure 9. The ROC curve plots sensitivity (true positives) versus specificity (true negatives) for each level of the predictor variable. A widely used measure of accuracy of the ROC curve is the value of the area under the curve (AUC), which represents the predictive power of the model. AUC values range from 0 to 1.0, and the higher the AUC value, the better the model can discriminate between classifications. An AUC value of 0.5 does not offer discriminatory classification whereas an AUC over 0.8 is considered good, and over 0.9, excellent (Nahm, 2022)

We conducted a three-fold cross validation so that the model was split into a training and test set, using different subsets of the data. The AUC estimate, including cognitive predictors of nonverbal intelligence, working memory and the speed of automatization was 0.9 which indicates excellent predictive properties of our model. Table 14 shows the confusion matrix which represents the prediction summary, and Figure 10 demonstrates classification accuracy based on different combinations of the predictors in the logistic regression model.

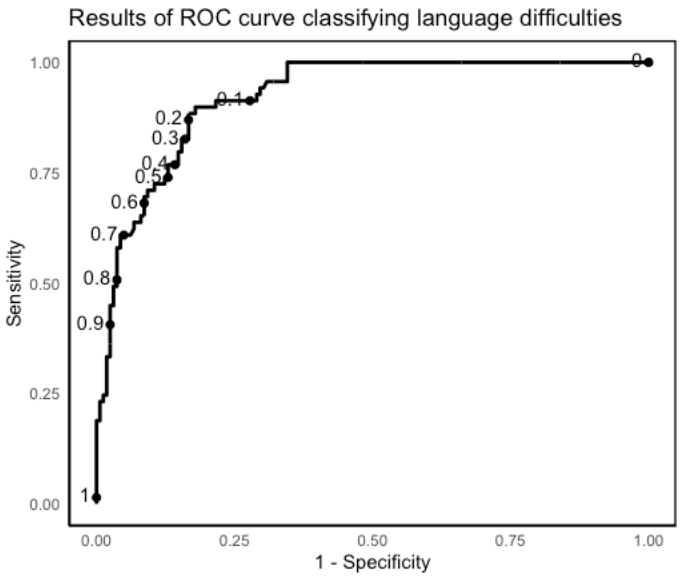


Figure 7. ROC curve demonstrating the predictive ability of the logistic regression model

Table 14. Confusion matrix

Prediction	**LT**	**DLD**
LT	47	6
DLD	7	17

**Based on a threshold of 0.5*

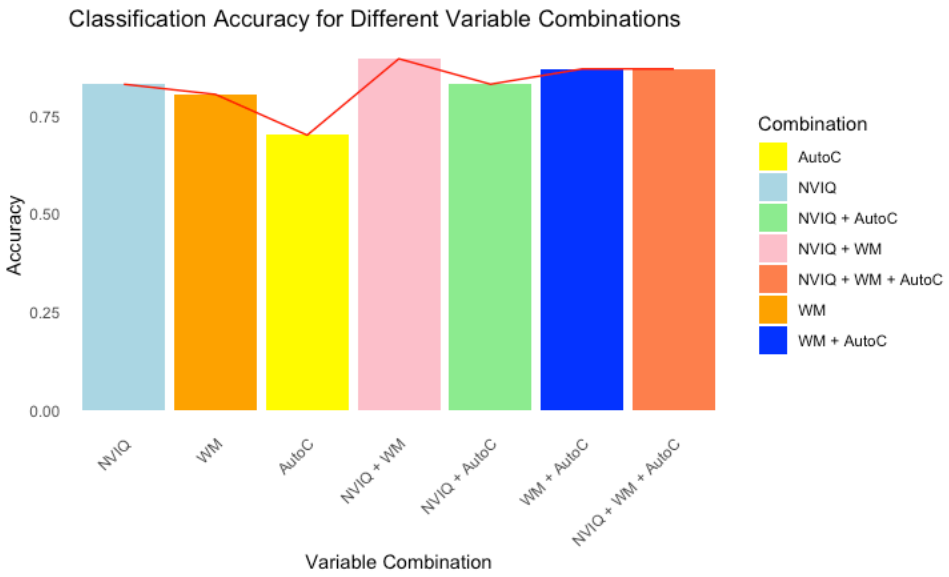


Figure 8. Classification accuracy of cognitive predictors

Discussion

The present study investigated cognitive predictors of linguistic performance in children with and without DLD. We examined specific cognitive abilities that determine group status (i.e., language typical or DLD). Whilst children with DLD generally performed below language typical children in linguistic tasks (by definition, DLD is determined by primary difficulties with language), we were particularly interested in the underlying cognitive processes that support linguistic proficiency. In addition to predictors of nonverbal intelligence, working memory, and implicit statistical learning, we investigated the speed of automatization using the MToH task. To the best of our knowledge, ours is the first study to explore automatization in children with DLD using this paradigm.

Nonverbal IQ is the strongest predictor of language outcomes in our sample, followed by working memory. It is interesting to note that in the LT group, the correlations between IQ and receptive grammar, as well as between IQ and receptive vocabulary, are similar. However,

in the DLD group, the relationship between IQ and grammar is much higher than that between IQ and vocabulary. Although the comparison of these correlations did not reach statistical significance (Fisher's Z test: $Z = 0.712$, $p = 0.47$), it is noteworthy that the relationship between IQ and receptive vocabulary, by comparison, is close to zero. As discussed in the introduction, consensus in the field is that nonverbal IQ is strongly related to vocabulary, but not applicable to grammar, because it is widely accepted that grammar develops through implicit processes. A similar relationship has been reported in adult native speakers (Dąbrowska, 2018), and we observe similar results in a large sample of children with DLD (Lancaster & Camarata, 2019). Our results suggest that grammar relies on problem solving abilities in both groups, but additionally shows that children with DLD process language differently to their LT peers. Grammar is a particular area of difficulty for children with DLD. According to the PDH hypothesis, children with DLD compensate for impaired procedural learning by using declarative memory for learning explicit rules and memorizing complex forms in chunks (Ullman et al., 2020; Ullman & Pierpont, 2005; Ullman & Pullman, 2015). Whilst this could be the case in our sample, our results are also in line with usage-based models which do not delineate grammar and lexicon, and instead suggest that both explicit and implicit processes interact at different stages of linguistic development and in differing degrees (Dąbrowska, 2010a; Divjak et al., 2022).

Our results show negative correlations between ISL and language outcomes in the DLD group. However, performance in the ISL task was at chance across the board. Some children were unable complete the task due to task difficulty and therefore we cannot interpret our findings with confidence. Developing suitable ISL tasks for children is a challenge and we are not alone in reporting task difficulty. Arnon (2020) examined the reliability of three statistical learning measures in adults and children (mean age 8;2), with a 2-month gap between testing

sessions. This included a visual ISL task modelled on the Arciuli and Simpson (2011) embedded triplet task that we adapted for our study. Their results indicated moderate reliability in adults (using internal consistency, split-half reliability, and test-retest reliability across sessions), but in children, whilst accuracy levels indicate that learning did take place, reliability was below psychometric standards. Arnon (2020) propose that measures used to detect statistical learning abilities in children may not be sensitive or stable enough to capture variability in ISL adequately, or potentially, that ISL is affected by other cognitive processes such as memory and attention. Other studies report that forced-choice tasks reveal weaker effect sizes in children (Isbilen et al., 2022), possibly due to the fact that they are required to reflect upon prior learning, which adds noise to the data (Christiansen, 2019). Conversely, a recent longitudinal study exploring the relationship between ISL and language in primary school children measured auditory and visual ISL every six months over four time points. Visual ISL was measured using the embedded triplet design that we modelled our ISL task from. At the first time point, scores did not indicate acceptable reliability; however, both accuracy and internal consistency increased over subsequent time points (Kidd et al., 2023).

The between-group differences showed larger effect sizes for our measures of linguistic ability, whereas we only found group differences in two of our cognitive measures: nonverbal intelligence and working memory. We suggest two possible reasons for this. The first reason is that cognitive effects are secondary, that is, they are already starting to occur in children with DLD in our sample. For example, Botting (2005) report a 23-point drop in nonverbal IQ scores in children with SLI between the ages of 7 and 14 years of age. Another longitudinal study investigating a slightly older age range, from 7 to 17 years, did not find a decline in nonverbal skills for the group, but when analyzing sub-groups, a decline in growth was found in nonverbal skills in some individuals with SLI (Conti-Ramsden, St Clair, Pickles, & Durkin, 2012). Whilst

the children in our study are slightly younger than the groups referred to in these two studies, it could be that there is slower growth in nonverbal skills by school-going age.

The second reason we propose is that language difficulties in children with DLD may be additive, caused by a grouping of cognitive factors. In our study, 61% (n = 14) of children in the DLD group have a combination of low nonverbal IQ, low working memory, and slow speed of automatization. These findings suggest that language difficulties can be explained by an association of cognitive difficulties. We suggest that if children have strengths in one of two areas of cognition, this provides a protective buffer in terms of language development, but a combination of these cognitive processes could result in linguistic difficulties. For example, strengths in nonverbal IQ would enable a child to process and analyze information quickly, reducing the cognitive load on other areas such as working memory. Similarly, good working memory capacity would support the temporary storage and processing of lexical units making it easier to retrieve words and form a grammatical sentence with ease. Speed of automatization is also important to linguistic outcomes, to facilitate fast and effortless processing which is especially important in forming grammatical structures. However, weaknesses in more than one area, say for example, a combination of low working memory capacity and slow speed of automatization, may cause an additive effect. In this case, the child may not have a large vocabulary knowledge from which to retrieve lexical items, or poor working memory capacity might mean that they are unable to remember what has been said to them. Furthermore, slower speed of automatization would make it difficult for the child, not only to retrieve the correct words, but also to draw upon syntactic rules to form an appropriate grammatical sentence.

The speed of automatization is underexplored in DLD and to the best of our knowledge, ours is the first study using the MToH paradigm in children with DLD. Our overall results

suggest that automatization plays a role in language acquisition for children with DLD. In this study, we planned to assess interference effects in the MToH task but this was not possible as many children found the secondary task challenging. As described earlier, the secondary task involved reverse tapping where children were instructed to continue to solve the puzzle whilst tapping in response to a recording: two taps in response to the word "one" and one tap in response to the word "two". Many children with DLD found this too difficult; this may indeed reflect their challenges with automatization. On the other hand, it could be indicative of motor difficulties commonly seen in children with DLD (Hill, 2001; Finlay & McPhillips, 2013). Another possibility is that reverse tapping was difficult for children with DLD due to executive function deficits, particularly affecting response inhibition (Pauls & Archibald, 2016). In this case, some children may have struggled to suppress the 'correct' response to the numbers "one" or "two", when asked to provide a different response.

There are practical approaches to supporting children with DLD with weak automatization. Kamhi (2019) provides specific recommendations for clinical interventions. In addition to standardized-tests, Kamhi recommends that clinicians could evaluate conversational, narrative and expository discourses to determine specific forms, structures and more complex units of discourse that can potentially be proceduralized. In this way, Kamhi suggests that children can be encouraged to learn specific phrases by rote so that these structures become routinized. According to the theory of cognitive skill learning, continual practice of conversational interaction will only help to improve language skills (Chater & Christiansen, 2018).

Further research is needed to explore these cognitive predictors in more depth and establish whether there are subcomponents that could further explain language difficulties in

children with DLD. Our results suggest that a combination of cognitive processes potentially underlie language difficulties commonly seen in DLD.

There are several potential limitations to be considered in light of our results. We did not assess different aspects of working memory, such as phonological working memory, which may be relevant for children with DLD and could contribute further to our findings. Additionally, the small sample size of DLD participants in this study may limit the statistical power of the results. While small sample sizes are common in DLD research, the findings reported here warrant further investigation, particularly in studies with larger sample sizes. In addition, we encourage further research regarding automatization in DLD populations, building upon the MToH paradigm.

Conclusion

In conclusion, our results suggest that specific cognitive abilities could act as protective factors in language development, namely nonverbal IQ, working memory and automatization. From a practical perspective, these findings suggest the importance of including general cognitive skills in therapies to support children with DLD, and as recommended by Kamhi (2019), specific interventions to support automatization of language (e.g., specific clauses and longer phrases and structures).

Acknowledgements: We thank the children who took part in our study and their families for their support. We also thank the many schools and organizations who helped in sharing information regarding our study during COVID-19. We also thank Justin Kamerman for advice and guidance on statistical analysis in R. This project was funded by an Alexander von Humboldt Professorship (ID-1195918) awarded to the second author.

4.4 Study two: Blake, Dąbrowska, and Llompart (under review).

This section presents a manuscript entitled:

Towards an understanding of developmental language disorder in adults: investigating the relationship between cognitive abilities and linguistic outcomes

This paper refers to study two of this thesis and builds upon the findings from the previous study which focused on DLD in children. While a wide body of literature explores the developmental trajectory of DLD in children, less is known about how nonverbal cognitive abilities relate to linguistic outcomes in adulthood. This study seeks to address that gap by examining whether the cognitive abilities identified in childhood continue to influence linguistic outcomes in adults with DLD. In this study, two novel cognitive abilities are explored; the role of language aptitude/language analytic ability, and the speed of automatization, two areas that have not previously been investigated in the context of adults with DLD. 30 LT adults and 30 adults with DLD (mean age; 38 years) took part in an online study and completed a battery of cognitive tasks investigating nonverbal IQ, language aptitude/language analytic ability, implicit statistical learning, pseudoword learning, and speed of automatization. Linguistic outcomes of interest were grammar (using a grammaticality judgment task and a picture selection task), and vocabulary.

TOWARDS AN UNDERSTANDING OF DEVELOPMENTAL LANGUAGE DISORDER IN ADULTS: INVESTIGATING THE RELATIONSHIP BETWEEN COGNITIVE ABILITIES AND LINGUISTIC OUTCOMES

Ashley Blake[1], Ewa Dąbrowska[1,2], Miquel Llompart [2,3]

[1]University of Birmingham, [2]Friedrich-Alexander University, Erlangen-Nuremburg, Germany,

[3]Universitat Pompeu Fabra

Abstract

There is a growing body of research regarding Developmental Language Disorder (DLD) in adulthood but very few studies focus on the role of cognitive abilities later in life. Here, we explore the relationship between cognition and language in adults with and without DLD. In the present study, 30 adults with language difficulties and 30 adults with typical language (LT) completed a battery of online cognitive and language assessments. Accuracy and reaction time measures were obtained for all tasks. Our results show group differences in linguistic measures, demonstrating that language remains an area of difficulty for people with DLD in adulthood. The ability to recognize the grammatical functions of sentence elements is a significant predictor of grammatical and vocabulary knowledge in our LT group, and it predicts grammar in the DLD group. We find small or no differences for all the other cognitive abilities; this is in line with previous studies which report a relative 'catch up' in cognitive abilities in adults with DLD. Response times for the DLD group were slower in both linguistic and non-linguistic tasks, indicating that adults with DLD have slower processing speed.

Keywords: developmental language disorder, language aptitude, processing speed, nonverbal IQ, implicit statistical learning, automatization.

Introduction

Background

People with developmental language disorder (DLD) have persistent difficulty understanding and producing language (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE consortiums 2016, 2017; Leonard, 2014a). Whilst language difficulties in DLD are sometimes resolved during development (Bishop & Edmundson, 1987), for many people, problems are persistent, with far-reaching consequences in other areas of life such as social and emotional health, education, employment, and independent living (Conti-Ramsden et al., 2018; Dubois et al., 2020; Johnson et al., 2010). However, many adults who experience language difficulties are not formally diagnosed with DLD. This could be due to a general lack of awareness, as DLD is not very well-known or recognised. Despite a high prevalence rate of over 7% in childhood (Norbury et al., 2016; Tomblin et al., 1997), estimates for adult populations are not known. Additionally, assessment tools used to identify DLD in children are not sensitive enough to detect higher-order language difficulties often seen in adults with DLD, and most do not have normative scores beyond adolescence (Del Tufo & Earle, 2020; Plante, Shenkman, & Clark, 1996).

Prior research examining the trajectory of language skills into adolescence and adulthood provide evidence of long-term language deficits in DLD (Botting, 2020; Clegg et al., 2005; Conti-Ramsden et al., 2018; Del Tufo & Earle, 2020; Earle et al., 2018; Law et al., 2009; Whitehouse et al., 2009). Word learning difficulties, which are evident in children with DLD, continue into adulthood. For example, a series of studies investigating word learning in adults with DLD suggest that weaknesses arise in the encoding stage, which is the initial phase of memory formation where a memory trace connects a word form with its meaning, yet they show strengths in the retention of new words (McGregor et al., 2013, 2017, 2017, 2020). Adults

with DLD have difficulties with written narratives (Suddarth et al., 2012), and have problems understanding legal language such as Miranda warnings, which are read to individuals at the time of arrest. This language is said to draw upon higher-order comprehension skills due to complex syntax, vocabulary, and pragmatics (Rost & McGregor, 2012). Additionally, in a study exploring DLD and dyslexia, Del Tufo and Earle (2020) established that college students with DLD have difficulty with verbal working memory, as well as reading fluency and comprehension under timed conditions, potentially affecting exam performance or rapid notetaking.

As shown above, language remains a persistent obstacle for many adults with DLD and therefore it is important to gain an understanding of the underlying cognitive abilities that support language. Previous research suggests a changing pattern of nonverbal skills in adults with a history of DLD from childhood (Botting, 2020; Clegg et al., 2005; Tomblin et al., 1992), yet this area is largely underexplored in the literature. Consequently, we lack an understanding as to how cognitive abilities support language in people with DLD later in life. Studying adult populations with DLD is difficult, partly because there is a great deal of heterogeneity in terms of language characteristics and severity of DLD. Additionally, because many adults are undiagnosed, it is difficult to recruit participants to studies.

In the present study, we explore several cognitive abilities to determine how they predict linguistic outcomes in adults with and without DLD. Our motivation for discussing these processes is discussed below.

Nonverbal IQ

Nonverbal IQ in DLD has received considerable attention in the literature, partly due to the change in terminology from Specific Language Impairment (SLI) to DLD. One of the reasons for this change of nomenclature was due to diagnostic criteria for SLI, which stipulated that verbal language should be below age expected levels (with scores of at least 1.25 SD below the mean on standardized language tests), and nonverbal IQ scores of 85 or higher (Reilly et al., 2014). This requirement has been viewed as exclusionary as it left many children ineligible for diagnosis and potentially without speech and language therapy (Dockrell et al., 2006). The new recommended terminology of DLD no longer considers nonverbal IQ scores as part of diagnostic criteria and thus encompasses a wider range of abilities (Bishop et al., 2017). In spite of this, it is worth noting that previous research shows that IQ does play a role in DLD, and generally children with DLD have lower scores than language typical children on tests measuring nonverbal cognition (Gallinat & Spaulding, 2014, although see Lancaster et al., 2024 for possible reasons for this).

Evidence from longitudinal studies report a difference in nonverbal IQ scores over different time points, with lower scores in adolescence in comparison to childhood assessments (Botting, 2005; Conti-Ramsden et al., 2012). Similarly, Tomblin et al. (1992) report a downward trajectory of nonverbal IQ scores in a group of 35 adults with DLD, with lower nonverbal IQ scores in their early twenties, to those recorded in childhood. In contrast, Clegg et al. (2005) show a different pattern of results. In a longitudinal study of a group of men with a history of DLD in childhood, they found that performance on IQ measures remained stable from early childhood to adolescence but a dramatic decrease followed in early adulthood: participants scored 14 points lower in nonverbal IQ than when they were tested during

adolescence. When the group were reassessed in their mid-thirties, performance IQ scores were similar to those recorded in childhood, demonstrating a potential 'catch up' (Clegg et al., 2005). Similarly, Botting (2020) reports a small but significant increase in nonverbal IQ scores in adults with DLD between the ages of 16 and 24 years. However, both Botting and Clegg acknowledge that differing scores could be attributed to the fact that different cognitive assessments were used to measure nonverbal IQ over different time points (see Lancaster et al., 2024).

Language aptitude/language analytic ability

The term 'language aptitude' is normally used to refer to the ability to learn a foreign language in instructional settings (Carroll, 1981, Carroll & Sapon, 2002). According to Carroll (1964, 1990), language aptitude comprises a set of specific cognitive abilities, namely phonetic coding ability (applicable to phonological learning of unfamiliar sounds), associative memory, and grammatical sensitivity and inductive language learning ability. The latter two tests are particularly relevant for grammar learning and are sometimes subsumed under the term language analytic ability (Li, 2015; Skehan, 2002). Traditional language aptitude tests such as the Modern Language Aptitude Test (MLAT: Carroll and Sapon, 1959), and the Pimsleur Language Aptitude Battery (PLAB: Pimsleur, Reed, & Stansfield, 1966) require slow, effortful processing and are reliant on metalinguistic awareness: in other words, they measure explicit language learning ability (Li, 2015).

However, language aptitude is not considered relevant to first language (L1) acquisition as it is generally accepted that L1 grammar develops through implicit processes (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). Yet recent research in L1 adult populations

suggests that language aptitude is indeed relevant for grammar. For example, Dąbrowska (2018) reports a significant correlation between language analytic ability and L1 grammatical comprehension in adults (r = .46). This result is supported by a recent studies by Llompart and Dąbrowska (2023), which also found a robust relationship between language analytic ability and tasks measuring grammatical comprehension and grammatical judgement (with similar effect sizes ranging from .43 to .48). An even higher correlation between grammatical sensitivity and receptive grammar in adult native speakers (r = .68) is reported by Winckel and Dąbrowska (2024). Prela, Dąbrowska and Llompart (2022) observed a slightly lower correlation (.37), possibly because performance on the grammar task in many participants was close to ceiling. It is striking that the correlations between foreign language aptitude and grammar observed in these studies are actually higher than the average correlation between these variables reported in a recent meta-analysis by Li (r = .31: Li, 2015).

The relationship between grammar and explicit learning is interesting from the perspective of DLD. As discussed earlier in this section, language aptitude involves explicit, metalinguistic awareness, for example, grammatical sensitivity (as measured by the MLAT), assesses the ability to determine the relationship between words and how a word fits into a sentence. Therefore, if language aptitude is relevant to L1 grammar, as shown above, then it is certainly plausible that it is relevant to language outcomes in adults with DLD.

Grammar is considered a particular area of difficulty in people with DLD, and according to the Procedural Circuit Deficit Hypothesis (Ullman et al., 2020), grammatical deficits are explained by impairment to neural regions underlying procedural (implicit) memory. Yet, it is suggested that declarative (explicit) memory is intact and that this system plays a compensatory role in DLD (Ullman et al., 2020; Ullman & Pierpont, 2005; Ullman &

Pullman, 2015). It has been found that children with DLD compensate for grammatical deficits (such as the English past tense inflection), by relying on explicit learning (Pinker, 1991; Pinker & Ullman, 2002). If this were the case, then we would expect to find a relationship between language aptitude and grammar in people with DLD, but not in language typical individuals.

Implicit statistical learning

Implicit statistical learning (ISL) refers to the ability to detect regularities in the environment and is evident in a broad range of domains and modalities, including language (see meta-analysis by Isbilen & Christiansen, 2022). Recent meta-analyses demonstrate relatively small, but significant correlations, for example: Boeve et al. (2023) report an effect size of r = .16 between ISL performance and language skills (vocabulary, syntax, and comprehension). Ren et al. (2023) investigate ISL and two outcomes: language-related and reading-related skills, with a similar effect size for both (r = .236 for language and r = .239 for reading).

A number of studies suggest that people with DLD have deficits in ISL (see meta-analyses: (Lammertink et al., 2017; Lum et al., 2014; Obeid et al., 2016). However, research findings are mixed, and some studies provide contradictory results. For example, Lammertink et al. (2020) did not find evidence of a visual statistical learning deficit in DLD, and West et al. (2021) suggest that there is insufficient evidence to suggest that impaired procedural learning is a causal factor in DLD and dyslexia. Additionally, a recent meta-analysis demonstrates limited support for a relationship between procedural learning and language and literacy measures (r = .06 for the overall sample which included language typical, DLD and dyslexic participants, but these relationships did not retain significance when examined

separately). The authors suggest that their findings are linked to poor psychometric properties of the serial reaction time task which is frequently used to measure ISL and procedural learning (Oliveira et al., 2023).

An issue that has been highlighted by a number of researchers is that there are different approaches used to measure the relationship between ISL and language, which has led to considerable variability between studies (Arnon, 2020; Bogaerts et al., 2021; Siegelman et al., 2017; Siegelman & Frost, 2015). Furthermore, evidence suggests that ISL may be modality specific (Conway & Christiansen, 2006), although results from Boeve et al.'s (2023) meta-analysis provide support for cross modal stimuli as a promising method of exploring ISL. In view of these issues, here we investigate ISL using a visual statistical learning (VSL) task developed by Siegelman, Bogaerts, and Frost (2017). The authors account for some of the issues in ISL tasks by revising their original design (see Siegelman & Frost, 2015), to include sufficient trials, minimal repeated patterns, and varying difficulty of test items. In a study using their revised VSL task, adults achieved 60% learning and test-retest reliability is reported at .68 (Siegelman et al., 2017).

The ability to memorize form-meaning pairings/phonological encoding skills

Very few studies have assessed word learning in adulthood. Most evidence to date is derived from a series of studies that explore memory processes underlying word learning in a group of college students with and without DLD (McGregor et al., 2013, McGregor, Gordon et al., 2017, McGregor, Arbisi-Kelm, et al., 2017, McGregor et al., 2020). These studies examine whether difficulties arise from encoding verbal information (referring to the initial memory trace which links a new word form and its meaning), or if difficulties occur in long-

term retention. Participants were trained on a set of novel words and their referents, with word recall assessed immediately after training and again one week later. Collectively, the results from these studies reveal that the encoding stage is the bottleneck to learning new words, yet long-term retention is a strength. McGregor et al. (2020) suggest that the pattern of word learning in adults with DLD is consistent with the PDH (Ullman et al., 2020), as encoding phonemic sequences could be dependent upon procedural memory. However, further research in adult populations is needed. A recent scoping review investigating word-learning abilities in people with DLD identified only five studies focused on young adults (out of a total of 70 studies), and most of these studies focused on adults up to 25 years of age (Jackson et al., 2019b).

The speed of automatization

In order to speak a language fluently, a speaker relies on the ability to process a large amount of information rapidly, and typically without much conscious effort. The term 'automaticity' is therefore typically associated with fluency (Segalowitz, 2003), and is widely explored in the context of second language learning (DeKeyser, 1997; DeKeyser, 2001; Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998; DeKeyser, 2015). DeKeyser (2015) considers automaticity in L2 learning from the approach of skill acquisition theory, where automaticity is reached through practice, similar to learning skills such as riding a bicycle, playing tennis, or driving a car. Also relevant to L2 research, automaticity has been explored by Dąbrowska and Blake (2022) in adolescents learning English as an L2. This study measured the speed of automatization by analysing performance on the Multiple-trial Tower of Hanoi task (MToH). Their results demonstrate a relationship between automatization in the MToH

task, and 'decorative' grammar (aspects of grammar which have abstract and generally redundant meaning).

However, automaticity is not usually investigated in typical L1 acquisition. In research regarding neurodevelopmental disorders, an automaticity deficit has been explored in the context of dyslexia (Nicolson & Fawcett, 1990, 2007, 2019; Vakil et al., 2015), but to the best of our knowledge, it has not been explored in DLD populations.

Processing speed

Several studies have established that children and adolescents with DLD have slower response times than typical controls in both linguistic and non-linguistic tasks (Kail, 1994; Leonard, 2014; Leonard et al., 2007; Miller et al., 2001, 2006; Windsor & Hwang, 1999). According to Kail (1994), children with DLD show a generalised slowing deficit, where each component of a task is processed more slowly, resulting in an overall slower processing speed. Kail found that school-age children with DLD responded 33% more slowly overall in linguistic and non-linguistic tasks, than typically developing children (Kail, 1994). A replication study by Windsor & Hwang (1999), found that children were one fifth slower overall, in comparison to one third as reported by Kail (1994). Similarly, Miller et al. (2001) report slower response times in children with DLD (although not applicable to all children with DLD), and in adolescents (Miller et al., 2006). These robust findings have led to the suggestion that slow processing speed may be a reliable clinical marker for DLD (Park et al., 2015, 2020).

Research regarding processing speed in adults is limited. Miller & Poll (2009) conducted a small-scale study of college students with a history of language difficulties and

found that those with language difficulties scored lower in language assessment tasks and had slower response times. In a study investigating memory trace decay by Poll, Watkins and Miller (2014), adults with DLD were slower than language typical controls to detect target words in a sentence, but their performance improved when sentences were presented more slowly. Poll, Miller and van Hell (2016) explored sentence repetition accuracy and processing limitations in adults with and without DLD. The sentence repetition task included sentences of varying lengths, comprising of argument and adjuncts, reliant on judgements of grammaticality (for examples, see Poll et al., 2016, p.305). Overall, adults with slower processing speed had more difficulty with the long argument condition sentences which were more cognitively demanding, consistent with Kail's generalised slowing hypothesis (1994). A recent meta-analysis explored reaction times (RT) to ascertain whether slower processing in DLD is domain general (Zapparrata et al., 2023). Studies included in the analysis involved tasks such as simple RT (motor response to a stimulus), choice RT (evaluating simple stimuli through a forced-choice response), naming of verbal stimuli, or interference control (attentional and interference-based tasks). The overall results indicate longer mean RTs in the DLD group (g = .47). Longer RTs were evident across all tasks, stimulus types, modalities, and responses, suggesting that slower processing is indeed domain general and not simply language related.

The present study

The present study is novel in that it explores certain areas that are not typically investigated in DLD research, such as language aptitude/language analytic ability, and the speed of automatization. We also evaluate more well-explored cognitive predictors such as nonverbal IQ, pseudoword learning, and implicit statistical learning. In addition, we measure reaction times to explore whether there are differences in processing speeds in both linguistic and non-

linguistic tasks. We investigate how these processes predict vocabulary and grammar (measured by a grammaticality judgement and a receptive grammar task), in a group of adults with and without DLD.

Method

Participants

30 adults with DLD (20 female and 10 male; mean age: 38;77 years), and 30 adults with typically developing language (21 female and 9 male; mean age 38;50 years), took part in an online study. The adults with DLD were recruited via social media. Inclusion criteria specified: 1) that adults should be 18 years of age or older, 2) speak English as a first language and, 3) have no known diagnosis of autism spectrum disorder or a hearing impairment. As we are aware that many adults with DLD are undiagnosed, we stated that a formal diagnosis of DLD was not required. 17 participants in our DLD group were formally diagnosed with DLD. For the rest of adults who did not report a diagnosis but were included in the DLD group, their inclusion was motivated by their responses to a series of questions determining their eligibility. These questions are provided in Appendix 3. The control group was recruited via Prolific and matched for age, gender, and education. Participants received a voucher as a token of thanks for participating in our study.

Materials and procedures

All tasks were presented on the Gorilla Experiment Builder platform (https://gorilla.sc; Anwyl-Irvine et al., 2020. Participants completed three linguistic tasks (assessing grammar and vocabulary), and five cognitive tasks (to assess language aptitude, nonverbal IQ, pseudoword

learning, implicit statistical learning, and the speed of automatization). These measures are described below.

Background questionnaire

Participants completed a questionnaire which included questions regarding their age, gender, educational level, occupation, first language (and details of any additional languages spoken), whether or not they had a formal diagnosis of DLD and, if they did, any additional information regarding their diagnosis.

Cognitive assessments

Language aptitude (Sentence Pairs task)

To measure language aptitude, we use the Sentence Pairs task, which measures metalinguistic awareness through sensitivity to grammatical structure without using specific terminology such as: adjective, verb, noun. The task was developed by Llompart and Dąbrowska (2023) and is modelled on the Words in Sentences sub-test of the MLAT (Carroll & Sapon, 1959). Participants are presented with pairs of sentences and in the first sentence, one word is printed in capital letters. In the second sentence, five words are highlighted, and participants are asked to select the word in the second sentence that has the same (grammatical) function as the word in capital letters in the first sentence. There are four practice trials with feedback given, i.e., if an incorrect response is selected, a red cross is displayed (with an example of the correct sentence), and a green tick is displayed for a correct response. There are 32 test trials.

Nonverbal IQ (The Block Design task)

Nonverbal IQ was assessed using the Block Patterns subtest from the Shipley Institute of Living Scale (Shipley-2; Shipley, Gruber, Martin, & Klein, 2009), adapted for online use. In this task, participants are required to replicate patterns of black-and-white shapes. Each trial comprises a pair of images with two versions of the same design, but part of the second image has one or more squares shaded in grey. Participants are asked which pattern completes the image and they select their answer from a multiple-choice format displayed underneath the images. There are 12 questions with trials 5-12 involving several subparts such that there are 26 questions in total.

Pseudoword Learning task

This task was designed by Prela, Llompart, & Dąbrowska (2022), and is adapted from a novel word learning paradigm developed by Llompart & Reinisch (2017, 2020, 2021) to test word-picture associations (explicit, associative memory). Participants are presented with ten objects selected from a novel object database (Horst & Hout, 2016), and ten nonce words formed following English phonotactics, recorded by a native speaker of English. There is a training stage and two test phases. In the training stage, participants see four pictures on the screen and listen to an audio recording of the novel 'target' word. Their task is to select the item that corresponds to the word by clicking on one of the images. Each target word is presented together with three distractors, and feedback is provided after entering a response (a green tick for correct answers and a red cross for incorrect answers). The training phase consists of ten blocks, each of which consists of one presentation of each target word. After the training phase, there are two test phases. Test one involves the same format as the training phase but without feedback. In test two, participants listen to words and are asked whether the word they hear is one of the novel words heard previously. The words comprise of recordings of the words

they have learned in the training phase, presented together with foils. This is a forced-choice task and participants click on either ‘yes’ or ‘no’.

Visual Implicit Statistical Learning task (VSL)

The visual implicit statistical learning task is modelled on an original design by Siegelman, Bogaerts, and Frost (2017), which in turn uses shapes from Fiser and Aslin (2001). The task involves a familiarisation phase and test phase. Participants see 16 shapes (labelled 1 to 16, for the sake of exposition), which form eight triplets. The triplets differ in their transitional probabilities (TPs) as follows:

1. Four triplets with TPs of .33 (shapes 1-2-3, 2-1-4, 4-3-1, and 3-4-2).
2. Four triplets with TPs of 1 (shapes 5-6-7, 8-9-10, 11-12-13, and 14-15-16).

During the familiarisation phase, eight triplets are displayed one-by-one, in random order. Each triplet appears 24 times and the order is controlled so that the same triplet does not appear twice in a row. Participants are advised to pay close attention to the shapes that appear so that they can answer questions relating to the shapes later in the task. A cover task involves pressing the spacebar (this occurs every 16 trials) to ensure that participants maintain attention.

After the familiarisation phase, participants are asked whether they have seen:

1. Some fixed sequence of two shapes
2. Some fixed sequence of three shapes
3. Some fixed sequences of four shapes
4. Some fixed sequences of six shapes

5. No fixed sequences of shapes
6. If you noticed a different sequence, please describe this below.

The test phase involves two alternative-forced-choice trials in which participants are asked to choose between two triplets. One triplet was presented during the familiarisation stage (low TPs =.33 or TP = 1), and a foil triplet that had not appeared previously (TP = 0). There are two types of test items, pattern recognition trials (see Figures 11 and 12), and pattern completion trials (see Figure 13). Participants are asked to choose the pattern that appears most familiar to them or they are asked to complete a missing shape in a pattern. The 42 test items comprise of either two, three, or four forced-choice questions.

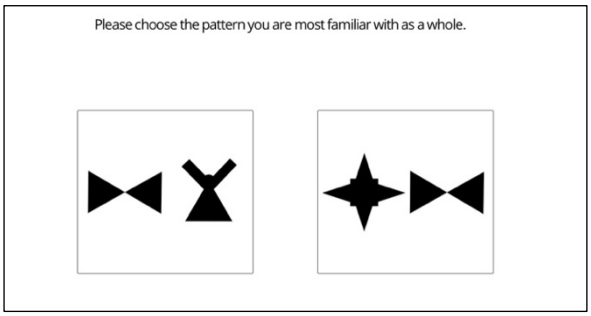


Figure 9. Example of a pattern recognition trial in the VSL task (pair recognition)

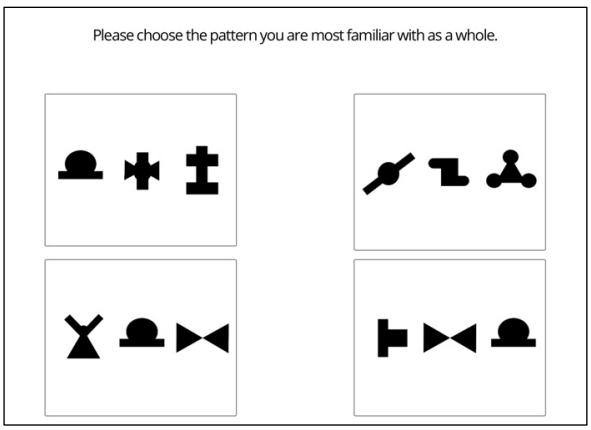


Figure 10. Example of a pattern recognition trial in the VSL task (triplet recognition)

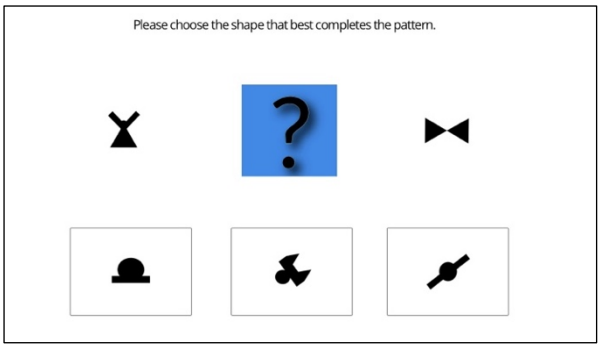


Figure 11. Example of a pattern completion trial in the VSL task

Speed of automatization (The Multiple-trial Tower of Hanoi task: MToH)

The Tower of Hanoi puzzle consists of three vertical rods set on a horizontal base and several discs ranging in size. We use the Multiple-trial Tower of Hanoi task (Dąbrowska & Blake, 2022). This is a four-disc version of the puzzle and is adapted for online use with the support of Gorilla Scripting Consultancy (see Figure 14). At the beginning of the task, the four discs are stacked on the leftmost rod and participants are instructed to move the discs, one at a time, to the rightmost rod, until all four discs are stacked in ascending order. To help participants get started, they are advised of the first move: to move the smallest disc to the middle rod. Two short videos demonstrate legal and illegal moves (participants can only move one disc at a time, and they cannot place a larger disc on top of a smaller disc). Participants are instructed that they will solve the puzzle 25 times and while doing so, they will advance through various levels. The master level (trials 21-25) involves a secondary task where participants are asked to do two tasks simultaneously, complete the puzzle and respond to shapes that appear on the screen. There are three practice trials for the secondary task with feedback provided. The number of moves is recorded for each trial.

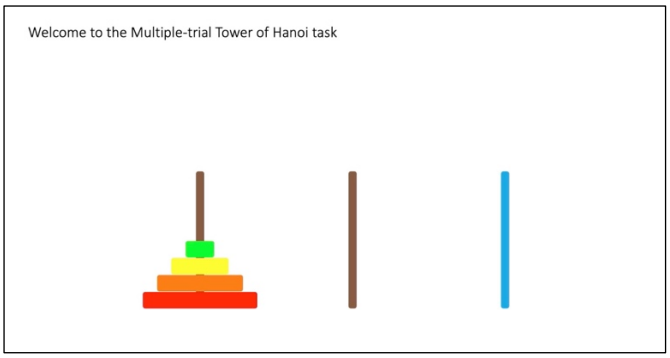


Figure 12. Multiple-trial Tower of Hanoi task

Language assessments

Receptive grammar (Pictures and Sentences task)

This is an online version of a task developed by Dąbrowska (2018). Participants read a sentence on the screen which corresponds to two pictures. Their task is to select the picture that matches the sentence that they have read. There are 80 items representing 10 different constructions (for an example, see Figure 15). The 10 constructions (listed in Table 15) are presented in a random order so that no two items of the same construction appear consecutively.

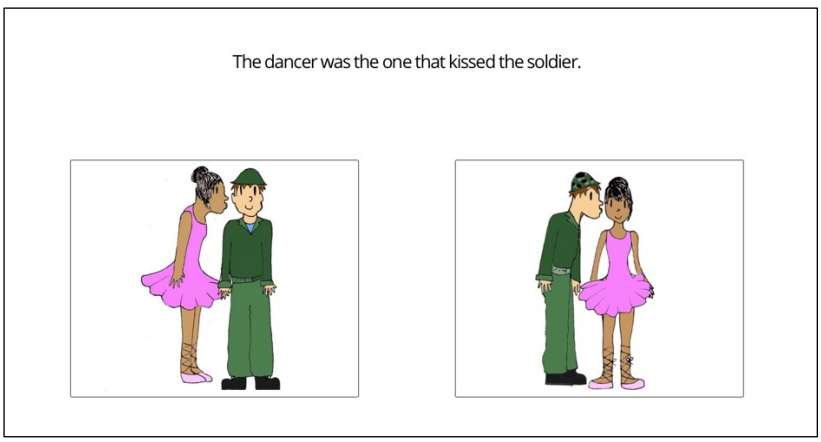


Figure 13. Example of a test item from the Pictures and Sentences task

Table 15. Examples of constructions in the Pictures and Sentences task

Construction	*Example*
Passive	The dancer was scratched by the boy.
Subject cleft	It was the boy that scratched the dancer.
Object cleft	It was the dancer that the boy scratched.
Subject relative	The boy was the one who scratched the dancer.
Object relative	The dancer was the one that the boy scratched.
Quantifier + is	Every lamp is on a table.
Quantifier + has	Every table has a lamp on it.
Postmodifying prepositional phrase	The lamp on the table is white.
Tough movement	The doll is hard to see.
Complex noun phrase	The wheelbarrow on the path with the brick was broken.

Grammaticality Judgment task (GJT)

The items in the grammaticality judgment task are based on stimuli used by DeKeyser (2000) who utilised stimuli developed by Johnson and Newport (1989). Our task includes 80 of the 200 items used by DeKeyser (2000), so that each of 8 categories were represented, with ten trials of each. (see Table 16). Half the stimuli in the task are grammatical and the other half ungrammatical. Participants are presented with a sentence and are asked to click on 'yes' if they think the sentence is grammatical, and 'no' if ungrammatical.

Table 16. Examples of constructions in the Grammaticality Judgment task

Structure	*Examples of grammatical and ungrammatical sentences (ungrammatical items are in italics)*
Agreement attraction	The steps listed on the website make applying sound very easy.
	The legs of the dining-hall table needs to be repaired.
Participle clause	The alarm went off while he was sitting in the waiting room.
	Loving the view, the sea was so calm and colourful
Subcategorization	Seven judges of the British supreme court handle corruption cases.

	The old lady finally decided selling the house to her neighbours.
Word order	Sonia decided to ask Pete what his favourite meat dish was.
	Josh was very concerned about what would be the consequences of his brother's actions.
Stranded Wh-Questions	What did the fugitive hope that the police would do?
	What does the lady reckon what her husband bought?
Double tense	Did the owner actually work in the supermarket?
	Did James forgot that you were waiting for him?
Object agreement	Which film are the critics praising?
	Which city has the ambassadors visited this week?
That trace	What do you think is dangerous?
	What does she feel that put her in a bad position with her employer?

Vocabulary (Shipley's vocabulary task)

We use an implementation of the vocabulary component of the Shipley Institute of Living Scale (Shipley-2: Shipley et al., 2009). This is a receptive vocabulary test comprising 40 words with multiple-choice answers. In each trial, the participant is presented with a word on the screen in uppercase letters (e.g., HAPPY) with four multiple-choice options in lowercase. Each response is numbered (e.g., 1. sombre, 2. joyful, 3. kind, 4. pensive). Participants are asked to select the correct answer by clicking on the item that is most similar to the word presented in uppercase.

Results

Data pre-processing

All data were analysed with R Statistical Software (ver. 4.2.1; R Core Team, 2022). In the DLD group, we excluded two participants for the following reasons: one participant did not meet our inclusion criteria for language difficulties, and one participant could not commit to all three research sessions. Because of the online format of the study, we inspected reaction

time (RT) data to check that participants were engaged with each of the tasks. This resulted in specific exclusions. In the DLD group, one participant's data was excluded from the Sentence Pairs task (due to low consecutive RTs of less than 1800ms in over half the trials and overall accuracy of 9%, suggesting that they were not focused on the task), one participant in the pseudoword learning task who did not engage with training trials, and one participant in the visual implicit statistical learning task (with very low RTs of less than 950ms in 20 consecutive trials). In the LT group, we excluded data from one participant in the pseudoword learning task for lack of engagement in the training trials. Missing values were imputed using the `mice` package (ver 3.14.0; van Buuren & Groothuis-Oudshoorn, 2011). After specific exclusions (as above), we examined outliers in RT data and removed outliers +/- 3 standard Median Absolute Deviations (Leys, Ley, Klein, Bernard, & Licata, 2013), above and below the median RT (per participant, on accurate trials only). Information regarding the percentage of outliers removed (per task), is reported in Appendix 4.

Descriptive statistics

As the MToH is a new task designed in our lab, we use two exploratory measures to explore automatization. The first measure is the difference score (AutoDiff), which is calculated by subtracting the average number of moves in the trials in which there was a secondary task (trials 21-25), from the average number of moves in the last five standard trials (trials 16-20). Whilst we expect interference across the board, if participants have automatized the solution for solving the puzzle, then we would expect to see less interference in the secondary task. Therefore, negative scores indicate interference effects. The second measure, the coefficient of variation (AutoCV), is calculated by dividing the standard deviation of the number of moves by the mean number of moves, and is based on previous work by Segalowitz

and colleagues (Hulstijn et al., 2009; Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998). We expect to find less variability in automatized procedures and a decrease in the coefficient of variation over repeated trials.

For the pseudoword learning task we report composite measures for the training phase, test one and two. For all other tasks, we report accuracy as proportion scores, by dividing the number of correct responses by the total number of test items.

Table 17 reports the descriptive statistics, using the `psych` package (version 2.3.9; Revelle, 2023). Reliability estimates (Spearman-Brown corrected) were computed using the `splithalf` package (version 0.8.2; Parsons, 2021), utilizing 5000 random splits. All reliability estimates are reported in Table 17, with exception of those pertaining to the MToH task, which are detailed below.

For the MToH task, we performed an odd/even split-half reliability analysis. This accounts for the stability of participants' scores which tend to become consistent from approximately the mid-point of the overall task (although it should be pointed out that scores do not always stabilise for every individual). The data was divided into two subsets: one for odd and one for even trials. The `splithalf` function was applied to each subset. In the LT group, the reliability estimates were as follows: 0.75, 95% CI [0.64, 0.84] for odd trials, and 0.87, 95% CI [0.78, 0.93] for even trials. In the DLD group, the reliability estimates were 0.83, 95% CI [0.68, 0.93] for the odd trials and 0.81, 95% CI [0.64, 0.92] for the even trials.

Examining accuracy scores across language assessments, the GJT task shows variability in performance, although the LT group generally has higher scores. In the pictures

and sentences task (receptive grammar), while some participants in both groups reached ceiling, there are a broader range of scores in the DLD group, suggesting greater variability in performance on this task. For vocabulary, the LT group demonstrated higher accuracy overall with ceiling effects evident in some participants. However, there is more variability in the DLD group, indicating differences in vocabulary knowledge.

The range of scores in cognitive measures is similar between groups with the exception of the sentence pairs task, where there is more variability in the DLD group. Notably, there is overlap in performance on this task, with some individuals with DLD performing similarly to the LT group. However, there were no significant differences between groups in all other cognitive measures, including nonverbal IQ, implicit statistical learning, pseudoword learning, and speed of automatization.

Table 18 provides group comparisons which were conducted using Welch two-sample *t*-tests (Welch, 1947). T-tests are provided to demonstrate differences between the two groups, but they are not corrected for multiple comparisons as more in-depth statistical analysis follows. Cronbach's coefficient alpha was computed using the `effsize` package (version 0.8.1.; Torchiano, 2016). As can be seen in Table 18, there are differences between groups in all linguistic tasks, but these are smaller for the pictures and sentences task. With regards to our assessments of cognitive abilities, there are group differences in language aptitude/language analytic ability (sentence pairs) but no notable differences in the other cognitive measures. There are significant differences between groups for response times, with the DLD group showing slower performance across linguistic and cognitive tasks.

Table 17. Mean, standard deviation (SD), range, reliability

Variables	Mean: DLD (SD)	Range: DLD	Reliability: DLD [CI]	Mean: LT (SD)	Range: LT	Reliability: LT [CI]
Language measures						
GJT accuracy	0.64 (0.10)	0.45 – 0.85	0.72 [0.56, 0.85]	0.73 (0.10)	0.53 – 0.90	0.78 [0.66, 0.88]
GJT RT	7935.11 (4958.69)	2344 - 24383	-	4787.84 (1866.98)	2666 - 10569	-
PST accuracy	0.89 (0.08)	0.65 – 1.00	0.85 [0.75, 0.93]	0.93 (0.06)	0.78 – 1.00	0.82 [0.61, 0.90]
PST RT	6247.64 (2510.93)	2982 - 12192	-	3552.54 (749.62)	2347 - 5718	-
Vocabulary accuracy	0.73 (0.14)	0.42 – 0.90	0.76 [0.62, 0.87]	0.82 (0.10)	0.65 – 1.00	0.64 [0.41, 0.81]
Vocabulary RT	4676.02 (2422.47)	2117 - 13988	-	2901.25 (1505.92)	1887 - 10075	-
Cognitive measures						
Sentence Pairs accuracy	0.49 (0.21)	0.15 – 0.81	0.86 [0.78, 0.92]	0.62 (0.18)	25.00 – 94.00	0.79 [0.67, 0.88]
Sentence Pairs RT	14364.32 (8001.69)	2136 - 35994	-	7787.22 (4160.10)	2278 - 19489	-
Nonverbal IQ accuracy	0.51 (0.21)	0.07 – 0.96	0.81 [0.72, 0.91]	0.56 (0.22)	0.11 – 0.96	0.85 [0.76, 0.92]
Nonverbal IQ RT	8951.92 (4809.09)	3121 - 24628	-	6156.68 (2420.93)	2221 - 12103	-
ISL accuracy	0.58 (0.13)	0.30 – 0.81	0.70 [0.53, 0.84]	0.56 (0.19)	0.28 – 95.00	0.86 [0.78, 0.93]
ISL RT	5546.90 (1673.81)	2854 - 9876	-	4504.67 (2639.02)	752 - 12820	-
Pseudoword accuracy	0.82 (0.08)	0.65 – 0.94	0.92 [0.86, 0.96]	0.86 (0.07)	0.68 – 0.94	0.95 [0.93, 0.98]
Pseudoword RT	4442.39 (956.11)	2846 - 6627	-	3640.88 (665.03)	2906 - 5685	-
AutoCV	0.18 (0.22)	0.00 – 0.71	-	0.17 (0.19)	0.00 – 0.68	-
AutoDiff	-0.43 (6.05)	-15.40 – 19.00	-	-2.07 (4.65)	-11.8 – 5.60	-

Language measures: GJT: Grammaticality judgement task; PST: Pictures and Sentences task; Vocabulary: Shipley's vocabulary task. Cognitive measures: Sentence pairs (language aptitude); Nonverbal IQ: Shipley's block design, ISL; implicit statistical learning - visual statistical learning task; pseudoword learning: pseudoword learning task; AutoCV: coefficient of variation (standard deviation divided by the mean number of moves on MToH task); AutoDiff: difference in moves between the last 5 standard trials and the secondary task of the MToH. All confidence intervals (CI) are at the 95% level.

Table 18. Group comparisons

Variables	*t*-test	Cohen's *d*	*p*-value
Language measures			
GJT accuracy	$t(57) = 3.4$	0.87	**
GJT RT	$t(35) = -3.3$	-0.84	**
PST accuracy	$t(54) = 1.9$	0.49	***
PST RT	$t(34) = -5.6$	-1.45	***
Vocabulary accuracy	$t(52) = 2.6$	0.67	*
Vocabulary RT	$t(48) = -3.4$	-0.88	**
Cognitive measures			
Sentence Pairs accuracy	$t(55) = 2.5$	0.64	*
Sentence Pairs RT	$t(43) = -4.0$	-1.03	***
Nonverbal IQ accuracy	$t(57) = 0.83$	0.22	0.4
Nonverbal IQ RT	$t(42) = -2.84$	-0.73	**
ISL accuracy	$t(51) = -0.43$	-0.11	0.7
ISL RT	$t(49) = -1.82$	-0.47	0.07
Pseudoword learning accuracy	$t(56) = 1.92$	0.50	0.06
Pseudoword RT	$t(52) = -3.76$	-0.97	***
AutoCV	$t(56) = -0.08$	-0.02	0.9
AutoDiff	$t(54) = -1.17$	0.30	0.2

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Correlations

This section provides a summary of key findings from the correlation analyses by participant group (Pearson correlation matrices can be found in the appendices: see Appendix 1 for the LT group and Appendix 2 for the DLD group).

Firstly, there are strong relationships between our measure of language aptitude/language analytic ability (sentence pairs), and all language variables in both groups. Interestingly, language aptitude and IQ have a strong correlation in the control group: (LT*: r* =

.73) but not in the DLD group (DLD: $r = .20$). A Fisher's Z-test revealed a significant difference between these correlations ($Z = 2.667$, $p < .001$). Nonverbal IQ correlates with all language measures in both groups (although weakly with vocabulary in the DLD group). In the LT group, ISL correlates with all language measures, yet in the DLD group, it is significant for pictures and sentences ($r = .44$) and vocabulary ($r = .43$) only. There are no relationships between pseudoword learning and language measures in the LT group, yet in the DLD group we find significant correlations with all language measures.

Lastly, our two measures of automatization provide mixed results. We find negative correlations between AutoCV and grammar (for the pictures and sentences task), and vocabulary in the LT group, and in the DLD group, a negative relationship between AutoCV and vocabulary. AutoDiff on the other hand, correlates with pictures and sentences in both groups, although the relationship is positive where we would typically expect to find a negative correlation.

Statistical models

Regression analyses were conducted for the whole group (n = 60) and then for each group separately (DLD and LT). We examined the effects of five cognitive predictors (language aptitude, nonverbal intelligence, implicit statistical learning, pseudoword learning and automatization), on three linguistic measures (grammaticality judgement task, pictures and sentences task and vocabulary). For whole group analyses, we included interactions between group (i.e., LT or DLD), and our predictors.

Following the method recommended by Crawley (2013), we fitted a maximum model first. We then inspected the model summary, removing the highest-order, non-significant interaction first. If an interaction was significant, then we retained the main effects too. Model comparisons were conducted, using ANOVA and if there were no statistical difference between models, then the newer model was retained. This method was continued, removing non-significant highest order terms until we obtained the best model fit. The 'relative importance' of predictors was assessed using the 'lmg' metric from the `relaimpo` package (version 2.2-6; Grömping, 2006). By using the calc.relimp function, we provide the proportion of variance of each regressor's influence to the model.

Model assumptions were checked using the `performance` package (version 0.10.1; Lüdecke et al., 2021). Heteroscedasticity was detected in the model for the pictures and sentences task. We used the Eicker-Huber-White method to adjust the values for standard errors, *t*-values and *p*-values (White, 1980). A summary of the best-fitting models obtained after model comparisons is provided in Table 19 (for the whole group), Table 20 (for the LT group), and Table 21 (for the DLD group).

The regression models for the combined LT and DLD groups (Table 19) demonstrate that the two most significant predictors for grammar and vocabulary are group (LT or DLD), language aptitude/language analytic ability (the sentence pairs task), and pseudoword learning. Implicit statistical learning and AutoDiff predict grammatical comprehension (the pictures and sentences task), and AutoCV is relevant to vocabulary. Following these results, we conducted further regression analyses for each group, and these results are discussed accordingly.

For the LT group (Table 20), performance on the sentence pairs task predicts grammar and vocabulary. As can be seen, the model for the grammaticality judgment task is predicted by sentence pairs, accounting for 17% of variance. Similarly, for receptive grammar (the pictures and sentences task), sentence pairs task explains 25% of variance in the model, and AutoCV 14%, although the latter is not statistically significant. Performance on the sentence pairs task accounts for 17% of the variance in vocabulary scores.

In the DLD group (Table 21), a slightly different pattern emerges with both the sentence pairs task and pseudoword learning predicting linguistic outcomes. For example, the model for grammaticality judgment shows that 40% of overall variance is accounted for by both sentence pairs (26%), and pseudoword learning (14%). Three predictors account for 53% variance on the pictures and sentences task, these are pseudoword learning (30%), sentence pairs (12%), and AutoD (11%). Lastly, pseudoword learning accounts for 18% of variance in vocabulary scores for the DLD group.

Table 19. Regression models for the whole sample (LT and DLD groups combined)

Grammaticality Judgment

Variable	Estimate (β)	std.error	t value	*p* value	95% CI	lmg
(Intercept)	0.72	0.02	42.31	***	0.69, 0.76	
Language group	-0.06	0.02	-2.36	*	-0.09, 0.00	0.11
Pseudoword learning	-0.01	0.02	-0.73	0.46	-0.05, 0.02	0.07
Sentence pairs	0.04	0.01	2.99	**	0.02, 0.07	0.16
Group x pseudoword learning	0.05	0.02	2.24	*	0.02, 0.12	0.06
Model R^2						**0.40**

Adjusted $R^2 = 0.35$

$F(4,55) = 9.05$, $p = < 0.001$

****** p < 0.001; ** p < 0.01; * p < 0.05***

Observations: 60

Pictures and Sentences

Variable	Estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.93	0.00	97.18	***	0.91, 0.95	
Language group	-0.03	0.02	-1.71	0.09	-0.05, 0.00	0.04
ISL	0.02	0.01	2.38	0.05	-0.00, 0.03	0.07
Pseudoword learning	-0.02	0.01	-1.83	0.07	-0.03, 0.00	0.10
Sentence Pairs	0.02	0.01	2.79	*	0.00, 0.03	0.14
AutoDiff	0.02	0.01	2.67	**	0.00, 0.03	0.08
Group x pseudoword learning	0.05	0.01	3.91	**	0.02, 0.07	0.10
Model R^2						**0.53**

Adjusted $R^2 = 0.48$

$F(6,53) = 10.1$, $p = < 0.001$

**** p < 0.001; ** p < 0.01; * p < 0.05*

Observations: 60

★ denotes Eicker-Huber-White correction for heteroscedasticity.

Vocabulary

Variable	Estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.81	0.02	39.16	***	0.76, 0.84	
Language group	-0.05	0.03	-1.70	0.09	-0.11, 0.00	0.07
Pseudoword learning	-0.01	0.02	-0.36	0.72	-0.05, 0.03	0.07
Sentence Pairs	0.04	0.02	2.26	*	0.00, 0.06	0.12
AutoCV	-0.03	0.01	-1.73	0.09	-0.05, 0.00	0.05
Group x pseudoword learning	0.05	0.03	1.65	0.65	-0.01, 0.10	0.03
Model R^2						**0.34**

Adjusted $R^2 = 0.28$

$F(5,54) = 5.61$, $p = < 0.001$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 60

Table 20. Regression models for the LT group

Grammaticality Judgment

Variable	Estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.73	0.02	42.03	***	0.69, 0.76	
Sentence Pairs	0.04	0.02	2.41	*	0.00, 0.07	
Model R^2						**0.17**

Adjusted $R^2 = 0.14$

$F(1,28) = 5.81$, $p = 0.023$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Pictures and Sentences

Variable	Estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.93	0.01	99.30	***	0.91, 0.94	
Sentence Pairs	0.03	0.01	2.83	**	0.00, 0.05	0.25
AutoCV	-0.02	0.01	-1.77	0.08	-0.03, 0.00	0.14
Model R^2						**0.39**

Adjusted $R^2 = 0.34$

$F(2,27) = 8.73$, $p = 0.001$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Vocabulary

Variable	Estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.82	0.02	48.71	***	0.78, 0.85	
Sentence Pairs	0.04	0.02	2.41	*	0.00, 0.07	
Model R^2						**0.17**

Adjusted $R^2 = 0.14$

$F(5,79) = 5.79,\ p = 0.023$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Table 21. Regression models for the DLD group

Grammaticality Judgment

Variable	estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.64	0.01	44.35	***	0.61, 0.67	
Pseudoword learning	0.05	0.02	3.05	**	0.01, 0.07	0.26
Sentence Pairs	0.03	0.02	1.95	0.06	-0.00, 0.06	0.14
Model R^2						**0.40**

Adjusted $R^2 = 0.35$

$F(2,27) = 8.93,\ p = 0.001$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Pictures and Sentences

Variable	estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.89	0.01	80.34	***	0.87, 0.91	
Pseudoword learning	0.04	0.01	3.61	**	0.01, 0.06	0.30
Sentence Pairs	0.02	0.01	1.91	0.06	-0.00, 0.04	0.12
AutoDiff	0.03	0.01	2.23	*	0.00, 0.04	0.11
Model R^2						**0.53**

Adjusted $R^2 = 0.48$

$F(3,26) = 9.86,\ p = < 0.001$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Vocabulary

Variable	estimate (β)	std.error	*t* value	*p* value	95% CI	lmg
(Intercept)	0.74	0.02	32.00	***	0.68, 0.78	
Pseudoword learning	0.06	0.02	2.48	*	0.00, 0.10	
Model R^2						**0.18**

Adjusted $R^2 = 0.15$

$F(1,28) = 6.07$, $p = 0.020$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Observations: 30

Discussion

The primary aim of this study was to investigate the relationship between cognition and language in adults with DLD. Our study is novel in that we explored areas not typically examined in DLD, such as language aptitude/language analytic ability and speed of automatization. We also investigated cognitive processes known to be relevant to linguistic outcomes in DLD, for example, nonverbal IQ, pseudoword learning, and implicit statistical learning. Reaction times were measured across all tasks. Our findings are discussed accordingly.

Firstly, our results show differences between groups in linguistic measures, indicating that language remains a primary area of difficulty in adults with DLD. This outcome is consistent with a wide body of research which has established that language and literacy skills remain an area of difficulty in adulthood (Botting, 2020; Clegg et al., 2005; Conti-Ramsden et al., 2018; Del Tufo & Earle, 2020; Earle et al., 2018; Law et al., 2009; Whitehouse et al., 2009).

Regression analyses show that our measure of language aptitude (the sentence pairs task) predicts grammaticality judgement and pictures and sentences in both groups. Therefore,

grammar depends on language aptitude in typically developing language and in DLD. These results are interesting from two perspectives. Firstly, as discussed in in the introduction, language aptitude is not usually explored in the context of L1 grammar acquisition, as grammar is considered to be dependent upon implicit learning and not reliant on explicit processes. However, recent studies report a significant relationship between language aptitude and L1 grammatical proficiency (Dąbrowska, 2018; Llompart & Dąbrowska, 2023; Prela et al., 2022; Winckel & Dąbrowska, 2024), and our results are consistent with these findings.

Secondly, to the best of our knowledge, ours is the first study to show that language aptitude plays a role in DLD. The measure of language aptitude used in our study, the sentence pairs task, measures grammatical sensitivity and relies on explicit processes, i.e., inductive reasoning and metalinguistic awareness. This evidence strongly suggests that explicit reasoning abilities are relevant in L1 acquisition, both in LT populations and also in adults with DLD. Proponents of the procedural/declarative model have argued that adults with DLD use explicit processes (declarative memory), to compensate for deficits in procedural learning which underlie grammar (Ullman et al., 2020; Ullman & Pierpont, 2005; Ullman & Pullman, 2015). Yet our results show that language aptitude is also relevant to grammar in the LT group.

Nonverbal IQ is not a significant predictor in regression models in either the LT or DLD group. Where children with DLD often score below their peers in nonverbal skills (Gallinat & Spaulding, 2014), this may not be the case in adulthood. Our results align with previous longitudinal studies which report similar findings (Botting, 2020; Clegg et al., 2005). Whilst these outcomes are positive in supporting a potential ‘catch-up’, Botting (2020) tentatively suggests that nonverbal IQ scores could in fact fluctuate over time and points to the fact that they may not follow a stable trajectory in adults with DLD. Another reason for

changing scores in these two studies (and acknowledged by the authors), could be that different IQ tests were used at different time points (see discussion points raised by Lancaster et al., 2024).

We find that nonverbal IQ and language aptitude are strongly correlated in the LT group (r = .73), but not in the DLD group (r = .20). The absence of a relationship in the DLD group suggests that IQ doesn't "translate" into grammatical sensitivity in adults with DLD and could be due to a bottleneck in phonological encoding. Encoding verbal information is an area of difficulty in children with DLD (Jackson et al., 2021). This continues into adulthood – a series of studies show that college students with DLD require more exposure than LT adults to encode word forms into long-term memory, yet retention is a strength. The authors conclude that the encoding stage (referring to the initial memory trace linking a word form and referent), is a bottleneck for learning new words (McGregor et al., 2013, 2017a, 2017b, 2020), and our findings align with this theory. Additionally, our results from the pseudoword learning task are interesting. Both the LT and DLD groups show similar accuracy. Yet group differences are much more pronounced for reaction time than for accuracy. This suggests that poor processing speed contributes to the challenge of learning new words in DLD. Additionally, regression analyses show that pseudoword learning is a significant predictor of both grammar and vocabulary in the DLD group, yet it is not significant for the LT group. Our results are in line with Del Tufo and Earle (2020) who report that college students with DLD have deficits in nonword repetition (reliant on phonological processing). However, to date there are very few studies exploring adult populations with DLD.

We used two exploratory measures to explore automatization, AutoCV and AutoDiff. Regression analyses showed that AutoCV predicts pictures and sentences in the LT group

whereas AutoDiff is significant for pictures and sentences in the DLD group. Automatization does appear to be relevant for language in our sample, but our measures may not be sensitive enough to pick up differences in adult populations, and so we interpret our findings with caution. The other concerning factor is that our two measures of automatization do not correlate with each other so it is possible that they measure different things. Upon closer examination of our results, many participants did not automatize a solution to the MToH, and we believe that this could be due to the online format of the task. To automatize a solution to the puzzle, participants repeat the puzzle over numerous trials. Finding a solution to the puzzle is difficult in the earlier stages of the task and it is possible that if participants do not solve the puzzle quickly, some may have become frustrated and demotivated by the online format. The speed of automatization is relevant to children with DLD (Blake, Dąbrowska, & Riches, in preparation), and to L2 learning in adolescents (Dąbrowska & Blake, 2022). However, both these studies used a physical version of the puzzle, and it could be argued that a tangible version of the task is not only more motivating but could draw on different cognitive processes (for discussion regarding the differences in modalities, see Blake & Dąbrowska, 2024).

Implicit statistical learning correlates with grammar and vocabulary in both groups. The relationship with vocabulary is unexpected, as lexical knowledge is assumed to rely primarily on explicit learning mechanisms. However, our ISL task also correlates strongly with both IQ and aptitude. Therefore, we suspect participants used explicit learning mechanisms in the familiarisation stage and it could be questioned whether the VSL task is a pure measure of implicit learning. It has been suggested that reflection-based measures could add 'noise' to research findings (Christiansen, 2019; Isbilen & Christiansen, 2022), and, as discussed earlier, the reliability of ISL tasks is a current topic of investigation in the literature (Arciuli, 2017;

Arciuli & Conway, 2018; Arnon, 2020; Bogaerts et al., 2021; Isbilen & Christiansen, 2022; Siegelman & Frost, 2015).

A significant finding to emerge from our study is that the DLD group have slower response times, evident in linguistic and non-linguistic tasks. This is in line with previous studies which report slower processing speeds in adults with DLD (Miller & Poll, 2009; Poll et al., 2014, 2016). These results are also consistent with data from a recent meta-analysis which reports longer RTs in people with DLD on simple processing tasks in both verbal and non-verbal contexts (Zapparrata et al., 2023). Therefore, slow processing is not limited to language and is evident across all domains, which is what we observe in our results. Furthermore, in our study, group differences for reaction times are larger than for accuracy which could indicate that poor accuracy is also due to slower processing in DLD. This partly fits with the findings of Del Tufo and Earle (2020), who assessed reading comprehension and fluency of college students with DLD and found that DLD participants did not have major comprehension difficulties per se, but they struggled when processing content under time-limited conditions.

There is limited research regarding adults with DLD, and this study offers new insights regarding the relationship between cognition and language in people with DLD, in later life. While the sample size in this study is small, this is a common challenge in DLD research, particularly among adults. However, consideration of our research findings is warranted in light of the number of predictors investigated, and further research with larger sample sizes is suggested to further validate and expand our findings.

Conclusion

Whilst the study described here is small-scale, there are several conclusions that warrant further investigation. The relevance of explicit reasoning abilities in L1 learning is important to theoretical knowledge regarding DLD in adulthood, and also from a clinical perspective, in developing intervention support for adults. Explicit processes are usually assumed to be relevant for lexical knowledge in L1 acquisition, but our data indicates that they are relevant for grammar too. Lastly, our results show that slow processing in adults with DLD is evident in linguistic and non-linguistic domains and this needs to be taken into consideration, not only in therapy, but also in wider day-to-day life.

Acknowledgements: This project was funded by an Alexander von Humboldt Professorship (grant number: ID 1195918) awarded to the second author. We extend our thanks to all participants for their time in taking part in our study.

4.5 Study three: Blake, Wright, Dąbrowska, Riches, and Winckel (in preparation)

This section presents a manuscript in preparation for publication entitled:

Non-verbal cognition as a predictor of linguistic outcomes in children with and without developmental language disorder

This paper refers to study three of this thesis and investigates cognitive abilities and language in children with and without DLD.

Building upon the results of study one (b), this study investigates the relationship between cognition and language in an older cohort of children (mean age 10;8 years). This study was pre-registered (https://osf.io/uc2qa) and focused on specific cognitive domains: the role of language aptitude/language analytic ability (using a grammatical sensitivity task), analogical reasoning, inhibition, and implicit statistical learning. These cognitive predictors were examined in relation to linguistic tasks assessing grammar, vocabulary and collocations. A total of 39 LT children and 34 children with DLD (mean age 10; 8 years), participated in an online, moderated study.

NON-VERBAL COGNITION AS A PREDICTOR OF LINGUISTIC OUTCOMES IN CHILDREN WITH AND WITHOUT DEVELOPMENTAL LANGUAGE DISORDER

Ashley Blake[1], Richenda Wright[2], Ewa Dąbrowska[2], Nick Riches[3], Elodie Winckel[2]

[1]University of Birmingham, [2]Friedrich-Alexander-Universität Erlangen-Nürnberg

[3]Newcastle University

Abstract

This study examined the relationship between cognitive abilities and linguistic outcomes in children with and without Developmental Language Disorder (DLD). We assessed performance in several cognitive abilities: analogical reasoning, language aptitude (using a grammatical sensitivity task), implicit statistical learning, and inhibition, and the extent to which performance relate to linguistic outcomes. 39 language typical (LT) children and 34 children with DLD (M = 10;8 years), took part in a moderated online study. Our results show between group differences in all linguistic outcomes. Regarding cognitive measures, there are large differences between groups in analogical reasoning and language aptitude (measured by a grammatical sensitivity task). There were no differences between groups in inhibition or implicit statistical learning, although we observe slower processing speeds in the DLD group. Analogical reasoning was the most significant predictor of receptive grammar in children with DLD, and grammatical sensitivity predicted all linguistic outcomes in the LT group. The grammatical sensitivity task used in this study is metalinguistic and is essentially a grammatical analogies test. Therefore, we propose that LT children may tap into language-specific analogical reasoning skills in the form of grammatical sensitivity to process all domains of language, whereas children with DLD rely on domain-general nonverbal analogical reasoning skills for receptive grammar.

Keywords: developmental language disorder, analogical reasoning, language aptitude, grammatical sensitivity, implicit statistical learning.

Introduction

Background

Developmental Language Disorder (DLD) is a persistent language disorder (Bishop et al., 2017; Leonard, 2014a), which occurs in approximately 7.5% of the population (Norbury et al., 2016; Tomblin et al., 1997). Although language is the primary area of difficulty, many children with DLD also have problems with domain-general cognitive processes including areas such as working memory (for reviews, see: Archibald, 2017; Henry & Botting, 2017; Montgomery et al., 2010), executive function (Pauls & Archibald, 2016), processing speed (meta-analysis: Zapparrata et al., 2023), temporal processing (Tallal & Piercy, 1973b, 1973a, 1974, 1975), and implicit statistical learning/procedural learning[13] (see meta-analyses: Lammertink et al., 2017; Obeid et al., 2016; Lum et al., 2014; Oliveira et al., 2023; West et al., 2021). In addition, a meta-analysis by Gallinat and Spaulding (2014) shows that children with DLD have lower scores in nonverbal IQ tests than their language typical (LT) peers (although see Lancaster et al., 2024 for possible reasons for this). This finding could indicate underlying deficits in nonverbal skills, or alternatively, a bidirectional relationship between language and cognition where weak nonverbal skills influence language, and vice versa (Griffiths et al., 2022).

Due to the heterogeneity observed in DLD, it has been suggested that DLD should be considered a spectrum disorder with different but overlapping strengths and weaknesses (Lancaster & Camarata, 2019). DLD is a complex, polygenic condition (Mountford et al., 2022), but despite extensive research, we do not have a comprehensive understanding of the

[13] Implicit statistical learning and procedural learning refer to different processes but are often assessed using the same tasks.

cognitive underpinnings that contribute to DLD. Therefore, the present study aims to explore this relationship further. We investigate several cognitive abilities, some of which are novel in the context of DLD (e.g., analogy and language aptitude - these will be discussed below), and first language (L1) acquisition. Firstly, we explore language from the perspective of usage-based approaches and in doing so, we discuss specific cognitive abilities that potentially contribute to linguistic outcomes in these populations.

According to usage-based theories of language acquisition, children's grammar develops through linguistic experience and depends upon domain-general cognitive capacities (Behrens, 2009; Bybee, 2010; Goldberg, 2006; Tomasello, 2003). These abilities include categorization, chunking, memory, analogy, and cross-modal association – the ability to link meaning with form (Bybee, 2010). The process of language acquisition is gradual and in the early stages of development, form-meaning pairings (constructions) are stored as chunks so that they can be retrieved as a single unit (Bybee, 2010). Language is intrinsically pattern based and children observe abstract patterns of similarities between utterances.

Analogy plays an important role in language acquisition and is a central component of this investigation. Through analogy, children generalize across utterances and map items of a similar meaning to create new abstract constructions of their own (Goldberg, 1995; Tomasello, 2003). For example, consider a busy morning routine where a child and their caregiver have daily discussions regarding breakfast and getting dressed. Caregivers may use phrases such as: *would you like some X? (cereal, milk, an apple),* or *let's brush X (your hair, teeth)*. These chunks are then generalized through slot-and-frame patterns to create new abstract utterances e.g., *gimme X (apple, milk),* or *brush (hair, teeth).* Or children may generalise across schematic representations to derive new schematic constructions, for example, a schema such as c*an I X?*

(get up, get down, throw it), could be slowly generalised, from *can you X?,* to *will you X?*, and finally, *will thing X?* (see examples in Dąbrowska, 2010b, p.698). Therefore, children's ability to generalise across patterns relies on analogy. Ultimately, the process of language development occurs in piecemeal fashion through repeated exposure until linguistic knowledge is entrenched.

The kinds of generalizations about which lexical items can fit into specific slots in a construction rely on analogy. It has been suggested that there is a reciprocal relationship between analogical reasoning ability and linguistic outcomes (Gentner & Christie, 2010). Analogical reasoning underlies generalization; therefore, if children have difficulty in establishing patterns across forms, then this may be a contributing factor to their language difficulties. Previous research provides evidence that children with DLD have analogical reasoning deficits across linguistic and non-linguistic tasks in comparison to their LT peers (Krzemien et al., 2017, 2019, 2020; Leroy et al., 2012, 2014; Nippold et al., 1988), although aside from these studies, this is an area that has not received much attention. Another factor for consideration is that there is a degree of circularity in that many nonverbal IQ tasks have an analogical reasoning component (e.g., Raven's Progressive Matrices, Raven, Raven, & Court, 1998), and children with DLD often score below their LT peers in nonverbal cognitive tasks, so this could be a contributing factor to potential deficits in analogical reasoning performance in children with DLD. For example, Nipoold et al. (1998) assessed analogical reasoning performance in children with DLD using three analogical reasoning tasks (a verbal task, perceptual reasoning task, and a functional analogical reasoning task). In all three tasks, children with DLD performed below their LT peers but when they controlled for nonverbal reasoning, group effects were no longer significant.

Another aspect that has been linked to performance in analogical reasoning tasks is the process of inhibition. Inhibitory control is a core component of executive function – being able to regulate attention enables individuals to focus on and attend to a specific task whilst being able to suppress irrelevant aspects (Diamond, 2013). Prior research has linked inhibitory control to analogical reasoning according to the principle that analogical reasoning involves adhering to similarities between structures as well as ignoring aspects that are not salient to meaning, i.e., potentially involving inhibition (Richland et al., 2006). If children with DLD have weaknesses in inhibition, as shown by previous research (Pauls & Archibald, 2016), then this could be a contributing factor to poor performance in analogical reasoning tasks (see Krzemien et al., 2020). On the other hand, this may not be the case in LT children. For example, Simms et al. (2018) examined executive functions related to analogical reasoning and found that working memory predicted analogical reasoning performance, but inhibitory control did not play a role.

Language aptitude is typically associated with success in foreign and second language learning (L2), measured by tests such as the Modern Language Aptitude Test (MLAT: Carroll & Sapon, 1959). There are several components of the MLAT, and of these, there are two subtests which are considered the most relevant for grammar. These are grammatical sensitivity and inductive learning ability, both of which can be subsumed under the term 'language analytical ability' (Li, 2015; Li et al., 2016; Skehan, 2002). The subtest that measures grammatical sensitivity does so without using part of speech terminology such as nouns and verbs. Instead, participants are presented with two sentences, and they are asked to identify a word in the second sentence that fulfils the same grammatical function as a given word in the first (see Figure 17). Therefore, this task essentially involves analogical reasoning in mapping

similarities between two sentences. In fact, Carroll (1990, p. 19) describes the grammatical sensitivity task as a 'grammatical analogies test'.

Language aptitude is not usually considered relevant to first language acquisition. This is because many researchers assume that L1 grammar acquisition relies almost exclusively on implicit learning mechanisms (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). However, in line with usage-based models, Dąbrowska (2010a), proposes that both explicit and implicit processes are involved in language learning, although at different stages and to varying degrees. This aligns with usage-based theories which propose a grammar-lexicon continuum (Fillmore & Kay, 1993; Fillmore, 1988; Goldberg, 1995). It is therefore certainly plausible that explicit knowledge, as measured by the grammatical sensitivity task, for example, is relevant to first language acquisition if rule induction relies on identifying analogical relationships of form and function (see Dąbrowska & Blake, 2022). Recent investigations show that analytic ability is related to grammar in adult L1 populations (Dąbrowska, 2018; Llompart & Dąbrowska, 2023; Prela et al., 2022; Winckel & Dąbrowska, 2024), in older LT children (Wright, Geertsema, le Roux, Winckel, & Dąbrowska, 2024), and in adults with DLD (Blake, Dąbrowska, Llompart, in preparation). To the best of our knowledge, language aptitude in the form of grammatical sensitivity has not been studied in children with DLD.

Sensitivity to the frequency of patterns is central to usage-based theories of language (Divjak, 2019; Ellis, 2002). A wide body of research has associated implicit statistical learning (ISL) to language acquisition using paradigms such as the serial reaction time (SRT) task (Nissen & Bullemer, 1987), as well as various visual and auditory statistical learning paradigms (for descriptions of different ISL tasks, see Bogaerts, Siegelman, Christiansen, & Frost, 2021; Boeve et al., 2023). Results from recent meta-analyses provide evidence of the relationship

between ISL and language, for example, examining auditory-linguistic ISL (Isbilen & Christiansen, 2022), ISL and linguistic abilities (Boeve et al., 2023), and between ISL and language and reading outcomes (Ren et al., 2023). Additionally, prior research reports procedural learning deficits in DLD population (Lammertink et al., 2017; Lum et al., 2014, 2014; Obeid et al., 2016). Some studies do not find evidence of deficits (Lammertink, Boersma, Rispens, & Wijnen, 2020), and West, Melby-Lervåg, and Hulme (2021) report mixed evidence relating to a number of tasks measuring procedural learning. Additionally, a recent meta-analysis conducted by Oliveira et al. (2023), shows a negligible relationship between procedural learning and language and literacy, as measured by the SRT task ($r = 0.06$), and no association was observed when analysing the DLD, dyslexic and LT groups separately. The variability in results has been linked to issues regarding task reliability, especially in children (Arnon, 2020). A new ISL task was designed for the present study, based on a visual search paradigm.

The above literature establishes the context of several cognitive abilities that are important to language in the context of pattern finding. In the next section, we consider specific linguistic outcomes that are of interest in the present study; these are vocabulary, grammar, and collocational knowledge.

Children with DLD acquire their first words later than language typical peers (Leonard, 2014a). There is evidence that children with DLD have problems in the encoding stage, referring to the initial memory trace linking a word form and its corresponding referent (Jackson, Leitão, Claessen & Boyes, 2021). While some research finds that DLD children have similar retention is similar to their LT peers (Haebig et al., 2017; Jackson et al., 2021; Leonard et al., 2019), other studies report poor retention (Rice et al., 1994; Riches et al., 2005). Overall,

children with DLD often have vocabularies that are not as broad or as in-depth as their LT peers (McGregor et al., 2013). This has a negative effect on grammar, as vocabulary and grammar are closely tied in early development. Children require a considerable amount of vocabulary knowledge to recognize grammatical patterns (Marchman & Bates, 1994). It is, therefore, not surprising that children with DLD have deficits in grammatical development, specifically with syntax (van der Lely, 2005), expressive morphosyntax (Leonard, 2014a), and tense marking (Rice et al., 1995, 1998; Rice & Wexler, 1996), which makes it difficult for them to understand and process language.

Whilst expressive and receptive vocabulary difficulties are well documented in DLD research, an area that has, to our knowledge, not been studied is collocational knowledge. Collocations, also referred to as 'words which go together' (Riches et al., 2022), refer to combinations of two or more words that occur together more frequently than can be expected by chance (e.g. Adj + NP combinations like '*heavy rain'*, or idioms like '*raining cats and dogs'*). In adults, collocational knowledge is related to performance in receptive vocabulary and grammar, suggesting that these aspects of language may also be supported by a learning mechanism which is sensitive to frequency (Dąbrowska, 2019; Dąbrowska, 2014; Llompart & Dąbrowska, 2020). It could be suggested that collocations provide a relatively pure measure of frequency in language.

However, there is less research regarding collocational knowledge in children. A systematic review by Smith and Murphy (2015) identified 16 collocation assessment tools that were mostly used for testing with L2 adults. Recently, Riches et al. (2022) assessed collocational knowledge in monolingual English children and children learning English as an Additional Language (EAL). Performance in the collocation task correlated with vocabulary

and grammar, although slightly weaker correlations were observed in the EAL group, possibly due to limited exposure, among other aspects. It has been shown that children with DLD may be reliant upon the frequency of multiword chunks (Kueser & Leonard, 2020), and often produce language that is limited in terms of syntactic creativity. Difficulties with receptive grammar in children with DLD may mean that they find it difficult to understand or remember formulaic phrases or that they may be less sensitive to frequently occurring patterns.

Aims and research questions

We address four pre-registered predictions (https:// osf.io/uc2qa), to examine the differences between LT and DLD children in specific cognitive abilities: language aptitude, as (assessed by a grammatical sensitivity task), analogical reasoning, implicit statistical learning, and inhibition. Given that children with DLD have challenges in both linguistic and non-linguistic tasks, we anticipate that the DLD group will perform below LT children on these measures. We investigate the extent of these differences and how they predict language outcomes in both groups.

This study has several novel approaches. Firstly, language aptitude (using a grammatical sensitivity task), is not typically assessed in L1 acquisition or in DLD populations. Secondly, we investigate analogical reasoning, which has not been extensively researched in the literature. We also explore implicit statistical learning and inhibition. The linguistic measures of interest are vocabulary, receptive grammar, sentence recall, and collocational knowledge. While vocabulary and grammar are known to be problematic for children with DLD, we are not aware of any research exploring collocational knowledge in DLD populations.

Method

Participants

Participants were recruited through advertising with organizations such as: 'Engage with Developmental Disorder', (E-DLD: https://www.engage-dld.com), and The DLD Project in Australia (https://thedldproject.com), and via social media platforms (e.g., Facebook groups 'Raising awareness for DLD', 'Developmental Language Disorder (formerly SLI) Support', and 'RADLD: Raising Awareness for DLD'). We invited children with and without DLD to participate and specified that they should be: 1) aged between 9 and 12 years of age, and 2) monolingual. In the case of DLD, we asked for a confirmed diagnosis of DLD and specified that children should not have hearing impairment or co-occurring biomedical conditions such as autism spectrum disorder, or Down syndrome. Our study was open to English speaking children in the United Kingdom, Australia, New Zealand, Canada, and the United States of America.

Overall, we recruited 86 children (43 LT children and 40 DLD children). We excluded three participants (two LT and one DLD) who completed less than half the tasks in the study. Parents reported whether their child had a diagnosis of DLD, and we conducted an additional classification measure to verify accuracy of our groupings. This involved measuring children's performance on the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals: CELF-5 (Wiig et al., 2013). Sentence repetition tasks are considered reliable clinical markers of DLD (Conti-Ramsden et al., 2001), and the difficulties with sentence repetition persist into adulthood (Poll, Miller, & van Hell, 2016), making it a stable diagnostic criterion. Participants with DLD were only included in the DLD group if they scored more than 1.25 SDs below the mean on this measure, which excluded six participants. Four additional participants originally in the LT group were excluded because they also scored more than 1.25

SDs below the mean and, thus, partially meet the criteria for the diagnosis of a language disorder. Therefore, our final sample, post-exclusion was: 39 LT children and 34 DLD children, with a mean age of 10;8 (range: 8;11 – 12;11). The mean age in the LT group was 10.79 years and 10.64 in the DLD group. The difference between the groups was not significant (p = .59).

Materials

Language tasks

Receptive grammar

The Pictures and Sentences task (Dąbrowska, 2018) is a forced-choice receptive grammar task with each test item comprising two pictures and a sentence that is read aloud by the experimenter. The child's task is to select the picture that matches the sentence. There are 88 items in the test, representing 11 different constructions which are presented in a random order so that no two items of the same construction appear consecutively. See Figure 16 for an example trial. Table 22 provides a list of constructions included in the task. The raw score was converted to a percentage score.

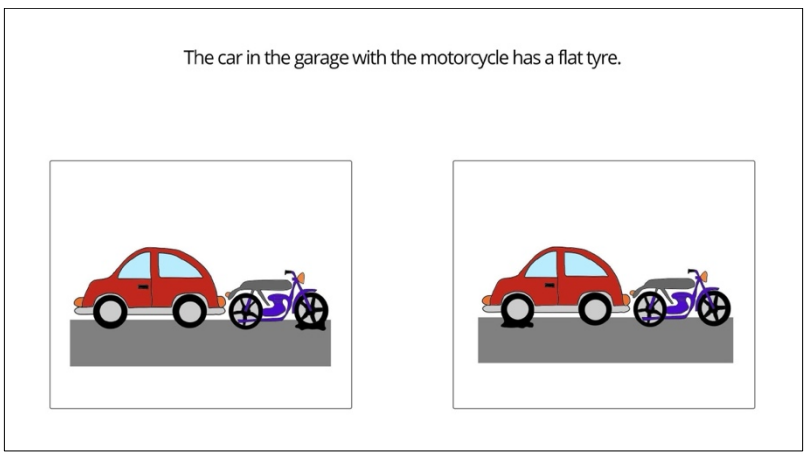


Figure 14. Test item from the Pictures and Sentences task

Table 22. Constructions in the Pictures and Sentences task

Construction	Example
Active	The boy scratched the dancer.
Passive	The dancer was scratched by the boy.
Subject cleft	It was the boy that scratched the dancer.
Object cleft	It was the dancer that the boy scratched.
Subject relative	The boy was the one that scratched the dancer.
Object relative	The dancer was the one that the boy scratched.
Quantifier + is	Every lamp is on a table.
Quantifier + has	Every table has a lamp on it.
Postmodifying prepositional phrase	The lamp on the table is white.
Complex postmodifying preposition phrase	The window in the room with the chair is broken.
X-Is-Difficult-to-Answer	The doll is hard to see

Recalling sentences

The Recalling sentences subtest of Clinical Evaluation of Language Fundamentals (CELF-5; Wiig et al., 2013) was used. In this task, the child listens to sentences and their task is to repeat the sentence back to the experimenter. The sentences increase in length and complexity during the task. The test is discontinued after five consecutive sentences containing four or more errors in each. The maximum possible score is 78.

Vocabulary

The British Picture Vocabulary Scale (BPVS3) (Dunn, Dunn, Styles, & Sewell, 2009) is a standardized receptive vocabulary task based on a picture selection format. Each test item comprises four illustrations, and the participant is asked to select the picture that best matches a word read aloud by the experimenter. There are 14 sets of 12 items, increasing in difficulty

as the test progresses. The test is discontinued after eight or more errors in a set. The maximum possible score is 168.

Collocations

The Language Detective game (Riches et al., 2022) is used to measure collocational knowledge. The task is presented as a detective game and the child is instructed that they will listen to some phrases, some of which are 'good', and some that are 'not so good'. Their task is to decide which phrase sounds better. Each of the 89 trials consists of two expressions: one target and one distractor (e.g. *'do homework'* or *'do a picture'*), each with an accompanying visual representation. There are three practice trials in which experimenter reads two sentences out loud, such as *kick ball* and *ball kick,* and the child is asked 'which one sounds better?'. The child sees two pictures, each representing one of the two sentences, and indicates their choice by repeating their selected sentence or pointing to the picture that represents their choice. The maximum score is 89 and the raw score was converted to the percentage of correct responses.

Cognitive Tasks

Language aptitude/language analytic ability (assessed by a grammatical sensitivity task)

The Matching Words subtest of the Modern Language Aptitude Test: Elementary (MLAT-E; Carroll & Sapon, 1967) is a test of grammatical sensitivity that uses verbal analogy to find words that fulfil the same function across two sentences but does not make use of traditional part of speech terminology. Participants see 30 pairs of sentences and are asked to choose the word in the second sentence that "does the same job" as a key word in the first (see

Figure 17 for an example of an item[14]). The maximum possible score is 30 and the raw score was converted to the percentage of correct responses.

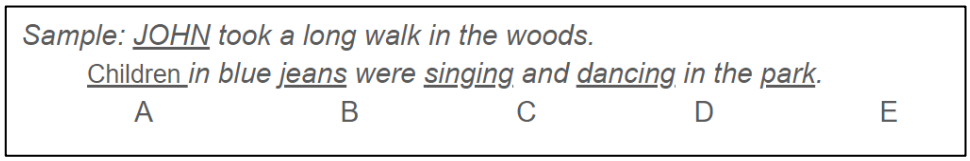


Figure 15. Sample item from the verbal analogical reasoning task.
The correct answer is A.

Analogical Reasoning

The Scene Analogy task (Richland et al., 2006) comprises of 20 scene analogy problems, adapted for online use with permission from the authors. The task involves pairs of pictures and the child's task is to observe the role of an object in the first picture (the top picture) and use this information to select an object in the second picture (the bottom picture), that plays the same role. The picture sets vary in the number of relational mappings (one relation versus two relation problems), and the presence of featural distractors (distractor versus no distractor problems). An example is shown in figure 18.

The task is presented as a PowerPoint using two functions on Zoom: screen sharing and mouse control. The experimenter read instructions to the child as per Richland et al. (2006: p. 263-264). Firstly, the experimenter explains the one-relation example picture set by saying: 'There is a certain thing that happens in the top picture, and the same thing happens in the bottom picture, but it looks different [...]. I am going to put a blue star on one thing in the top picture, and your job is to tell me *what is the same thing happening* in the bottom picture.'

[14] The example given is available on the internet as a sample item and is not an item from the test (https://lltf.net/mlat-sample-items/mlat-part-iv/)

The two-relation problems are explained as follows: 'Now sometimes what is happening will have two parts, like the one you just saw [as per first example] and sometimes there will be three parts to what is happening. Let me show you what I mean.' The experimenter then explains both pictures, and says: 'If I put the blue star on [...], what is the same thing happening in the bottom picture? The task begins with two practice trials in which the experimenter verifies the child's understanding before the child begins. For every set of pictures, the experimenter says: "If I put my blue star on [...] in the top picture, where will you put your yellow star in the bottom picture? What is the same thing happening?" The example in figure 18 shows the target item in the top picture (indicated by a blue star), with the correct item in the bottom picture (indicated by a yellow star).

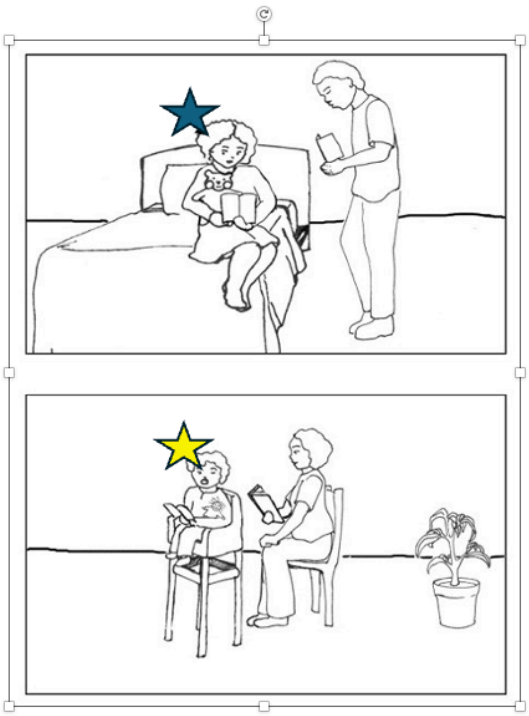


Figure 16. Example of a sample item from the Scene Analogy task
Two relations/no distractor (with permission from Richland, Morrison, & Holyoak, 2006).

Inhibition

Inspired by Krzemien et al. (2020), we designed a Go/No-Go (GNG) game modelled on an original design with permission from the authors (Geurten, Catale, & Meulemans, 2016). Children are introduced to 'Nellar', a friendly cartoon alien (see Figure 3). The storyline is presented as two aliens who are playing hide-and-seek. When the child sees Nellar (a green alien), they should help him to hide by pressing the spacebar. But when they see Nellar's cousin, Neera (a red alien), they should not press anything. The task begins with a practice block of eight trials with feedback provided for correct and incorrect trials (for example, if the child does not click on Nellar, they hear a noise and see the words: '*Uh oh, remember when you see Nellar, press the space bar'*). The child is advised that they will gather gold coins as they play, and they will see their score at the end. There are 40 test trials, and no feedback is given in the test phase. As per Geurten et al. (2016), there are 15 Go trials and 25 No-Go trials. The stimuli are presented for 350ms, with varied interstimulus intervals ranging from 1900 to 4100ms so that the child could not anticipate the presentation of the stimulus. Responding to the distractor stimulus is a commission error whereas not responding to a target stimulus is an omission error. The maximum score is 40 and the raw score was converted to the percentage of correct responses.

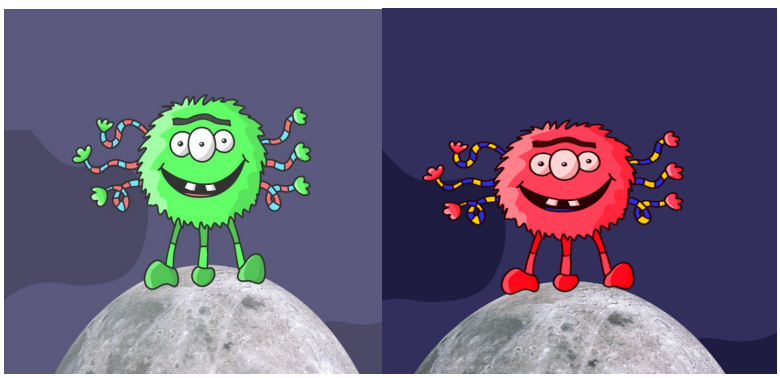


Figure 17. Example trials from the Go/No-Go task.
(A 'go' trial is depicted on the left with the green alien, and a 'no-go' trial on the right with a red alien).

Implicit statistical learning

Where's Nellar? (Wright, Prela, Riches, Blake & Dąbrowska, in preparation) is an implicit statistical learning task based on a visual search paradigm. Children meet Nellar, the friendly alien, once again, and they also see Nellar's friends (distractor aliens). The children are told that they will play a game of hide-and-seek, and they should try and catch Nellar by clicking on him as fast as they can. The child is then presented with a 3 x 3 grid of nine aliens, including Nellar, who moves positions in each trial (see Figure 4). During the practice trials, if the child clicks on an incorrect alien, they hear a noise and see the words '*Uh oh'*. If they click correctly, they hear a sound and see the words '*Well done'*. After three practice trials, they are informed that they will try to catch Nellar on every planet. The task is divided into six planets, representing artificial levels between blocks of trials to keep participants motivated to complete the task. Each planet consists of four blocks, with a total of 24 blocks in the task. Each block contains four triplets (a fixed sequence of three locations where Nellar appears e.g. locations 1-7-3 in the 3x3 grid) and the presentation of each triplet is preceded by the appearance of Nellar in a random filler position, separating the triplets. Thus, each sequence of a filler-triplet-combination is a sequence of two random and two predictable locations given that the filler position is random, as is the move to the first position in the triplet. The following two locations are predictable from the initial triplet position. There were no consecutive appearances in the same location. In the first block, participants start off with an additional three fillers to ensure that mistakes on the initial trials do not influence the results. Participants start on level 1 and, after every four blocks, see a level-up screen to mark progress and motivate the participants. A timeout of 2000ms per trial was imposed, after which participants heard negative audio feedback before Nellar moved to the next position. Both a motor learning and a sequence learning score was obtained from this task.

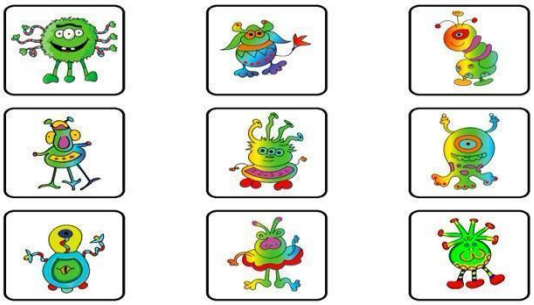


Figure 18. Example trial from 'Where's Nellar?'
The target alien is in the top left location.

Procedure

Before taking part in the study, parents and caregivers were asked to complete a consent form and a background questionnaire which included information regarding the child's language ability, parents' education, and employment. They were then invited to book the online sessions which were conducted using the videoconferencing platform, Zoom (https://zoom.us). There were two moderated sessions, with two experimenters hosting one session each. Moderation of the sessions allowed the experimenters to verify task understanding, and children could ask questions or take a break if required. The only tasks that were not moderated were two games hosted on the Gorilla Game Builder platform (https://gorilla.sc: Anwyl-Irvine et al., 2020). These were the GNG and 'Where's Nellar?' tasks. In these tasks, written instructions were provided together with additional audio-recorded instructions that played automatically throughout the task to support comprehension.

In session one we administered the vocabulary test, the analogical reasoning task, and the recalling sentences task. After completing this session, participants received the login information to complete the two online tasks for inhibition and implicit learning. The final session consisted of the receptive grammar, collocations, and grammatical sensitivity tasks.

Data Processing

The tests for recalling sentences, vocabulary, and grammatical sensitivity were administered and scored according to the testing manuals.

For the analogical reasoning task, we recorded all responses. Errors were coded according to three categories: distractor errors, relational errors, or 'other errors' (as per Richland et al., 2006). The score for the analogical reasoning task is the number of incorrect responses, reported as a negative value from a baseline of zero.

For the receptive grammar task (Pictures and Sentences) and the task for collocations (the Language Detective Game), we recorded the number of correct responses. This was converted to a percentage of correct responses.

For the measure of inhibition, we extracted a score based on the number of correct responses (i.e. responding or not responding as required by the type of trial). Six trials were excluded from data processing for being above 4500ms (longer than the timeout built-in to the experiment) or below 5ms (impossibly short).

In the implicit statistical learning task, 172 trails with reaction times over 3000ms were excluded. These trials were more than 1000ms over the programmed timeout for the online task. As the task was conducted online, connection or server problems can occasionally result in reaction times that should not be possible as they are longer than the programmed timeout. An additional 254 trials under 100ms were excluded as these are likely mis-clicks. Furthermore

1128 trials that were more or less than three times the mean absolute deviation by participant-specific mean reaction time were excluded. After exclusions a total of 655 timeouts remained, of which 329 occurred in random trials and 326 occurred in predictable trials. One participant with ten or more consecutive timeouts was excluded as were an additional ten participants with 30 or more timeouts in total (14.78% of the trials) as these likely indicate poor engagement with the task. A sequence learning score was derived from the difference between the mean reaction time on random and predictable trials across all blocks. Thus, a positive score means that the movement patterns were learned implicitly, and that the individual was faster on predictable trials. Following Divjak et al. (2022) and Llompart and Dąbrowska (2020), we calculated a motor learning score by calculating individual slopes for the first 15 blocks. The reaction time data was transformed by inverting and scaling to improve normality. Trial numbers were standardized, and a linear mixed-effects model was fitted to predict the transformed reaction time with random slopes and intercepts for participants. Coefficients from the model were extracted to obtain individual participant intercepts and slopes.

Results

Data was imported to R (ver. 4.2.1, R Core Team, 2022) for processing.

Descriptive statistics and reliability

Descriptive statistics and reliability for the assessment measures are presented in Table 23. For scored measures, the reliability was calculated using 5000 random splits using the `splithalf` package (Parsons, 2021), and the reliability reported is the Spearman-Brown corrected reliability estimate. For the sequence learning, the reliability reported is the correlation between learning (random – predictable reaction time) along an odd/even block

split, following the methodology proposed by Farkas et al. (2024). Reliability was not calculated for the standardised measures or for motor learning. The low reliability for analogical reasoning in the LT group is due to ceiling effects, there is too little variance in the data.

The between group t-tests and effect sizes for the different measures are reported in Table 23. There are large between-group differences on all language measures as well as on nonverbal analogical reasoning. These differences are reflected in the distribution plots in Figure 21. Figure 22 illustrates that there are no significant differences in performance on the motor learning, sequence learning, and inhibition measures. The analogical reasoning scores reflect the number of errors made, with the DLD group have a wider range of scores.

Table 23. Mean, median, standard deviation (SD), reliability, *t*-tests

Measure	Mean: LT	SD: LT	Reliability: LT	Mean: DLD	SD: DLD	Reliability: DLD [95% CI]	*t*-test	Cohen's *d*	*p*-value
Recalling sentences (raw)	58.41	7.01	-	29.38	8.14	-	*t*(-16.20) 65.65	-3.84	<.001
Recalling sentences (standardized)	11.70	2.83	-	5.29	1.62	-	*t*(-12.00) 62	-2.72	<.001
Vocabulary (raw)	139	12.45	-	115.2	16.02	-	*t*(-7.28) 61.98	-1.74	<.001
Vocabulary (standardized)	103.00	11.97	-	81.80	9.18	-	*t*(-8.6) 70	-1.98	<.001
Receptive grammar	89.48	6.04	.72 [.56, .84]	73.06	8.59	.71 [.55, .83]	*t*(-9.32) 58.17	-2.24	<.001
Collocations	80.84	7.29	.68 [.51, .81]	65.86	7.79	.59 [.38, .77]	*t*(-8.45) 68.12	-1.99	<.001
Grammatical sensitivity	62.56	19.53	.83 [.73, .89]	33.63	10.65	.36 [.02, .63]	*t*(-7.99) 60.26	-1.81	<.001
Analogical reasoning	-2	1.23	.06 [-.29, .44]	-3.24	2.02	.47 [.16, .07]	*t*(-3.92) 50.19	-0.95	<.001
Inhibition	85.45	9.71	.69 [.52, .82]	86.54	9.23	.7 [.51, .83]	*t*(0.94) 70.45	0.12	.62

Motor learning	-0.03	0.03	-	-0.02	0.02	-	*t*(1.35) 59.27	0.33	.18
ISL: Sequence learning	23.01	45.52	.64 [.40, .80]	9.57	31.01	-.18 [-.53, .24]	*t*(-1.44) 59.60	-0.35	.16

Nonverbal analogical reasoning: number of errors made on the task. Sequence learning: mean random RT – mean predictable RT. Motor learning: individual slope. Inhibition: percentage correct responses. Verbal analogical reasoning: percentage correct responses. Collocations: percentage correct responses. Receptive grammar: percentage correct responses (raw and standardized scores). Vocabulary (raw and standardized scores).

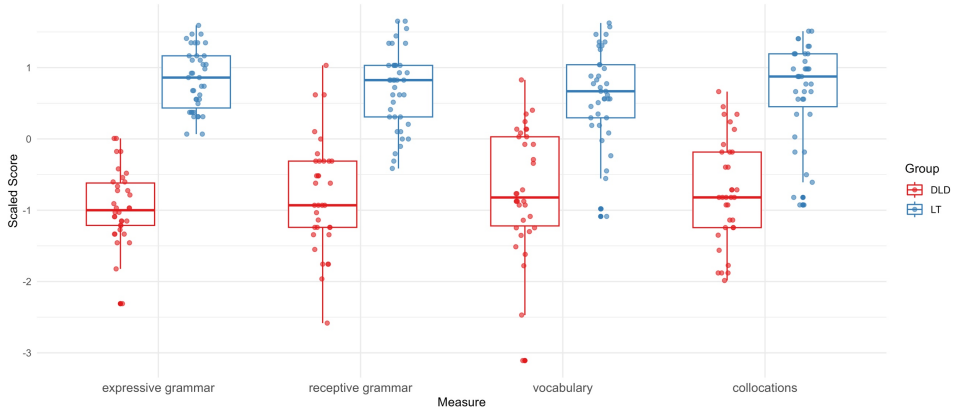


Figure 19. Distribution by group on the language tasks

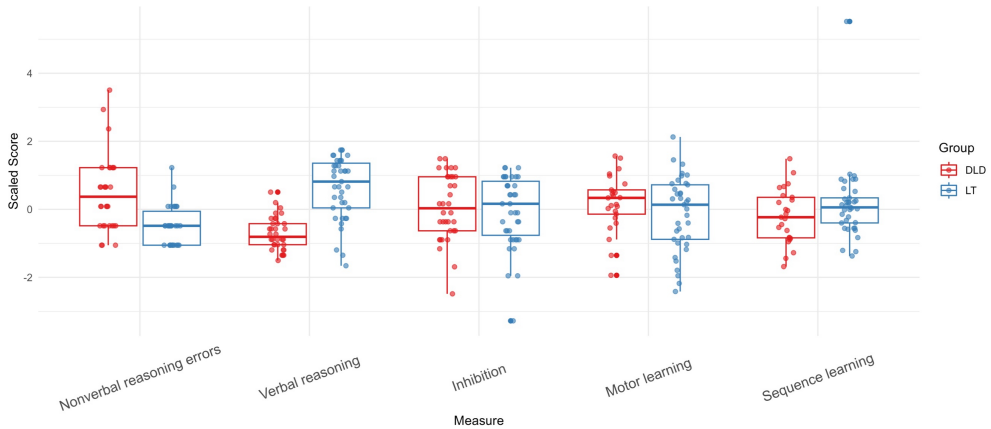


Figure 20. Distribution by group on the cognitive measures

Analogical reasoning

In analysing the results from the analogical reasoning task, we followed the method from Richland et al. (2006) to look at accuracy across different trial types. Figure 23 depicts the scores for each of the four trial types. Trials vary by number of relations and the presence or absence of a distractor. The LT group had higher accuracy across all trial types and gaps between performance for the two groups are largest when a distractor is present. The majority of errors made by both groups were relational errors (75.44% of all errors), followed by distractor errors (15.20%) and 'other' errors (9.36%). Errors are made more often in scenes that are more complicated to process in terms of the number of relations and the presence of a distractor.

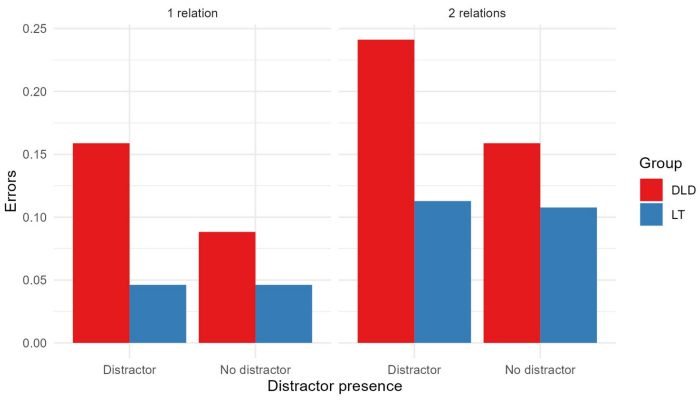


Figure 21. Error rates by trial type in the analogical reasoning task

Receptive grammar

Figure 24 shows the mean score of each group by construction. The maximum score for each construction is 8 and chance is indicated by the dashed line. Both groups are at or near ceiling on actives, the control condition. The group differences vary across constructions, with groups performing similarly on subject relatives and subject clefts and larger differences visible for complex NP, object relatives, and object clefts. The LT group performed at chance only on the *Q-has* construction, while the DLD group performed below chance on both the *Q-has* and object relative constructions.

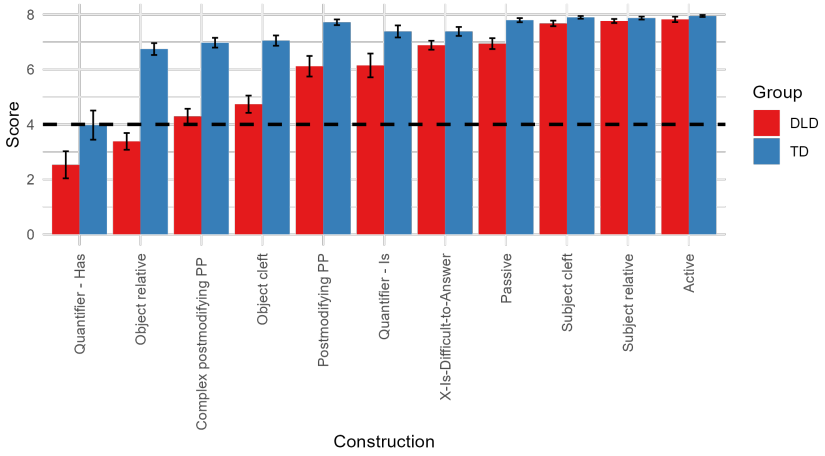


Figure 22. Accuracy rates by construction for the receptive grammar task

Implicit statistical learning RT

Figure 25 shows the mean reaction time by block number for all participants, separated by trial type. There is a visible downward slope in reaction time from Block 1 to Block 9 for both random and predictable trials, likely due to practice effects or possibly due to motor learning taking place. From block 12, the RTs on unpredictable trials are longer than for predictable trials, indicating that participants have implicitly learned the sequences.

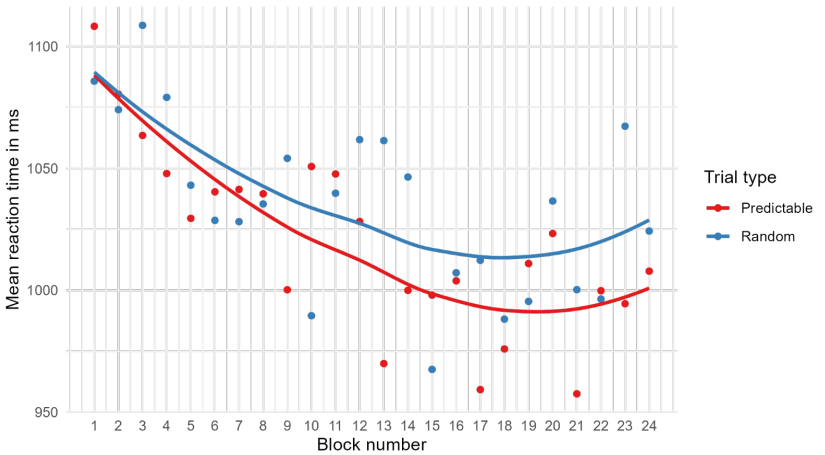


Figure 23. Reaction time on the ISL task by block for all participants

Figure 26 shows the mean reaction time by block number separated by trial type for LT participants, and participants with DLD. The reaction times of the LT participants show a

similar pattern to the one for all participants, although the divergence of the two trial types appears to take place sooner. When compared to Figure 25, it is visible that the divergence of the two reaction time regression lines takes place from Block 11 onwards for the participants with DLD. The participants show a similar learning effect, just later. Level-up screens were provided after every four blocks to motivate participants in the task. Blocks 1, 5, 9, 13, 17, and 21 were immediately after a level indication.

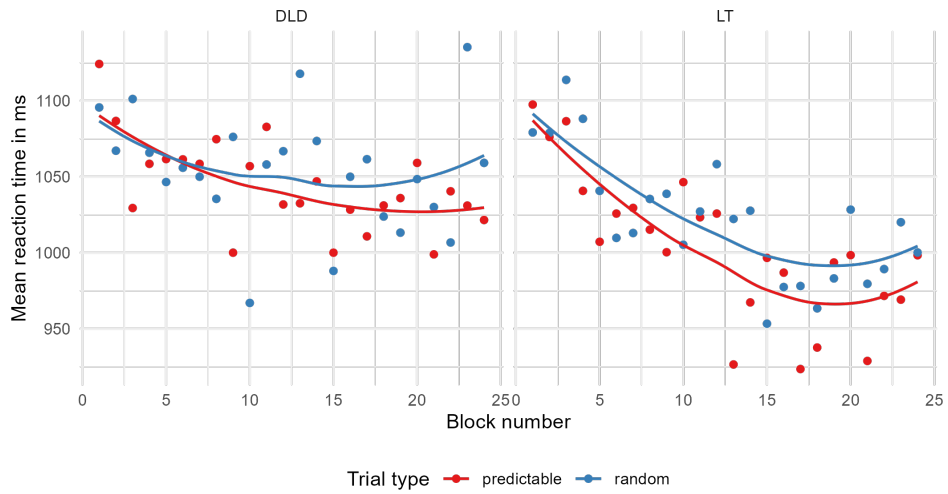


Figure 24. Reaction time on the ISL task by block for LT and DLD participants

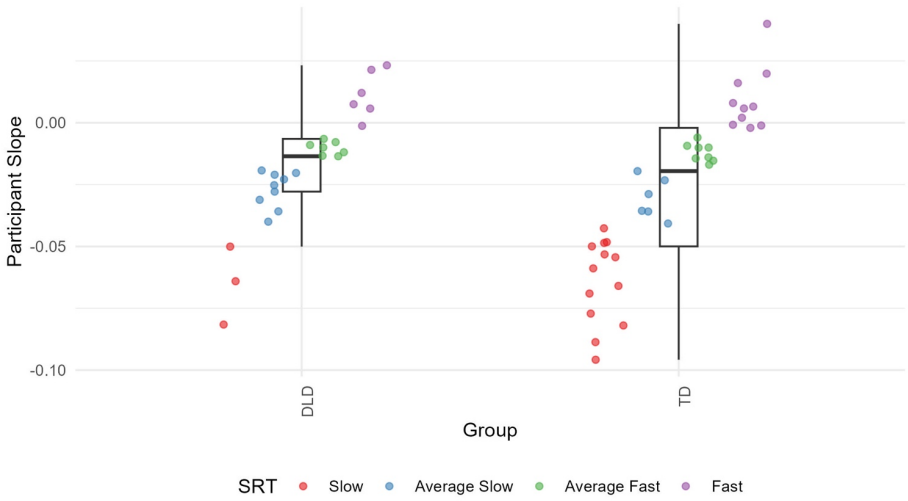


Figure 25. Motor learning distribution
(based on individual participant slopes by slope quantiles)

Collocations

The collocations task consists of three kinds of collocations and the accuracy for each kind is depicted in Figure 28. Collocation refers to the kind found in words that go together at a statistically significant frequency. There are three types: binomials, word pairs, and idioms. Within the word pairs position there are three types of word pairs; VP + NP, VP + Adj, Adj _ NP (Riches et al., 2022). Idioms refer to fixed expressions e.g. it fits like a glove. Binomial collocations typically have a "NP and NP" structure that occurs in one order much more frequently than in the other e.g. salt and pepper. Both groups were above chance (0.5) on all three types of collocations. The gap between the LT and DLD groups was smallest for idioms, and this was the type of collocation in which both groups had the lowest score.

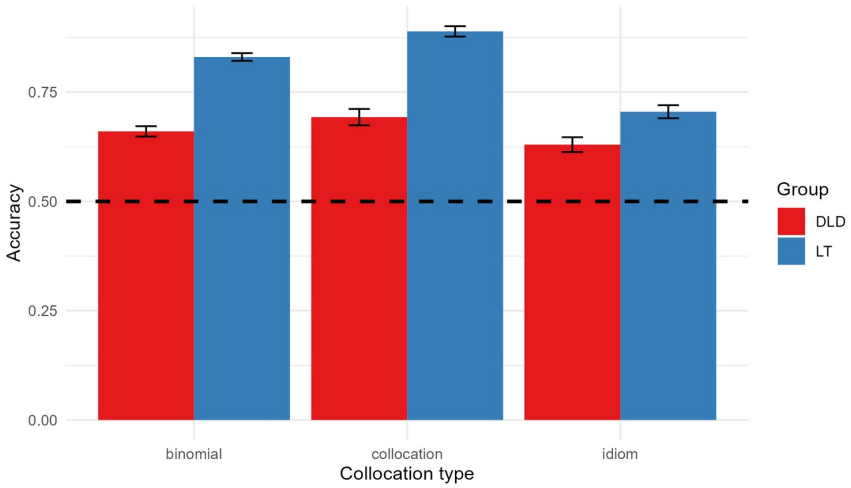


Figure 26. Accuracy by collocation type
(chance line is indicated by a dotted line)

Correlation analysis

Correlations between the measures are reported for the DLD group in Table 24 and for the LT group in Table 25. In the LT group, grammatical sensitivity is correlated with all the language measures as well as inhibition. In the DLD group, grammatical sensitivity is only correlated with analogical reasoning. Analogical reasoning is not related to any language or cognition measures in the LT group.

Table 24. Correlations for the DLD group (n = 34)

	Age	Vocab	RS	Coll	RGr	GrS	AR	ISL	ML	Inhibition
Age	-									
Vocabulary	.50**									
Recalling sentences	.31	.42*								
Collocations	.35*	.29	.51**							
Receptive grammar	-.20	.12	.22	.26						
Grammatical sensitivity	-.06	.07	.19	.20	.21					
Analogical reasoning	-.04	.26	.07	.14	.60**	.46**				
ISL - Sequence learning	-.10	-.36*	-.14	-.09	-.15	.1	-.33			
Motor learning	.11	.18	.21	-.01	.42*	.15	.25	-.33		
Inhibition	-.11	.13	-.23	-.46**	.12	-.06	.13	-.03	.02	-

*Note: correlations of .44 and above are significant at p < .01** and correlations of .34 and above are significant at p < .05*.*

RS: Recalling sentences; Coll: collocations; RGr: receptive grammar; AR: analogical reasoning (error scores), GrS: grammatical sensitivity, analogical reasoning, ISL: implicit sequence learning; ML: motor learning. Raw scores are reported throughout.

Table 25. Correlations for the LT group (n = 39)

	Age	Vocab	SR	Coll	RGr	GrS	AR	ISL	ML	Inhibition
Age	-									
Vocabulary	.44**									
Recalling sentences	.47**	.59**								
Collocations	.59**	.52**	.48**							
Receptive grammar	.32*	.56**	.47**	.41**						
Grammatical sensitivity	.48**	.47**	.44**	.48**	.54**					
Analogical reasoning	0	.19	.07	.11	.11	.24				
ISL - Sequence learning	-.20	-.34*	-.09	-.23	-.28	-.1	-.09			
Motor learning	.09	0	-.1	.13	.26	.22	.03	-.07		
Inhibition	-.02	-.01	.1	.16	.39*	.49**	-.15	-.02	.19	-

*Note: correlations of .41 and above are significant at p < .01** and correlations of .32 and above are significant at p < .05*.*

RS: Recalling sentences; Coll: collocations; RGr: receptive grammar; AR: analogical reasoning (error scores), GrS: grammatical sensitivity, analogical reasoning, ISL: implicit sequence learning; ML: motor learning. Raw scores are reported throughout.

Inferential analysis

Numerical predictor variables were center scaled. Although we found between-group differences we lacked the power to run complex models with interaction and instead ran separate models for each of the two participant groups. We predicted the language outcomes (vocabulary, receptive grammar, collocations, and recalling sentences) from grammatical sensitivity, analogical reasoning, inhibition, and implicit statistical learning. The relative contribution (lmg) of predictors to the variance in the outcome variable was calculated with the `relaimpo` package (Grömping, 2006). The summaries of the model outputs are combined for both groups for vocabulary (Table 26), receptive grammar (Table 27), collocations (Table 28), and recalling sentences (Table 29).

As shown in Table 26, the overall model for vocabulary accounts for 59% of variance in the LT group, where grammatical sensitivity is the most robust predictor, contributing 35% of the total variance. Sequence learning and inhibition further contribute an additional 13% and 7% respectively. In contrast, there are no significant predictors in the DLD group. The model accounts for 15% of the total variance in vocabulary scores but the adjusted R^2 for the DLD group is negative (-0.13), highlighting potential limitations in the model, likely due to the small sample size.

Table 27 presents the model output for receptive grammar. The model accounts for 46% of the variance in the LT group, while in the DLD group, it explains 66% of the overall variance. Grammatical sensitivity emerges as a significant predictor for the LT children, explaining 27% of the variance in the model. For the DLD group, the majority of variance (48%) is accounted for by analogical reasoning, with motor learning contributing an additional 13%.

Table 28 provides the models for collocations. In the LT group, the model accounts for 25% of the variance, with grammatical sensitivity explaining 18% of variance. However, for the DLD group, the model accounts for 32% of the overall variance, yet none of the predictors reach statistical significance. Additionally, the adjusted R^2 in the DLD group is quite low (0.08), suggesting a weaker model fit for this group.

A similar pattern is observed in recalling sentences (Table 29). In the LT group, the overall model explains 42% of the variance, with grammatical sensitivity accounting for 28% of the variance. In contrast, the model for the DLD group accounts for 16% variance, but the predictors do not reach significance. The negative adjusted R^2 of -0.12 indicates that the model does not provide a good fit, potentially due to limitations related to sample size and the number of predictors examined in this study.

Table 26. Model outputs for Vocabulary

	LT Estimate (Standard error)	lmg	DLD Estimate (Standard error)	lmg
(Intercept)	133.61 ***		112.52 ***	
	(2.30)		(7.03)	
Grammatical sensitivity	11.21 ***	.35	0.85	.02
	(2.36)		(8.76)	
Nonverbal analogical reasoning	1.55	.02	-1.15	.02
	(3.57)		(3.99)	
Sequence learning	-3.55 *	.13	-5.92	.08
	(1.38)		(5.28)	
Motor learning	-2.23	.03	-0.18	.004
	(1.70)		(4.75)	
Inhibition	-4.85 *	.07	3.11	.04
	(1.87)		(3.76)	
N	29		21	
R^2	0.59		0.15	
Adjusted R^2	0.50		-0.13	
F (5,23) 6.71, p = < 0.001			F (5,15) 0.53 , p = 0.75	

**** $p < 0.001$; *** $p < 0.01$; * $p < 0.05$.*

Table 27. Model outputs for Receptive Grammar

	LT Estimate (Standard error)	lmg	DLD Estimate (Standard error)	lmg
(Intercept)	87.95 ***		73.10 ***	
	(1.21)		(2.73)	
Grammatical sensitivity	3.81 **	.27	-4.56	.02
	(1.24)		(3.41)	
Nonverbal analogical reasoning	2.41	.02	-6.99 ***	.48
	(1.87)		(1.55)	
Sequence learning	-1.05	.07	2.02	.02
	(0.72)		(2.05)	
Motor learning	1.05	.04	4.17 *	.13
	(0.89)		(1.85)	
Inhibition	-0.02	.06	0.00	.01
	(0.98)		(1.46)	
N	29		21	
R^2	0.46		0.66	
Adjusted R^2	0.34		0.54	
F (5,23) 3.9, p = 0.01			F (5,15) 5.75 , p = 0.003	

**** $p < 0.001$; *** $p < 0.01$; * $p < 0.05$.*

Table 28. Model outputs for Collocations

	LT Estimate (Standard error)	lmg	DLD Estimate (Standard error)	lmg
(Intercept)	78.57 ***		69.22 ***	
	(1.87)		(3.04)	
Grammatical sensitivity	4.32 *	.18	2.88	.06
	(1.92)		(3.79)	
Nonverbal analogical reasoning	0.93	.004	-1.15	.03
	(2.90)		(1.72)	
Sequence learning	-1.35	.05	-0.84	.01
	(1.12)		(2.28)	
Motor learning	-0.45	.003	-0.51	.001
	(1.38)		(2.05)	
Inhibition	-0.66	.01	-3.36	.21
	(1.52)		(1.62)	
N	29		21	
R^2	0.25		0.32	
Adjusted R^2	0.08		0.08	
F (5,23) 1.54, p = 0.21			F (5,15) 1.38 , p = 0.28	

**** $p < 0.001$; *** $p < 0.01$; * $p < 0.05$.*

Table 29. Model outputs for Recalling Sentences

	LT Estimate (Standard error)	lmg	DLD Estimate (Standard error)	lmg
(Intercept)	55.56 ***		27.48 ***	
	(1.51)		(3.48)	
Grammatical sensitivity	5.64 **	.28	-1.84	.01
	(1.54)		(4.34)	
Analogical reasoning	1.98	.01	0.32	.01
	(2.34)		(1.98)	
Sequence learning	-0.33	.004	-0.07	.002
	(0.90)		(2.61)	
Motor learning	-1.71	.08	1.25	.02
	(1.11)		(2.35)	
Inhibition	-2.10	.04	-2.77	.13
	(1.22)		(1.86)	
N	29		21	
R^2	0.42		0.16	
Adjusted R^2	0.29		-0.12	
F (5,23) 3.29, p = 0.02			F (5,15) 0.56 , p = 0.72	

Discussion

The aim of this study was to explore differences in specific cognitive abilities (language aptitude as measured by a grammatical sensitivity task, analogical reasoning, implicit statistical learning and inhibition), in children with and without DLD. If language is dependent upon these abilities (as discussed in the introductory section), we expect children with DLD to have weaknesses in these areas. Secondly, we examine whether differences in cognitive abilities predict linguistic outcomes in the two groups.

Language measures

Given that children with DLD are diagnosed according to their language difficulties, it is unsurprising that those assessments which explicitly assessed linguistic skills were weakest in the DLD group. The most striking difference is evident in the sentence repetition task where children with DLD perform considerably lower than their LT peers (d = -3.84). Sentence repetition tasks are considered a stable and reliable clinical marker of DLD (Conti-Ramsden et al., 2001; Poll et al., 2016), and our results reflect this. We also find large differences between the group for receptive grammar, vocabulary and collocational knowledge. The differences are larger for collocations (d = -1.99) than vocabulary (d = -1.74), but smaller than for grammar (d = 2.24). Performance in each of the language tasks is discussed below.

Receptive grammar

Results from the receptive grammar task show that the two groups perform similarly on frequent structures, including actives, subject relatives, and subject clefts. The disparities between the groups become apparent in the less frequent structures, such as object relatives, object clefts, and complex postmodifying prepositional phrases. Both groups performed at (LT)

or below chance (DLD) on the *Q-has* construction (Figure 24). Street and Dąbrowska (2010) report that Q-is is more frequent than Q-has, even in the written modality. They tested adults on these structures and their participants with low academic attainment performed best on actives (mean score of 97%), followed by passives (mean score of 88%), *Q-is* (mean score of 78%), and *Q-has* (mean score of 43%). Our participants with DLD performed below the adults in that study, but followed a similar pattern of results, hinting at exposure, especially in educational settings, as a possible factor in their receptive grammar knowledge. Individuals who have difficulties with language processing may read less and, as a result, have less cumulative exposure to the more complex structures and structures found predominantly in written language. In fact, children with general reading difficulties are less accurate in interpreting object relatives than children without reading difficulties (Cilibrasi et al., 2019).

Another possibility is that children with DLD have difficulty processing more difficult structures due to working memory deficits (see Montgomery et al., 2010; Henry & Botting, 2017; Archibald, 2017). Working memory also plays a role in filler-gap dependencies such as relative clauses; thus, object relative clauses, where the distance between the filler and the gap is relatively large, are more difficult to process than subject relatives (Gibson, 1998, 2000).

Collocations

There are very few studies that examine collocational knowledge in children (see Riches et al., 2022; Smith & Murphy, 2015), and to the best of our knowledge, collocations have not been explored in children with DLD. Collocational knowledge generally refers to the ability to use word combinations that occur together more frequently than expected by chance. This knowledge is related to receptive vocabulary and grammar in LT adults (Dąbrowska, 2014, 2019; Llompart & Dąbrowska, 2020), and children (Riches et al., 2022). Knowing which

and how words go together supports the fluent production of natural-sounding language and may aid language comprehension by speeding up the processing of frequently occurring phrases or chunks.

Another possibility is that performance in the collocation task is linked to vocabulary knowledge. Our results show that the children in the DLD group perform below the LT group in vocabulary. In the LT group, there is a strong relationship between collocations and vocabulary (r = .52), whereas in children with DLD, this relationship is much weaker (r = .29). The relationship between collocational knowledge and vocabulary is possibly bidirectional. It could be that children with DLD are just as good at inferring meanings of new words from context (see Dąbrowska, 2009), but perhaps they find it more difficult to track probabilities. Therefore, they rely primarily on inference whereas LT children rely on both inference and the co-occurrence of statistics.

Vocabulary

Our results showed large between-group differences for vocabulary (Cohen's d = -1.74), however we found that vocabulary is strongly correlated with age in both the DLD and LT groups (r = .50, .44 respectively). Children with DLD are late in acquiring their first words and once they begin talking, their language ability tends to be similar to children of a much younger age (Leonard, 2014a). We know from prior research that children with DLD tend to use less sophisticated vocabulary than that of their LT peers, in terms of volume and complexity (see McGregor et al., 2013).

Cognitive abilities

The DLD group show weaker performance in the analogical reasoning task in comparison to LT children, with between-group differences of d = -0.95. We observe large between group differences in language aptitude (grammatical sensitivity) (d = -1.81). There are no significant differences in performance in the remaining two cognitive abilities assessed in this study, implicit statistical learning and inhibition. These will be discussed later on in this section.

Analogical reasoning

Analogical reasoning is the most significant predictor of receptive grammar in the DLD group, accounting for 48% of variance in the model. The fact that we find such a strong relationship is surprising considering the low reliability for this task, which suggests that the effect could in fact be even higher. In the LT group, performance in the analogical reasoning task is at ceiling, resulting in low variability in scores and a low predictive value of the task.

Language aptitude (grammatical sensitivity)

In the LT group, language aptitude (grammatical sensitivity) predicts 27% of variance in receptive grammar. It also predicts vocabulary (35% of variance), collocations (18%), and recalling sentences (28% of variance in the model). Grammatical sensitivity is therefore central to linguistic outcomes in LT children, but it does not predict language outcomes in the DLD group. As discussed in the introduction, the language aptitude subtest used in this study (the grammatical sensitivity task), involves detecting and mapping linguistic similarities between two sentences and requires metalinguistic awareness to identify and map the relationship

between structures. Analogical reasoning is central to this process and therefore this task can be considered a verbal analogical reasoning task.

With this in mind, the most striking effect from our results is that analogical reasoning predicts language outcomes in both groups. In DLD children analogical reasoning predicts receptive grammar, and in LT children, verbal analogical reasoning (as in grammatical sensitivity), predicts all linguistic outcomes. There is also a strong positive correlation between grammatical sensitivity and nonverbal analogical reasoning in the DLD group *(r* = .46). We propose that the reason these tasks predict language differently between the two groups is due to the fact that there are ceiling effects in the LT group on the analogical reasoning task whereas in the DLD group, results in the grammatical sensitivity task are just above chance (mean score of 33%). Although performance in the grammatical sensitivity task is much higher in the LT group (mean score of 62%), the range of scores (13-86%) suggest that the task is also difficult for LT children. Weak performance in the grammatical sensitivity task in DLD children is reflective of their difficulty with analogical reasoning.

Implicit statistical learning

Turning to performance in implicit statistical learning (ISL), we expected to find a difference between the two groups in our ISL task, based on evidence from prior research which suggests an ISL deficit in children with DLD. Our results show similar amounts of sequence learning in both groups, although learning occurs earlier in the LT group. Response times between the groups diverge early in the task (from approximately block eleven onwards), and from thereon, children with DLD are consistently slower throughout the task. The fact that children with DLD appear to have overall slower processing speed in this task is consistent with studies that show slower RTs in ISL tasks in children with DLD (Gabriel et al., 2013;

Lum et al., 2010; Tomblin et al., 2007). We assessed both sequence learning and motor learning in ISL and it is interesting to note a correlation between motor learning and receptive grammar in the DLD group (r = .42). In the regression model for receptive grammar, motor learning contributes to 13% of the variance in the model for DLD children. These findings are consistent with the Procedural (circuit) Deficit Hypothesis which suggests that procedural memory plays a role in motor and sequence learning, and neural deficits in procedural learning account for grammatical difficulties which occur in DLD (Ullman, 2001; Ullman & Pierpont, 2005; Ullman et al., 2020).

We anticipated that ISL would predict language - however, this hypothesis was not met, and we found that neither of the groups appeared to rely on ISL for their overall linguistic knowledge. Our data shows that ISL contributed to performance on vocabulary in the LT group, but the effect was relatively modest (13% of variance). Some studies find minimal or no evidence of an ISL deficit in DLD (Lammertink, Boersma, Rispens, & Wijnen, 2020; Oliveira et al., 2022; West et al., 2021), and many researchers express concerns regarding poor reliability of ISL tasks, especially in children (Arnon, 2020). We designed a new ISL task for this study and report reliability of .64 for the LT group, yet only .18 for the DLD group. The low reliability on this task in the DLD group could explain the inconsistent results reported in the literature and we propose that this should be pursued in future research.

Inhibition

We expected that children with DLD would show effects of interference in the Go/No-Go task (a measure of inhibitory control). This hypothesis was not met, and our data revealed similar performance in both groups. These results are contrary to prior research which suggests that inhibition could be related to analogical reasoning performance in children with DLD

(Krzemien et al., 2020). In the LT group, inhibition did not correlate with any of the linguistic measures and in children with DLD, inhibition approached significance as a predictor of collocational knowledge, explaining 21% of the variance.

Hypotheses summary

We pre-registered four hypotheses pertaining to this study. Firstly, we predicted that we would find between group differences in all cognitive predictors. This hypothesis was partially met, with large between-group differences in the grammatical sensitivity task (d = -1.81), and the analogical reasoning task (d = -.95). Our remaining two hypotheses were not met, as we did not find interference effects in the DLD group based on our inhibition task, and there were no significant differences between groups on implicit statistical learning. Both groups showed similar amounts of learning in ISL, although the DLD group were slower overall. The key finding to emerge from the present study is that analogical reasoning predicts language outcomes in both groups.

Limitations

There are limitations that should be considered in generalizing the results of this study. We did not assess nonverbal IQ and working memory and it is highly possible that these may be mediating factors for some of the tasks (as per Blake, Dąbrowska, & Riches, in preparation).

The language difficulties of children with DLD were considered in the assessment process and both visual and auditory instructions were given for all tasks. The vast majority of the tests did not rely on reading as a skill and where reading was required in the grammatical sensitivity task, the researcher was present to ensure that the sentence was read correctly. Although we did not examine literacy-based outcomes, including a measure of reading ability

would have allowed for a more fine-grained analysis of the language abilities of the participants. Reading ability is closely tied to print exposure with the direction of causation likely running from ability to exposure (Van Bergen et al., 2018). Print exposure is also an important source of exposure to lexically rich and syntactically complex language. This means that the extent to which reading difficulties are present may have a significant impact on the quantity of complex language that children are exposed to. This effect is hinted at in the receptive grammar task where children with DLD show weaker performance on the structures that are more common in written language.

Lastly, with 73 participants in this study (39 LT children and 34 children with DLD), we recognise the limitation of statistical power, particularly given the number of predictors examined. Furthermore, whilst children with DLD were matched to language typical children based on age, future research should consider including an additional language matched group. This would enhance comparability and strengthen the findings reported in this study.

Conclusion

Overall, the results of this study add to our understanding regarding the cognitive processes underlying language acquisition children with and without DLD. The most significant finding to emerge from this study is the relationship between analogical reasoning and language in both groups, with analogical reasoning predicting receptive grammar in children with DLD and grammatical sensitivity (as a verbal analogical reasoning task) predicting all linguistic outcomes in LT children. According to usage-based theories, analogy is key to children developing novel abstract utterances of their own. Although children with

DLD perform below LT children in analogical reasoning, this ability is fundamental to their processing of grammar.

Acknowledgements: The authors thank the children who participated in our study, and their families for their help in arranging and supporting the online sessions. The online games were presented on Gorilla Game Builder with a subscription and tokens funded by the 'Bishop' prize awarded to the first author. This study was funded the German Research Foundation (DFG) [Research Training Group "Dimensions of Constructional Space", project number 468527017], and by an-Alexander von Humboldt Professorship (ID-1195918) awarded to Ewa Dąbrowska.

PART THREE: GENERAL DISCUSSION AND CONCLUSION

5.0 Discussion

This doctoral programme of research examined the cognitive predictors of individual differences in language acquisition in individuals with and without DLD. As discussed in the introductory section, the goal of this research was to develop a greater understanding regarding the interrelation between cognition and language in DLD. To address this aim, a pilot study and three empirical studies were conducted: two exploring cognition and language in children (encompassing two different age groups; study one: mean age of 8;1 years, and study three: mean age of 10;8 years), and one study investigating adult populations (study two). This section outlines the key findings from each of the studies that comprise this thesis, and then discusses clinical implications, possible limitations and recommendations for future research.

5.1 Summary of studies

5.1.1 Pilot study: Dąbrowska and Blake (2022)

The pilot study examined the validity of the MToH task by exploring automatization and grammatical proficiency in L2 learners (Dąbrowska & Blake, 2022). We examined the relationship between the MToH and language aptitude, and two grammar tasks. These were picture selection task (measuring features of 'functional' grammar, where grammar is essential to derive meaning), and a grammaticality judgement task (to measure aspects of 'decorative grammar', where the role of grammar is potentially surplus to meaning, i.e., the sentence can be understood even if is ungrammatical). It was predicted that language aptitude would correlate with performance on functional grammar, and automatization with decorative grammar. Results showed that language aptitude was related to performance in both grammar

tasks. Automatization was not related to functional grammar but a significant relationship was found between performance in the interference task of the MToH and the grammaticality judgement task (decorative grammar). Most importantly, the results from this study determined the use of the MToH paradigm for the empirical studies in this thesis.

5.1.2 Study one: Blake, Dąbrowska, and Riches (1a and 1b: under review)

Study one examined the relationship between language and cognition in children with and without DLD. Four cognitive abilities were examined in this study: nonverbal IQ, working memory, implicit statistical learning, and the speed of automatization. Two papers discuss the outcomes of this research; the first paper (a) examines language development in LT children (Blake, Dąbrowska, & Riches, under review) and the second paper (b), evaluates cognitive predictors of language acquisition in children with and without DLD (Blake, Dąbrowska, & Riches, under review).

Study one (a): Blake, Dąbrowska, and Riches (under review) focused on the relationship between cognition and language in LT children. There results revealed that nonverbal IQ and working memory are important for LT acquisition (with smaller effect sizes for working memory). Two measures, automatization and ISL, did not predict linguistic outcomes. The most significant finding to emerge from this analysis was a strong relationship between nonverbal IQ and receptive grammar, with nonverbal IQ accounting for 39% of the variance in the regression model. Furthermore, simple correlations between nonverbal IQ and grammar were more significant than the relationship observed between nonverbal IQ and vocabulary. According to the literature, there is an established relationship between nonverbal IQ and vocabulary. However, it is commonly held that IQ is not relevant to grammar because grammar is thought to develop implicitly (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001). The results

from this study offer contrasting evidence by showing that nonverbal IQ is related to both vocabulary and grammar in language development.

Study one (b): Blake, Dąbrowska, and Riches (under review) evaluated cognitive predictors of language in children with and without DLD. Consistent with the definition of DLD, the results show that language is the most problematic area for children with DLD, with larger between group differences in linguistic tasks than cognition. Logistic regression revealed that of the four cognitive predictors examined, three hold significance to language acquisition in children with DLD; these are nonverbal IQ, working memory, and the speed of automatization. It is interesting to note a robust relationship between nonverbal IQ and receptive grammar in the DLD group (by comparison, there was no association between nonverbal IQ and receptive vocabulary). This finding suggests that problem solving skills are relevant to grammar in children with DLD. The results from this study suggest that certain cognitive abilities could act as a protective buffer for language development in children with DLD, yet a combination of weaknesses (in this case, nonverbal IQ, working memory and automatization) have an additive effect, which could lead to poor language outcomes. Given the small sample size of DLD participants in our study, this is potentially a bold assumption to make. However, all children in the DLD group showed deficits in working memory in comparison to LT peers, and over 60% of children with DLD presented with a combination of low nonverbal IQ, poor working memory, and slow speed of automatization.

5.1.3 Study two: Blake, Dąbrowska, & Llompart (under review)

The aim of this study was to investigate cognitive predictors of linguistic abilities in adults with and without DLD. We examined nonverbal IQ, (foreign) language aptitude/language analytic ability, ISL, pseudoword learning, and the speed of automatization. These abilities

were assessed in relationship to performance on two grammar tasks (a grammaticality judgement task and receptive grammar), and vocabulary. The results demonstrate that language remains a significant area of difficulty for people with DLD, continuing into adulthood. There is very little research regarding adults with DLD and this study was novel in that it explored language aptitude, using a grammatical sensitivity task. Results show a relationship between language aptitude and grammar in both groups. There were small to no differences between groups in other cognitive abilities assessed for this study (nonverbal IQ, ISL, pseudoword learning and automatization). These findings are in line with prior research which demonstrates a changeable pattern of nonverbal IQ in adults with DLD (Botting, 2020; Clegg et al., 2005; Tomblin et al., 2007, but see Lancaster et al., 2024). Interestingly, nonverbal IQ and language aptitude correlated in the LT group, but not in the DLD group, indicating that IQ does not 'translate' into language learning ability in adults with DLD. We suggest that this is potentially due to a bottleneck in phonological encoding, consistent with the work of McGregor et al. (McGregor et al., 2017; McGregor et al., 2013, 2017, 2020). Lastly, in the DLD group, slower response times were evident across language and cognitive tasks, indicative of slow processing speeds in adults with DLD (Miller & Poll, 2009; Poll et al., 2014, 2016; Zapparrata et al., 2023).

5.1.4 Study three: Blake, Wright, Dąbrowska, Riches, and Winckel (in preparation)

This study explored non-verbal analogical reasoning, language aptitude (using a grammatical sensitivity task for children), inhibition, and ISL in children with and without DLD. Consistent with the previous two studies in this thesis, the results showed that language is the most problematic area for children with DLD, with large between group differences in vocabulary, grammar and collocations. Between group differences were also evident in analogical reasoning (LT children performed at ceiling in this task). Analogical reasoning was significant

for receptive grammar in children with DLD (explaining 48% of variance in the regression model). Therefore, the results show that even though children with DLD perform below their peers in analogical reasoning, it is key to their processing of receptive grammar. Contrary to our hypothesis, inhibition did not affect analogical reasoning performance in either group. In LT children, language aptitude (grammatical sensitivity) was the most significant predictor of all linguistic outcomes - the language aptitude task used in this study requires analogy and so essentially it is a verbal analogical reasoning task. Performance in the ISL task revealed similar sequence learning in both groups (yet slower response times in the DLD group). Collectively, the results from this study demonstrate that analogical reasoning is key to language outcomes in children with and without DLD.

5.2 Overall summary of research findings

There are several general conclusions that emerge from this PhD thesis. Firstly, the results from three empirical studies show that the fundamental difficulty in DLD is language. This is not unexpected as the very definition of DLD relates to weak language skills. The strength of this thesis is that it investigates language in both children and adults, with results showing that language difficulties in DLD persist beyond childhood and continue into adulthood.

The overall results demonstrate that specific cognitive abilities are relevant for language. In LT children (study one: a), these are nonverbal IQ and working memory. In children with DLD, a cumulative risk model is presented (study one: b), comprising of nonverbal IQ, working memory and speed of automatization. In addition, the research findings from study three show that analogical reasoning predicts language in children with and without DLD. Exploration of cognitive predictors in adults with DLD reveal a different picture; language aptitude is significant to grammar in adults with and without DLD. We report similar performance in

LT/DLD groups in the other cognitive abilities assessed in this study, consistent with prior research which report changing patterns of nonverbal IQ in adults with DLD (Botting, 2020; Clegg et al., 2005). Lastly, we observe that adults with DLD require more time to process linguistic and non-linguistic tasks, this too is in keeping with the literature (see meta-analysis by Zapparrata et al., 2023).

A key finding that emerges from this thesis is a strong relationship between nonverbal IQ and grammar across the first two studies. This is significant in both children and adults, with and without DLD, with higher correlations between nonverbal IQ and grammar than those observed between nonverbal IQ and vocabulary. As discussed throughout this thesis, because many researchers consider that grammar develops purely implicitly in child language development (DeKeyser et al., 2010; Ellis, 1996; Ullman, 2001), nonverbal IQ is not typically explored or considered relevant to grammar. Therefore, the findings presented in this thesis offer new insights into the way in which language is processed in LT and DLD populations.

Further substantiation of explicit learning processes in grammar is demonstrated by the significant relationship observed between language aptitude/language analytic ability and grammar in adults with and without DLD (study two), and in LT children (study three). These results are consistent with a body of research which demonstrate that language aptitude is relevant to grammar in L1 speakers (Dąbrowska, 2018; Llompart & Dąbrowska, 2023; Prela et al., 2022; Winckel & Dąbrowska, 2024, Wright et al., 2024). The results reported in this thesis extend these findings by revealing that language aptitude is relevant to grammar in monolingual LT children, and to adults with and without DLD. It would be interesting to assess the relationship between language aptitude and grammar in adolescents with DLD to ascertain whether there are developmental changes between childhood and adulthood.

From a theoretical standpoint, the relationship between explicit and implicit processes in language is in accordance with usage-based models which do not draw a clear distinction between grammar and lexical knowledge (Behrens, 2009; Bybee, 2010). Instead, usage-based approaches propose that explicit and implicit processes are involved in language acquisition, albeit in different stages and to different extents (Dąbrowska, 2010a, 2019), and functioning according to a continuum (Divjak et al., 2022). Our results offer support for the use of explicit, metalinguistic strategies in language interventions for individuals with DLD. As shown, language interventions using explicit processes are beneficial to support areas of potential cognitive weakness (see Balthazar et al., 2020).

The results from this thesis report mixed findings regarding the investigation of implicit statistical learning abilities in children and adults. As several new tasks were designed to assess ISL, discussion is provided for each of the studies in turn. In study one, ISL was measured using an embedded triplet task, based on an original task designed by Arciuli and Simpson (2011, 2011), yet both LT and DLD children performed at chance. Unfortunately, our results are difficult to interpret because the task was potentially too challenging for children. Some children with DLD did not attempt or could not complete the task. Upon reflection, certain aspects of the task could be modified to improve its effectiveness.

Firstly, regarding stimulus design, while we incorporated different patterns on individual stimuli (framed as 'friendly aliens'), the same colours were used for all stimuli. This uniformity may have made it difficult for the children to identify salient features needed to distinguish between the aliens. Secondly, in the original task, the experimenter moderated the test phase; whereas in our study, as the task was performed online, children completed the test phase at their own pace. This may have impacted performance as some children may not have fully

understood what was required of them. Lastly, whilst the number of trials in the familiarisation stage followed the original task design, it may have been too lengthy, possibly causing the children to lose interest by the time they reached the test phase. Future iterations of the task should address these considerations. Additionally, incorporating gamification elements such as rewards and feedback could enhance children's engagement and enjoyment of the task.

Reliability for this task was assessed using the `splithalf` package in R (Parsons, 2021). Based on 5000 random splits, the Spearman-Brown corrected reliability estimate for the LT children in study 1a was 0.52 (95% CI [0.30, 0.68]). In study 1b, the reliability for the LT group was 0.60 (95% CI [0.34, 0.76]), while the reliability for the DLD group was -0.15 (95% CI [-0.60, 0.51]). In the DLD group, the reliability estimates are below acceptable standards which may be indicative of the small sample size and missing data.

Study two investigated ISL in adults with and without DLD using a visual statistical learning task which was modelled upon an original design by Siegelman, Bogaerts, and Frost (2017). In this study, Siegelman et al. revised a previous version of the visual statistical learning task (Siegelman & Frost, 2015), with specific improvements to the test phase including: a) revising the number of trials to include maximal number of trials with fewer repeated patterns, and b) ensuring that there were items of varying difficulty, comprising of pattern recognition trials and pattern completion trials. Their revised task had improved test-rest reliability of .68, in comparison to .58 reported by Siegelman and Frost (2015). Looking at the results from our task, the split-half reliability in our sample was 0.70 (95% CI [0.53, 0.84]) for the DLD group, and 0.86 (95% CI [0.78, 0.93]) in the LT group (30 participants in each sample). It should be noted that in our study, there was a correlation between performance on implicit statistical learning and grammar and vocabulary for both DLD and LT participants. This suggests the

learning that occurred was not purely implicit and raises the possibility of the involvement of explicit processes (see Batterink et al., 2015). Further research using this task could aim to delineate the specific stages at which implicit and explicit processes occur in statistical learning.

Lastly, a new child friendly ISL task was designed for study three. We used a visual search paradigm based on alternating random and predictable sequences of four triplets (in a 3x3 grid). The task was framed as a hide-and-seek game , where children were instructed to click on a specific stimulus (named "Nellar") as quickly as possible. The results from this study showed similar sequence learning and similar RTs between the two groups at the beginning of the task but learning in the LT group became consistently faster over the duration of the task. It is interesting to note that reliability for this task differs between the two groups. The split-half reliability in the LT group was 0.64 (95% CI [0.40, 0.80]), indicating acceptable reliability. In contrast, the DLD group exhibited low reliability, with a split-half reliability estimate of -0.18 (95% CI [-.053, 0.24]). Despite the variability in reliability between the groups, the overall results suggest that the task was generally effective. Further assessments are planned in our lab to evaluate reliability across different populations, comparing performance in children, adolescents, and adults.

As discussed throughout this thesis, concerns regarding reliability and task difficulty are not unusual in the field of ISL and many researchers have raised concerns regarding reliable assessment of ISL (Arciuli, 2017; Arciuli & Conway, 2018; Arnon, 2020; Bogaerts et al., 2021; Siegelman & Frost, 2015). Arnon (2020) reports lower reliability in tasks measuring implicit statistical learning in children, when compared to adults. Therefore, this raises the question of whether differences are related to the individual tasks or if they reflect developmental

differences in implicit statistical learning (see Smalle & Bogaerts, 2024). Another important factor is that most researchers have investigated ISL using tasks designed for group-level analysis, which may not adequately capture fine-grained individual differences in implicit statistical learning (Siegelman et al., 2017). Future research should focus on developing sensitive measures of ISL across both visual and verbal modalities, with an emphasis on investigating individual capacities for ISL. We are aware of current initiatives in this regard which are promising for the field (see Boeve et al., 2023).

The current programme of research is, to the best of the author's knowledge, the first to explore the speed of automatization in DLD using the MToH task. Firstly, a pilot study was conducted to determine the use of the MToH as a measure of automatization, with results showing a relationship between performance on the MToH task and grammar performance in L2 adolescents. In study one, automatization proved significant for children with DLD, but not LT children. In adults with DLD, results from regression analyses revealed an inconsistent pattern, with different measures of automatization relating to grammar, i.e., 'AutoCV' was relevant to receptive grammar in the LT group, and 'AutoDiff' to receptive grammar in the DLD group. These results indicate that the speed of automatization is potentially relevant for linguistic outcomes but suggest that the MToH paradigm requires further adjustment to ensure a more fine-grained measure of automatization.

It should also be taken into account that two different modalities of the MToH were used, the advantages and disadvantages of which are detailed in Blake and Dąbrowska (2024). For example, a physical puzzle was used to assess automatization in the pilot study and in study one with children, whereas an online version of the task was developed for use in adults in study two. The advantages of a tangible, physical puzzle is that it is more engaging. In the pilot

study (with adolescents) and study one (with children), the MToH was moderated by a researcher, providing the opportunity of encouraging participants, helping maintain attention, and celebrating success in solving the puzzle and improving performance. This certainly had a visible impact upon participants' positive engagement with the task. On the other hand, an online puzzle is cost effective as no physical resources are required and participants can complete the puzzle at a time and place that is convenient to them. However, it is not possible to monitor performance in real time and there is the possibility that participants may lose motivation with the repetitive element of the task. As discussed earlier, despite best efforts to include motivational aspects in the online measure of the MToH (study two), it is possible that the online version was less engaging.

Future work could aim to develop the online version of the MToH further by enhancing the motivational element of the task through gamification and incorporating rewards and feedback into the task. Another aspect which requires further evaluation is the number of trials in both the physical and online versions of the task. In the online version of the MToH, we used 25 trials, whereas the physical version included 40 trials for adolescents and 30 trials for children. We opted for fewer trials in the online version to mitigate online fatigue, but further research could evaluate the optimum number of trials for different populations and modalities of the MToH. Additionally, future research could evaluate the secondary task of the MToH. The reverse tapping task was particularly challenging for children with DLD, possibly due to motor difficulties or issues with inhibition (i.e., tapping twice in response to the spoken word 'one' would involve inhibiting the automatic response of tapping once).

Once participants have worked out the optimum solution to the ToH puzzle, their performance usually stabilises. While the number of trials taken to reach an optimum solution

to the puzzle varies per participant, performance is usually stable from the mid-point of the task (although it should be pointed out that not every participant reaches stability). Taking this into account, the reliability of the MToH task was assessed using the split-half method, by dividing the data into odd and even trials. Reliability scores are reported below.

Adults:

Group	Odd trials	95% CI	Even trials	95% CI
LT	0.75	0.64, 0.84	0.87	0.78, 0.93
DLD	0.83	0.68, 0.93	0.81	0.64, 0.92

Children:

Group	Odd trials	95% CI	Even trials	95% CI
LT (study 1a)	0.64	0.49, 0.76	0.62	0.43, 0.76
LT (study 1b)	0.67	0.51, 0.80	0.66	0.46, 0.82
DLD (study 1b)	0.53	0.19, 0.80	0.54	0.28, 0.74

These results indicate that the MToH task shows acceptable reliability in children, using a physical puzzle (scores between 0.50 and 0.70), and good reliability overall in adults with the online version of the task (scores between 0.70 - +0.80).

Therefore, the results reported in this thesis indicate that the MToH is a promising paradigm to assess automatization of linguistic outcomes in people with DLD. As discussed above, research is recommended to fine-tune assessment methods but these results offer a promising starting point for future work.

Lastly, this discussion section concludes by considering the overall aim of this thesis, which set out to explore the relationship between cognition and language in individuals with and

without DLD. Although a broad and ambitious aim, this research question was addressed by assessing well-known areas of cognition, such as nonverbal intelligence, working memory, and implicit statistical learning. In addition, this thesis investigated abilities that are less well-explored, especially in DLD, for example, analogical reasoning, language aptitude/analytic ability and speed of automatization.

The overarching question is how certain aspects of cognition are related to language? In early development, a child's nonverbal cognitive abilities drive their ability to learn language, and many of these abilities are assessed in this thesis – for example, nonverbal intelligence or problem-solving abilities, working memory, and implicit statistical learning. Later on in development, other aspects have an impact upon language, such as quality and quantity of language input from caregivers, teachers and peers, as well as the child's environment and their education (for example, a language rich curriculum with good quality phonics instruction in the early stages of teaching to aid the development of reading skills). However, the relationship between cognition and language is complex and possibly bidirectional. For example, the developmental theory of mutualism suggests that cognitive abilities and language have a positive, reciprocal relationship (see Griffiths et al., 2022; Kievit, 2020). This theory proposes that strong nonverbal cognitive skills could support language development through problem solving abilities and conceptual understanding. Similarly, good language skills would most certainly enhance nonverbal reasoning skills by bringing more sophisticated understanding to a given task. If we consider that language is inherently pattern based, then a child that has strong pattern recognition skills coupled with a good working memory has a distinct advantage when compared to a child who may have weaknesses in these areas.

The tasks used to assess many areas of cognition are rarely 'task pure'. Aside from the specific cognitive ability that they are designed to assess, many other factors are involved, such as attention and motivation as well as the individual's ability to understand a task at a given moment in time. These aspects can vary on a given day and at an individual level, especially when considering the heterogeneity that typically occurs in people with DLD. Furthermore, even when a task is predominantly nonverbal, it often includes elements of language, whether written or spoken. As a result, there is a possibility of a participant not fully understanding the instructions or the requirements of a task, which could impact their overall performance. These areas require careful consideration, especially when working with people with language difficulties. The empirical studies that were conducted to investigate the aims of this thesis used nonverbal tasks where possible, and studies involving children were moderated to ensure that children understood the tasks and could ask questions throughout the research sessions. This enabled the researcher to monitor participant engagement and where necessary, intervene to assess understanding, or try and encourage engagement.

However, no task is infallible and task-related variables may influence performance outcomes, these factors should be taken into consideration in interpreting overall findings. In evaluating the results from the various studies included in this thesis, consideration is given to unaccounted variance which may introduce 'noise' into the data. This could be due to environmental influences or other aspects such as measurement error, heterogeneity of DLD participants, and small sample sizes. As it is difficult to recruit participants with DLD, many of the studies in this thesis are slightly underpowered when considering the number of predictors that have been evaluated in each study. Another reason for potential noise could be attributed to group-level designs, where comparison is made between LT and DLD participants, which may not adequately capture individual differences. While group level

designs may show trends which suggest that certain aspects of language are challenging, the underlying reasons can vary greatly at an individual level (for example, working memory, processing speed, and many other aspects).

Future studies focusing on individual differences provide the opportunity to assess unique variance by considering certain aspects that can be overlooked in group-based study designs. There are several different methods for consideration. For example, conducting sub-group analyses could account for specific strengths and weaknesses in certain areas of language and cognition. Lancaster and Camarata (2019), propose conceptualizing DLD according to a continuum, similar to the approach adopted in ASD. Lastly, investigation of DLD according to a transdiagnostic approach investigates the underlying mechanisms and symptoms across neurodevelopmental conditions rather than assessment of distinct conditions, offering a more child-centred perspective. This is particularly promising given the comorbidity of DLD with conditions such as ADHD and dyslexia (see Astle et al., 2022).

5.3 Clinical implications

The ramifications and challenges of language difficulties in everyday life cannot be underestimated. It goes without saying that the aim of researching DLD is not only of theoretical importance, to inform our understanding regarding the cognitive abilities that underlie language in DLD, but also to transfer this knowledge into practical approaches in clinical interventions. However, it is concerning that it takes a considerable amount of time to integrate evidence from research into clinical practice – on average, this takes approximately 17 years (Balas & Boren, 2000; Grant, Green, & Mason, 2003). Recent research, however, provides promising recommendations for clinical support, as detailed below.

In relation to automatization, Kamhi (2019) outlines recommendations for therapeutic interventions by suggesting several areas of speech and language that could be proceduralized/automatized, arising from the perspective of skill learning. These include certain language structures that could be practiced, for example, younger children benefit from learning simple expressions that can be routinized, and leading on from this, interventions could include more detailed narrative and expository discourses. With this in mind, the results from this thesis are of importance in creating a knowledge base regarding the process of automatization in DLD, with the view of further research to ascertain how automatization could be incorporated into intervention support.

Furthermore, this thesis reports evidence of the interaction between implicit and explicit processes in language. This finding is evident in children and adults, both with and without DLD. This is interesting, not only from the perspective of typical language development, but also in the context of informing intervention support in DLD. While many interventions for DLD are designed to support implicit learning, a growing body of research highlights the importance of integrating explicit processes in intervention (Balthazar et al., 2020; Calder, Claessen, & Leitão, 2018, Calder, Claessen, Ebbels, & Leitão, 2020, 2021; Ebbels, 2007, 2014; Ebbels , Maric, Murphy, & Turner, 2014, Ebbels et al., 2024; Finestack, Engman, Huang, Bangert, & Bader, 2020; Finestack, 2018; Finestack & Fey, 2009; Kamhi, 2019). Therefore, the results reported in this thesis add support to this knowledge base. Aside from these recent studies, it is surprising to find that there is very little evidence relating to intervention research. A forum dedicated to intervention in DLD finds that as few as 4% of articles focused on treatment research in the past five years (this figure is based upon two journals: ASHAWire and PsycInfo) (Montgomery & Gillam, 2024). However, that said, it is promising to see that this forum reports several new studies on intervention in DLD, namely, vocabulary (Bedore et

al., 2024; Camarata, Stiles, & Birer, 2024; Gordon et al., 2024), grammatical morphology (Finestack, Ancel, Lee, Kuchler, & Kornelis, 2024; Leonard et al., 2024); syntax (Balthazar & Scott, 2024; Montgomery et al., 2024; Tarvainen, Launonen, & Stolt, 2021), naturalistic interventions (Camarata et al., 2024), and expository discourse (Ukrainetz, 2024). It is beyond the scope of this concluding section to provide an overview of these papers; however, it is interesting to note that the approaches in these studies focus on either implicit or explicit measures of intervention, or a combination of both.

Another encouraging development is an initiative led by Frizelle et al. (2023), who provide a call for international guidelines to inform intervention research in DLD. Their proposed consensus study will incorporate knowledge from people with DLD and their families, speech and language therapists, academics, and the education community. The CATALISE convention had enormous impact in terms of informing the diagnostic profile and criteria of DLD (Bishop et al., 2016, 2017), and similarly, this call for action promises benefits for individuals with DLD.

5.4 Strengths and limitations

Research regarding DLD is predominantly focused on children and although research regarding adult populations is growing, this remains an under-researched area. The strength of this thesis is that it provides new theoretical knowledge regarding DLD in both child and adult populations. To the author's knowledge, the research reported in this thesis is the first to explore the speed of automatization using the MToH paradigm in people with and without DLD. The results demonstrate the relevance of automatization to language outcomes in children with DLD, and, to a lesser extent, in adults with and without DLD. Another novel approach reported in this thesis is the role of language aptitude/language analytic ability in

children and adults with and without DLD, using a metalinguistic grammatical sensitivity task. Language aptitude is usually assessed in the context of L2 research and therefore, the results reported here offer unique theoretical contributions to L1 research and to knowledge regarding DLD.

Recruiting participants with DLD can be very difficult due to lack of awareness of the condition. This is challenging in child populations, but it is even more difficult to recruit adults with language difficulties to studies, partly due to lack of awareness regarding DLD, and also because many adults are undiagnosed and perhaps unaware of the underlying reasons for their language difficulties. A variety of recruitment methods were used to recruit participants to the studies reported in this thesis. For children this involved contacting schools and organisations specialising in DLD, speech and language therapists, as well as sharing research study information with families through online platforms. A positive aspect of online recruitment is that it is possible to reach a wider audience. On the other hand, online recruitment runs the risk of recruiting WEIRD (Western, educated, industrialized, rich, and democratic) participants (Henrich, Heine, & Norenzayan, 2010; Rad, Martingano, & Ginges, 2018). In this regard, there is concern that research studies tend to recruit participants who are more advantaged in the general population and, although unintentional, researchers may not reach participants who require intervention the most (Frizelle et al., 2023). With these issues in mind, it could be suggested that the research findings presented in this thesis under-represent the needs of children and adults with DLD. Additionally, as is often the case with DLD research, the sample sizes of DLD groups in this thesis are small. This highlights the importance of continued and targeted research to continue to raise awareness and increase knowledge of difficulties faced by people DLD, not only in children but also in adults, who remain under-represented in DLD research.

As discussed, the studies in this thesis were conducted online. For the first study, this was necessitated by the COVID-19 pandemic. However, after the success of this method, studies two and three also used online methods of data collection. Online sessions with children were moderated and guided by a researcher on Zoom, and the second study involving adult participants was run on Gorilla Experiment Builder with online instructions (although the researcher was available to answer questions via email at all times). Online methods have strengths and weaknesses (see Blake & Dąbrowska, 2024; Rodd, 2023). While online studies are particularly beneficial in reaching a wider audience for participation, Rodd (2023) raises certain factors which should be considered. For example, issues such as motivation, the possibility of misunderstanding task instructions, and potential confidence issues, i.e., when the participant is unsure whether or not they are performing the task correctly.

5.5 Conclusion

The overall results from this programme of empirical research contribute to previous knowledge regarding the interrelation between cognitive abilities and linguistic outcomes in children and adults with and without DLD.

This thesis provides evidence that the primary area of difficulty for people with DLD is language, with difficulties persisting into adulthood. The overall results show that specific cognitive abilities are important to support language in children with DLD, these are nonverbal IQ, working memory, automatization, and analogical reasoning. Yet we see a changing picture of cognition in adults with DLD to that seen in children; aside from language aptitude/language analytic ability, performance on other cognitive abilities do not differ significantly from LT adults. This is a growing area of research but there is still a great deal to be learned, and so these results offer important, novel findings to existing DLD research.

There are two original aspects of this PhD research. Firstly, to the best of the authors knowledge, this research is the first to demonstrate a relationship between language aptitude/analytic ability and language in adults with and without DLD, and in LT children (grammar, vocabulary, and collocational knowledge). These findings support the involvement of implicit and explicit processes in language development, adding to current knowledge regarding the underlying processes that support language. Secondly, this thesis discusses the role of automatization in linguistic outcomes. Whilst the measures taken to explore automatization in this thesis are exploratory, they demonstrate that the MToH is a promising paradigm to assess the speed of automatization in language learning. This is the first study to investigate the speed of automatization in DLD using the MToH, and the results described here provide a good foundation for future investigation with larger sample sizes. We recommend

that future studies explore possible adjustments to the measures of automatization described in this thesis, to increase sensitivity and reliability of the MToH paradigm.

This PhD thesis provides new knowledge regarding specific cognitive predictors that predict individual differences in language acquisition in children and adults, with and without DLD. It is hoped that further investigation of the research findings presented here will be undertaken in the future to help to improve the day to day lives of people with DLD.

References

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, *77*(5), 598–614. https://doi.org/10.1037/0021-9010.77.5.598

Ackerman, P. L., Beier, M. E., & Boyle, M. D. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*(4), 567–589. https://doi.org/10.1037/0096-3445.131.4.567

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

Adams, C. (2002). Practitioner Review: The assessment of language pragmatics. *Journal of Child Psychology and Psychiatry*, *43*(8), 973–987. https://doi.org/10.1111/1469-7610.00226

Adams, A., & Gathercole, Susan. E. (2000). Limitations In Working Memory: Implications For Language Development. *International Journal of Language & Communication Disorders*, *35*(1), 95–116. https://doi.org/10.1080/136828200247278

Adams, A. M., & Gathercole, S. E. (1995). Phonological Working Memory and Speech Production in Preschool Children. *Journal of Speech, Language, and Hearing Research*, *38*(2), 403–414. https://doi.org/10.1044/jshr.3802.403

Adlof, S. M., & Hogan, T. P. (2019). If We Don't Look, We Won't See: Measuring Language Development to Inform Literacy Instruction. *Policy Insights from the Behavioral and Brain Sciences*, *6*(2), 210–217. https://doi.org/10.1177/2372732219839075

Adlof, S. M., Scoggins, J., Brazendale, A., Babb, S., & Petscher, Y. (2017). Identifying Children at Risk for Language Impairment or Dyslexia With Group-Administered Measures. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3507–3522. https://doi.org/10.1044/2017_JSLHR-L-16-0473

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, *12*(10), 366–375. https://doi.org/10.1016/0166-2236(89)90074-X

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

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. https://doi.org/10.1177/0022219408315815

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The Cognitive and Behavioral Characteristics of Children With Low Working Memory. *Child Development*, *80*(2), 606–621. https://doi.org/10.1111/j.1467-8624.2009.01282.x

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

Alt, M., & Plante, E. (2006). Factors That Influence Lexical and Semantic Fast Mapping of Young Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *49*(5), 941–954. https://doi.org/10.1044/1092-4388(2006/068)

Alt, M., Plante, E., & Creusere, M. (2004). Semantic Features in Fast-Mapping: Performance of Preschoolers With Specific Language Impairment Versus Preschoolers With

Normal Language. *Journal of Speech, Language, and Hearing Research*, *47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (Ed.). (1993). *Rules of the mind*. L. Erlbaum Associates.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195324259.001.0001

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., Greeno, J. G., Reder, L. M., & Simon, H. A. (2000). Perspectives on Learning, Thinking, and Activity. *Educational Researcher*, *29*(4), 11–13. https://doi.org/10.3102/0013189X029004011

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology*, *45*(2), 153–219. https://doi.org/10.1016/S0010-0285(02)00002-6

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Archibald, L. M. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy*, *33*(1), 5–17. https://doi.org/10.1177/0265659016654206

Archibald, L. M. D., & Gathercole, S. E. (2006a). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675–693. https://doi.org/10.1080/13682820500442602

Archibald, L. M. D., & Gathercole, S. E. (2006b). Visuospatial Immediate Memory in Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *49*(2), 265–277. https://doi.org/10.1044/1092-4388(2006/022)

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, *14*(5), 919–924. https://doi.org/10.3758/BF03194122

Archibald, L. M. D., & Joanisse, M. F. (2009). On the Sensitivity and Specificity of Nonword Repetition and Sentence Recall to Language and Memory Impairments in Children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Archibald, L. M. D., Oram Cardy, J., Joanisse, M. F., & Ansari, D. (2013). Language, Reading, and Math Learning Profiles in an Epidemiological Sample of School Age Children. *PLoS ONE*, *8*(10), e77463. https://doi.org/10.1371/journal.pone.0077463

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160058. https://doi.org/10.1098/rstb.2016.0058

Arciuli, J., & Conway, C. M. (2018). The Promise—And Challenge—Of Statistical Learning for Elucidating Atypical Language Development. *Current Directions in Psychological Science*, *27*(6), 492–500. https://doi.org/10.1177/0963721418779977

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation: Statistical learning. *Developmental Science*, *14*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults: Cognitive Science. *Cognitive Science*, *36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arnon, I. (2019). Statistical Learning, Implicit Learning, and First Language Acquisition: A Critical Evaluation of Two Developmental Predictions. *Topics in Cognitive Science*, tops.12428. https://doi.org/10.1111/tops.12428

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. https://doi.org/10.3758/s13428-019-01205-5

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2), e1373. https://doi.org/10.1002/wcs.1373

Astle, D., Bassett, D. S., & Viding, E. (2022). *Capturing developmental dynamics within a transdiagnostic framework: Challenges and promises*. https://doi.org/10.31234/osf.io/jfpys

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, *63*(4), 397–417. https://doi.org/10.1111/jcpp.13481

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford University Press.

Baddeley, A. (1996). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5–28. https://doi.org/10.1080/713755608

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Hitch, G., & Allen, R. (2021). A multicomponent model of working memory. In Logie, R. H., Camos, V., & Cowan, N. (Eds.). (2021). *Working memory: State of the science* (First edition). (pp. 10–43). Oxford University Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Balas, E. A., & Boren, S. A. (2000). Managing Clinical Knowledge for Health Care Improvement. *Yearbook of Medical Informatics*, *1*, 65–70.

Balthazar, C. H., Ebbels, S., & Zwitserlood, R. (2020). Explicit Grammatical Intervention for Developmental Language Disorder: Three Approaches. *Language, Speech, and Hearing Services in Schools*, *51*(2), 226–246. https://doi.org/10.1044/2019_LSHSS-19-00046

Balthazar, C. H., & Scott, C. M. (2024). Sentences Are Key: Helping School-Age Children and Adolescents Build Sentence Skills Needed for Real Language. *American Journal of Speech-Language Pathology*, *33*(2), 564–579. https://doi.org/10.1044/2023_AJSLP-23-00038

Bates, E., Bretherton, I., & Snyder, L. S. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Beaunieux, H., Hubert, V., Witkowski, T., Pitel, A.-L., Rossi, S., Danion, J.-M., Desgranges, B., & Eustache, F. (2006). Which processes are involved in cognitive procedural learning? *Memory*, *14*(5), 521–539. https://doi.org/10.1080/09658210500477766

Becker, T. C., & McGregor, K. K. (2016). Learning by listening to lectures is a challenge for college students with developmental language impairment. *Journal of Communication Disorders*, *64*, 32–44. https://doi.org/10.1016/j.jcomdis.2016.09.001

Bedore, L. M., Peña, E. D., Shivabasappa, P., Perez, C., Vargas, A. G., Leachman, M., & Wang, J. (2024). An Exploration of the Nature of Vocabulary Change Following Language and Literacy Together Intervention. *American Journal of Speech-Language Pathology*, *33*(2), 611–626. https://doi.org/10.1044/2023_AJSLP-23-00042

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, *47*(2). https://doi.org/10.1515/LING.2009.014

Behrens, H. (2017). The Role of Analogy in Language Processing and Acquisition. In M. Hundt, Mollin, S., & S.E. Pfenninger (Eds.), *The Changing English Language* (1st ed., pp. 215–239). Cambridge University Press. https://doi.org/10.1017/9781316091746.010

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody Picture Vocabulary Test—Third Edition and Wechsler Adult Intelligence Scale—Third Edition with university students. *Journal of Clinical Psychology*, *57*(3), 417–422. https://doi.org/10.1002/jclp.1024

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). I. The Neurocognitive Profile of Williams Syndrome: A Complex Pattern of Strengths and

Weaknesses. *Journal of Cognitive Neuroscience*, *12*(Supplement 1), 7–29. https://doi.org/10.1162/089892900561959

Bishop, D. V. M. (2003). *Test for reception of grammar: TROG-2: manual* (version 2). Pearson.

Bishop, D. V. M. (2006). What Causes Specific Language Impairment in Children? *Current Directions in Psychological Science*, *15*(5), 217–221. https://doi.org/10.1111/j.1467-8721.2006.00439.x

Bishop, D. V. M. (2010). Which Neurodevelopmental Disorders Get Researched and Why? *PLoS ONE*, *5*(11), e15112. https://doi.org/10.1371/journal.pone.0015112

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD): Terminology and the case of DLD. *International Journal of Language & Communication Disorders*, *52*(6), 671–680. https://doi.org/10.1111/1460-6984.12335

Bishop, D. V. M., & Edmundson, A. (1987). Language-Impaired 4-Year-Olds: Distinguishing Transient from Persistent Impairment. *Journal of Speech and Hearing Disorders*, *52*(2), 156–173. https://doi.org/10.1044/jshd.5202.156

Bishop, D. V. M., & Hayiou-Thomas, M. E. (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain and Behavior*, *7*(3), 365–372. https://doi.org/10.1111/j.1601-183X.2007.00360.x

Bishop, D. V. M., 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. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Developmental Medicine & Child Neurology*, *37*(1), 56–71. https://doi.org/10.1111/j.1469-8749.1995.tb11932.x

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin*, *130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & and the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Blake, A., & Dąbrowska, E. (2024). Investigating the relationship between the speed of automatization and linguistic abilities: Data collection during the COVID-19 pandemic. *Linguistics Vanguard*, *0*(0). https://doi.org/10.1515/lingvan-2021-0145

Bliss, L. S. (2002). *Discourse impairments: Assessment and intervention applications* (1st ed). Allyn and Bacon.

Boerma, T., ter Haar, S., Ganga, R., Wijnen, F., Blom, E., & Wierenga, C. J. (2023). What risk factors for Developmental Language Disorder can tell us about the neurobiological mechanisms of language development. *Neuroscience & Biobehavioral Reviews*, *154*, 105398. https://doi.org/10.1016/j.neubiorev.2023.105398

Boersma, P. & Weenink, D. (2022). *Praat: doing phonetics by computer* [Computer program]. Retrieved 2022, from http://www.praat.org/

Boeve, S., Zhou, H., & Bogaerts, L. (2023). *A meta-analysis of 97 studies reveals that statistical learning and language ability are only weakly correlated*. [pre-print] https://doi.org/10.31234/osf.io/s8mwv

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2021). Is there such a thing as a 'good statistical learner'? *Trends in Cognitive Sciences*, S1364661321002825. https://doi.org/10.1016/j.tics.2021.10.012

Bogaerts, L., Siegelman, N., & Frost, R. (2021). Statistical Learning and Language Impairments: Toward More Precise Theoretical Accounts. *Perspectives on Psychological Science*, *16*(2), 319–337. https://doi.org/10.1177/1745691620953082

Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry*, *46*(3), 317–326. https://doi.org/10.1111/j.1469-7610.2004.00355.x

Botting, N. (2020). Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, *55*(2), 255–265. https://doi.org/10.1111/1460-6984.12518

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *36*(4), 421–432. https://doi.org/10.1080/13682820110074971

Botting, N., Durkin, K., Toseeb, U., Pickles, A., & Conti-Ramsden, G. (2016). Emotional health, support, and self-efficacy in young adults with a history of language impairment. *British Journal of Developmental Psychology*, *34*(4), 538–554. https://doi.org/10.1111/bjdp.12148

Botting, N., Toseeb, U., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2016). Depression and Anxiety Change from Adolescence to Adulthood in Individuals with and without Language Impairment. *PLoS ONE*, *11*(7), e0156678. https://doi.org/10.1371/journal.pone.0156678

Bull, M, Aylett, M.P. (1988). An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. *In 5th International Conference of Spoken Language Processing (ICSLP'98).* pp.1175-1178) ISCA.

Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, *19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903_3

Brown, R. 1973. *A First Language. The Early Stages*. Cambridge, MA: Harvard University Press.

Bybee, J. (2010). *Language, Usage and Cognition* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Calder, S. D., Brennan-Jones, C. G., Robinson, M., Whitehouse, A., & Hill, E. (2022). The prevalence of and potential risk factors for Developmental Language Disorder at 10 years in the Raine Study. *Journal of Paediatrics and Child Health*, *58*(11), 2044–2050. https://doi.org/10.1111/jpc.16149

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2020). Explicit Grammar Intervention in Young School-Aged Children With Developmental Language Disorder: An Efficacy Study Using Single-Case Experimental Design. *Language, Speech, and Hearing Services in Schools*, *51*(2), 298–316. https://doi.org/10.1044/2019_LSHSS-19-00060

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2021). The Efficacy of an Explicit Intervention Approach to Improve Past Tense Marking for Early School-Age

Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *64*(1), 91–104. https://doi.org/10.1044/2020_JSLHR-20-00132

Calder, S. D., Claessen, M., & Leitão, S. (2018). Combining implicit and explicit intervention approaches to target grammar in young children with Developmental Language Disorder. *Child Language Teaching and Therapy*, *34*(2), 171–189. https://doi.org/10.1177/0265659017735392

Camarata, S., Stiles, S., & Birer, S. (2024). Naturalistic Developmental Behavioral Interventions for Developmental Language Disorder. *American Journal of Speech-Language Pathology*, *33*(2), 627–641. https://doi.org/10.1044/2023_AJSLP-23-00116

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17–29

Carroll, J.B. (1964). The prediction of success in intensive foreign language training. In Robert Glaser (ed.), *Training research and education*, 87–136. Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T.S. Parry & C.W. Stansfield (Eds),. *Language aptitude reconsidered*. (pp.11-29). Englewood Cliffs, NJ: Prentice Hall Regents.

Carroll, J.B., & Sapon, S.M. (2002). *Modern language aptitude test, MLAT*: manual. Second Language Testing.

Carroll, J. B. & Sapon, S.M. (1959). *Modern language aptitude test*. Washington, DC: Second Language Testing Incorporated

Carroll, J. B., & Sapon, S. S. (1967). *The modern language aptitude test – Elementary*. New York: The Psychological Corporation. Operational test (2002, 2010 ed.), Rockville, MD: Language Learning and Testing Foundation.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193. https://doi.org/10.1037/h0059973

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Cattell, R. B., Feingold, S. N., & Sarason, S. B. (1941). A culture-free intelligence test: II. Evaluation of cultural influence on test performance. *Journal of Educational Psychology*, *32*(2), 81–100. https://doi.org/10.1037/h0058456

Chater, N., & Christiansen, M. H. (2018). Language acquisition as skill learning. *Current Opinion in Behavioral Sciences*, *21*, 205–208. https://doi.org/10.1016/j.cobeha.2018.04.001

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254. https://doi.org/10.1016/j.jml.2015.11.004

Chen, X. S., Reader, R. H., Hoischen, A., Veltman, J. A., Simpson, N. H., Francks, C., Newbury, D. F., & Fisher, S. E. (2017). Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Scientific Reports*, *7*(1), 46105. https://doi.org/10.1038/srep46105

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: M. I. T. Press

Chomsky, N. (1975). *Reflections on language*. New York, N.Y. Pantheon Books.

Chomsky, N. (1972). *Language and mind*. New York, N.Y. Harcourt Brace Jovanovich.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Berlin: Walter de Gruyter.

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Cilibrasi, L., Adani, F., & Tsimpli, I. (2019). Reading as a Predictor of Complex Syntax. The Case of Relative Clauses. *Frontiers in Psychology*, *10*, 1450. https://doi.org/10.3389/fpsyg.2019.01450

Clark, G. M., & Lum, J. A. G. (2017). Procedural memory and speed of grammatical processing: Comparison between typically developing children and language impaired children. *Research in Developmental Disabilities*, *71*, 237–247. https://doi.org/10.1016/j.ridd.2017.10.015

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. https://doi.org/10.1111/j.1469-7610.2004.00342.x

Cohen, N., & Squire, L. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207–210. https://doi.org/10.1126/science.7414331

Cole, K. N., Mills, P. E., & Kelley, D. (1994). Agreement of Assessment Profiles Used in Cognitive Referencing. *Language, Speech, and Hearing Services in Schools*, *25*(1), 25–31. https://doi.org/10.1044/0161-1461.2501.25

Conti-Ramsden, G. (2003). Processing and Linguistic Markers in Young Children With Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *46*(5), 1029–1037. https://doi.org/10.1044/1092-4388(2003/082)

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic Markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

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. https://doi.org/10.1044/jslhr.4205.1195

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The Extent to Which Psychometric Tests Differentiate Subgroups of Children With SLI. *Journal of Speech, Language, and Hearing Research*, *40*(4), 765–777. https://doi.org/10.1044/jslhr.4004.765

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, *53*(2), 237–255. https://doi.org/10.1111/1460-6984.12338

Conti-Ramsden, G., St Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental Trajectories of Verbal and Nonverbal Skills in Individuals With a History of Specific Language Impairment: From Childhood to Adolescence. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1716–1735. https://doi.org/10.1044/1092-4388(2012/10-0182)

Conway, C. M., Arciuli, J., Lum, J. A. G., & Ullman, M. T. (2019). Seeing problems that may not exist: A reply to West et al.'s (2018) questioning of the procedural deficit hypothesis. *Developmental Science*, *22*(4), e12814. https://doi.org/10.1111/desc.12814

Conway, C. M., & Christiansen, M. H. (2006). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*(3), 255–265. https://doi.org/10.1016/0028-3932(68)90024-9

Cowan, N. (2005). *Working memory capacity*. Psychology Press.

Cowan, N. (2022). Working memory development: A 50-year assessment of research and underlying theories. *Cognition*, *224*, 105075. https://doi.org/10.1016/j.cognition.2022.105075

Crawley, M. J. (2013). *The R book* (Second edition). Wiley.

Crisafi, M. A., & Brown, A. L. (1986). Analogical Transfer in Very Young Children: Combining Two Separately Learned Solutions to Reach a Goal. *Child Development*, *57*(4), 953. https://doi.org/10.2307/1130371

Curtiss, S. (2012). Revisiting Modularity: Using Language as a Window to the Mind*. In M. Piattelli-Palmarini & R. C. Berwick (Eds.), *Rich Languages From Poor Inputs* (pp. 68–90). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199590339.003.0005

Dąbrowska, E. (2004). *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh University Press Ltd.

Dąbrowska, E. (2009). Words as constructions. In V. Evans & S. Pourcel (Eds.) *New directions in cognitive linguistics* (Vol. 24, pp. 201–223). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.24.16dab

Dąbrowska, E. (2010a). Productivity, Proceduralization and Specific Language Impairment. *Human Development*, *53*(5), 278–286. https://doi.org/10.1159/000321290

Dąbrowska, E. (2010b). Formulas in the acquisition of English interrogatives: A case study. In Danuta Stanulewicz, Tadeusz Z. Wolanski, Joanna Redzimska (Eds.), *Lingua Terra Cognita II: A Festschrift for Roman Kalisz.* (pp. 675-702). Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253. https://doi.org/10.1075/lab.2.3.01

Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, *25*(4), 617–653. https://doi.org/10.1515/cog-2014-0057

Dąbrowska, E. (2015). Individual differences in grammatical knowledge. In Dąbrowska, E. & Divjak, D. (Eds.). *Handbook of Cognitive Linguisticss* (pp. 649-667). De Gruyter Mouton, Berlin. https://doi.org/10.1515/9783110626438-012

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Dąbrowska, E. (2019). Experience, aptitude, and individual differences in linguistic attainment: A comparison of native and nonnative speakers. *Language Learning*, *69*(Suppl 1), 72–100. https://doi.org/10.1111/lang.12323

Dąbrowska, E., Becker, L., & Miorelli, L. (2020). Is Adult Second Language Acquisition Defective? *Frontiers in Psychology*, *11*, 1839. https://doi.org/10.3389/fpsyg.2020.01839

Dąbrowska, E. & Blake, A. (2022). Speed of automatization predicts performance on "decorative" grammar in second language learning. In T. Piske, & A. K. Steinlen (eds.), *Cognition and second language acquisition: Studies on pre-school, primary*

school and secondary school children, pp. 285–309. Tübingen: Narr Francke Attempto.

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cogl*, *16*(3), 437–474. https://doi.org/10.1515/cogl.2005.16.3.437

Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*(6), 604–615. https://doi.org/10.1016/j.langsci.2005.11.014

Dale, P. S., Rice, M. L., Rimfeld, K., & Hayiou-Thomas, M. E. (2018). Grammar Clinical Marker Yields Substantial Heritability for Language Impairments in 16-Year-Old Twins. *Journal of Speech, Language, and Hearing Research*, *61*(1), 66–78. https://doi.org/10.1044/2017_JSLHR-L-16-0364

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*,*17*(4), 367–378. https://doi.org/10.1037/0012-1649.17.4.367

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Deary, I. J., Cox, S. R., & Hill, W. D. (2022). Genetic variation, brain, and intelligence differences. *Molecular Psychiatry*, *27*(1), 335–353. https://doi.org/10.1038/s41380-021-01027-y

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

DeKeyser, R (2015) Automaticity and automatization. In B. VanPatten, & J. Williams. (Eds.). (2015). *Theories in second language acquisition: An introduction* (Second Edition). (pp. 93-112). London: Routledge.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413–438. https://doi.org/10.1017/S0142716410000056

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing Second Language Morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195–221. https://doi.org/10.1017/S0272263197002040

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/S0272263100004022

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (1st ed., pp. 125–151). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.007

Del Tufo, S. N., & Earle, F. S. (2020). Skill Profiles of College Students With a History of Developmental Language Disorder and Developmental Dyslexia. *Journal of Learning Disabilities*, *53*(3), 228–240. https://doi.org/10.1177/0022219420904348

Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with Developmental Language Disorder. *Journal of Child Language*, *47*(3), 600–632. https://doi.org/10.1017/S0305000919000722

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. https://doi.org/10.1080/13803395.2016.1223279

Desmottes, L., Meulemans, T., & Maillart, C. (2016). Later learning stages in procedural memory are impaired in children with Specific Language Impairment. *Research in Developmental Disabilities*, *48*, 53–68. https://doi.org/10.1016/j.ridd.2015.10.010

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic Evidence for Bidirectional Effects of Early Lexical and Grammatical Development. *Child Development*, *74*(2), 394–412. https://doi.org/10.1111/1467-8624.7402005

Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge University Press.

Divjak, D., & Caldwell-Harris, C. L. (2015). Chapter 3: Frequency and entrenchment. In E. Dąbrowska & D. Divjak (Eds.), *Cognitive Linguistics—Foundations of Language* (pp. 61–86). De Gruyter. https://doi.org/10.1515/9783110626476-004

Divjak, D., Milin, P., Medimorec, S., & Borowski, M. (2022). Behavioral Signatures of Memory Resources for Language: Looking beyond the Lexicon/Grammar Divide. *Cognitive Science*, *46*(11), e13206. https://doi.org/10.1111/cogs.13206

Dixon, J. A., & Marchman, V. A. (2007). Grammar and the Lexicon: Developmental Ordering in Language Acquisition. *Child Development*, *78*(1), 190–212. https://doi.org/10.1111/j.1467-8624.2007.00992.x

Dockrell, J. E., Howell, P., Leung, D., & Fugard, A. J. B. (2017). Children with Speech Language and Communication Needs in England: Challenges for Practice. *Frontiers in Education*, *2*, 35. https://doi.org/10.3389/feduc.2017.00035

Dockrell, J. E., Lindsay, G., Letchford, B., & Mackie, C. (2006). Educational provision for children with specific speech and language difficulties: Perspectives of speech and language therapy service managers. *International Journal of Language & Communication Disorders*, *41*(4), 423–440. https://doi.org/10.1080/13682820500442073

Dockrell, J. E., Ricketts, J., Palikara, O., Charman, T., & Lindsay, G. A. (2019). What Drives Educational Support for Children With Developmental Language Disorder or Autism Spectrum Disorder: Needs, or Diagnostic Category? *Frontiers in Education*, *4*, 29. https://doi.org/10.3389/feduc.2019.00029

Dollaghan, C. A. (1987). Fast Mapping in Normal and Language-Impaired Children. *Journal of Speech and Hearing Disorders*, *52*(3), 218–222. https://doi.org/10.1044/jshd.5203.218

Donnarumma, F., Maisto, D., & Pezzulo, G. (2016). Problem Solving as Probabilistic Inference with Subgoaling: Explaining Human Successes and Pitfalls in the Tower of Hanoi. *PLOS Computational Biology*, *12*(4), e1004864. https://doi.org/10.1371/journal.pcbi.1004864

Doust, C., Fontanillas, P., Eising, E., Gordon, S. D., Wang, Z., Alagöz, G., Molz, B., 23andMe Research Team, Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Coker, D., Partida, G. C., Dhamija, D., Das, S., …

Luciano, M. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nature Genetics*, *54*(11), 1621–1629. https://doi.org/10.1038/s41588-022-01192-y

Dubois, P., Guay, F., & St-Pierre, M. C. (2023). Examining the contribution of motivation in the job search of youth with developmental language disorder. *Autism & Developmental Language Impairments*, *8*, 239694152311520. https://doi.org/10.1177/23969415231152094

Dubois, P., St-Pierre, M.-C., Desmarais, C., & Guay, F. (2020). Young Adults With Developmental Language Disorder: A Systematic Review of Education, Employment, and Independent Living Outcomes. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3786–3800. https://doi.org/10.1044/2020_JSLHR-20-00127

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British picture vocabulary scale* (3rd ed). GL Assessment Ltd.

Dunn, L. M., & Dunn, L. M. (2012). *Peabody Picture Vocabulary Test—Third Edition* [Dataset]. https://doi.org/10.1037/t15145-000

Durkin, K., Toseeb, U., Botting, N., Pickles, A., & Conti-Ramsden, G. (2017). Social Confidence in Early Adulthood Among Young People With and Without a History of Language Impairment. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1635–1647. https://doi.org/10.1044/2017_JSLHR-L-16-0256

Earle, F. S., & Ullman, M. T. (2021). Deficits of Learning in Procedural Memory and Consolidation in Declarative Memory in Adults With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *64*(2), 531–541. https://doi.org/10.1044/2020_JSLHR-20-00292

Ebbels, S. (2007). Teaching grammar to school-aged children with specific language impairment using Shape Coding. *Child Language Teaching and Therapy*, *23*(1), 67–93. https://doi.org/10.1191/0265659007072143

Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, *30*(1), 7–40. https://doi.org/10.1177/0265659013512321

Ebbels, S. H., Gadd, M., Nicoll, H., Hughes, L., Dawson, N., Burke, C., Calder, S. D., & Frizelle, P. (2024). The Effectiveness of Individualized Morphosyntactic Target Identification and Explicit Intervention Using the SHAPE CODING System for Children With Developmental Language Disorder and the Impact of Within-Session Dosage. *Language, Speech, and Hearing Services in Schools*, *55*(3), 803–837. https://doi.org/10.1044/2024_LSHSS-23-00098

Ebbels, S. H., Maric, N., Murphy, A., & Turner, G. (2014). Improving comprehension in adolescents with severe receptive language impairments: A randomized control trial of intervention for coordinating conjunctions. *International Journal of Language & Communication Disorders*, *49*(1), 30–48. https://doi.org/10.1111/1460-6984.12047

Eichenbaum, H. (2012). Memory Systems. In I. Weiner (Ed.), *Handbook of Psychology, Second Edition* (1st ed.). Wiley. https://doi.org/10.1002/9781118133880.hop203020

Eising, E., Mirza-Schreiber, N., De Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., Shapland, C. Y., Zhu, G., Wigg, K. G., Gerritse, M. L., Molz, B., Alagöz, G., Gialluisi, A., Abbondanza, F., Rimfeld, K., Van Donkelaar, M., Liao, Z., Jansen, P. R., Andlauer, T. F. M., … Fisher, S. E. (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proceedings of the National Academy of Sciences*, *119*(35), e2202764119. https://doi.org/10.1073/pnas.2202764119

Ellis, N. C. (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition*, *18*(1), 91–126. https://doi.org/10.1017/S0272263100014698

Ellis, N. C. (2002). Frequency effects in language processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. & Reinders, H. (2009_. (Ed.) *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of Construction Grammar*. *Review of Cognitive Linguistics*, *15*(1), 289–296. https://doi.org/10.1075/rcl.15.1.11cad

Ellis Weismer, S., Davidson, M. M., Gangopadhyay, I., Sindberg, H., Roebuck, H., & Kaushanskaya, M. (2017). The role of nonverbal working memory in morphosyntactic processing by children with specific language impairment and autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, *9*(1), 28. https://doi.org/10.1186/s11689-017-9209-6

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

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. https://doi.org/10.1044/1092-4388(2009/07-0189)

Farkas, B. C., Krajcsi, A., Janacsek, K., & Nemeth, D. (2023). The complexity of measuring reliability in learning tasks: An illustration using the Alternating Serial Reaction Time Task. *Behavior Research Methods*, *56*(1), 301–317. https://doi.org/10.3758/s13428-022-02038-5

Fawcett, A. J., & Nicolson, R. I. (1992). Automatisation Deficits in Balance for Dyslexic Children. *Perceptual and Motor Skills*, *75*(2), 507–529. https://doi.org/10.2466/pms.1992.75.2.507

Feldman, J., Kerr, B., & Streissguth, A. (1995). Correlational analyses of procedural and declarative learning performance. *Intelligence*, *20*(1), 87–114. https://doi.org/10.1016/0160-2896(95)90007-1

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Fillmore, C J & Kay P. (1993). *Construction Grammar Course Book*. UC Berkeley: Department of Linguistics.

Fillmore, Charles, J. (1988). *'The mechanisms of construction grammar'. In Annual Meeting of the Berkeley Linguistics Society (pp. 35-55).*

Finestack, L., Engman, J., Huang, T., Bangert, K. J., & Bader, K. (2020). Evaluation of a Combined Explicit–Implicit Approach to Teach Grammatical Forms to Children With Grammatical Weaknesses. *American Journal of Speech-Language Pathology*, *29*(1), 63–79. https://doi.org/10.1044/2019_AJSLP-19-0056

Finestack, L. H. (2018). Evaluation of an Explicit Intervention to Teach Novel Grammatical Forms to Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2062–2075. https://doi.org/10.1044/2018_JSLHR-L-17-0339

Finestack, L. H., Ancel, E., Lee, H., Kuchler, K., & Kornelis, M. (2024). Five Additional Evidence-Based Principles to Facilitate Grammar Development for Children With Developmental Language Disorder. *American Journal of Speech-Language Pathology*, *33*(2), 552–563. https://doi.org/10.1044/2023_AJSLP-23-00049

Finestack, L. H., & Fey, M. E. (2009). Evaluation of a deductive procedure to teach grammatical inflections to children with language impairment. *American Journal of Speech-Language Pathology*, *18*(3), 289–302. https://doi.org/10.1044/1058-0360%282009/08-0041%29

Finlay, J. C. S., & McPhillips, M. (2013). Comorbid motor deficits in a clinical sample of children with specific language impairment. *Research in Developmental Disabilities*, *34*(9), 2533–2542. https://doi.org/10.1016/j.ridd.2013.05.015

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, *18*(2), 168–170. https://doi.org/10.1038/ng0298-168

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning.* (pp. 243-285). London: Academic Press.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Frizelle, P., McKean, C., Eadie, P., Ebbels, S., Fricke, S., Justice, L. M., Kunnari, S., Leitão, S., Morgan, A. T., Munro, N., Murphy, C., Storkel, H. L., & Van Horne, A. O. (2023). Editorial Perspective: Maximising the benefits of intervention research for children and young people with developmental language disorder (DLD) – a call for

international consensus on standards of reporting in intervention studies for children with and at risk for DLD. *Journal of Child Psychology and Psychiatry*, *64*(3), 474–479. https://doi.org/10.1111/jcpp.13694

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*(1), 87–115. https://doi.org/10.1146/annurev.psych.49.1.87

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*(2), 336–343. https://doi.org/10.1017/S1355617710001724

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. https://doi.org/10.1017/S1355617712001270

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. https://doi.org/10.1017/S0142716413000490

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. https://doi.org/10.1044/1058-0360(2012/11-0044)

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*(1), 87–115. https://doi.org/10.1146/annurev.psych.49.1.87

Gallinat, E., & Spaulding, T. J. (2014). Differences in the Performance of Children With Specific Language Impairment and Their Typically Developing Peers on Nonverbal

Cognitive Tests: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1363–1382. https://doi.org/10.1044/2014_JSLHR-L-12-0363

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/S0142716406060383

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. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177–194. https://doi.org/10.1348/000709900158047

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. https://doi.org/10.1002/acp.934

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*(5), 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Gentner, D. (1988). Metaphor as Structure Mapping: The Relational Shift. *Child Development*, *59*(1), 47. https://doi.org/10.2307/1130388

Gentner, D., & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, *2*(2), 261–283. https://doi.org/10.1515/langcog.2010.011

Gentner, D., & Namy, L. L. (2006). Analogical Processes in Language Learning. *Current Directions in Psychological Science*, *15*(6), 297–301. https://doi.org/10.1111/j.1467-8721.2006.00456.x

Gentner, D. & Smith, L. (2012). Analogical reasoning. In V. S. Ramachandran (Ed.) *Encyclopedia of Human Behavior* (2nd Ed.). pp. 130-136. Oxford, UK: Elsevier.

Geurten, M., Catale, C., & Meulemans, T. (2016). Involvement of Executive Functions in Children's Metamemory. *Applied Cognitive Psychology*, *30*(1), 70–80. https://doi.org/10.1002/acp.3168

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants. *Language Learning*, *51*(1), 1–50. https://doi.org/10.1111/1467-9922.00147

Goldwater, M. B. (2017). Grammatical Constructions as Relational Categories. *Topics in Cognitive Science*, *9*(3), 776–799. https://doi.org/10.1111/tops.12272

Gordon, K. R., Storkel, H. L., Lowry, S. L., & Sultani, M. J. (2024). A Word-Learning Intervention Pilot Study Utilizing Principles of Retrieval- and Criterion-Based Learning for Children With Developmental Language Disorder. *American Journal of Speech-Language Pathology*, *33*(2), 530–551. https://doi.org/10.1044/2023_AJSLP-23-00037

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development*, *4*(3), 251–268. https://doi.org/10.1016/0885-2014(89)90008-7

Graybiel, A. M. (2008). Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience*, *31*(1), 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Grant, J., Green, L., & Mason, B. (2003). Basic research and health: A reassessment of the scientific basis for the support of biomedical science. *Research Evaluation*, *12*(3), 217–224. https://doi.org/10.3152/147154403781776618

Gray, S. (2003). Word-Learning by Preschoolers With Specific Language Impairment: What Predicts Success? *Journal of Speech, Language, and Hearing Research*, *46*(1), 56–67. https://doi.org/10.1044/1092-4388(2003/005)

Gray, S., Fox, A.B., Green, S., Alt, M., Hogan, T.P., Petscher, Y, Cowan, N. (2019). Working Memory Profiles of Children With Dyslexia, Developmental Language Disorder, or Both. *Journal of Speech, Language and Hearing Research*, *62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Graybiel, A. M., & Grafton, S. T. (2015). The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021691. https://doi.org/10.1101/cshperspect.a021691

Griffiths, S., Kievit, R. A., & Norbury, C. (2022). Mutualistic coupling of vocabulary and non-verbal reasoning in children with and without language disorder. *Developmental Science*, *25*(3). https://doi.org/10.1111/desc.13208

Grömping, U. (2006). Relative Importance for Linear Regression in *R*: The Package relaimpo. *Journal of Statistical Software*, *17*(1). https://doi.org/10.18637/jss.v017.i01

Guasti, M. T., White, M. J., Bianco, G., Arosio, F., Camilleri, B., & Hasson, N. (2021). Two clinical markers for DLD in monolingual Italian speakers: What can they tell us about second language learners with DLD? *Clinical Linguistics & Phonetics*, *35*(9), 829–846. https://doi.org/10.1080/02699206.2020.1830303

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Development Language Disorder II: A Comparison of Retrieval Schedules. *Journal of Speech, Language, and Hearing Research*, *62*(4), 944–964. https://doi.org/10.1044/2018_JSLHR-L-18-0071

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Hayiou-Thomas, M. E., Oliver, B., & Plomin, R. (2005). Genetic Influences on Specific Versus Nonspecific Language Impairment in 4-Year-Old Twins. *Journal of Learning Disabilities*, *38*(3), 222–232. https://doi.org/10.1177/00222194050380030401

Hayiou-Thomas, M. E., Kovas, Y., Harlaar, N., Plomin, R., Bishop, D. V. M., & Dale, P. S. (2006). Common aetiology for diverse language skills in 41/2-year-old twins. *Journal of Child Language*, *33*(2), 339–368. https://doi.org/10.1017/S0305000906007331

Hayiou-Thomas, M. E., Smith-Woolley, E., & Dale, P. S. (2021). Breadth versus depth: Cumulative risk model and continuous measure prediction of poor language and reading outcomes at 12. *Developmental Science*, *24*(1). https://doi.org/10.1111/desc.12998

Hebb, D. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford, UK: Blackwell.

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Tomblin, B. J., & 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. https://doi.org/10.1016/j.ridd.2011.07.026

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online Lexical Competition During Spoken Word Recognition and Word Learning in Children and Adults. *Child Development*, *84*(5), 1668–1685. https://doi.org/10.1111/cdev.12067

Hendricks, A. E., Adlof, S. M., Alonzo, C. N., Fox, A. B., & Hogan, T. P. (2019). Identifying Children at Risk for Developmental Language Disorder Using a Brief, Whole-Classroom Screen. *Journal of Speech, Language, and Hearing Research*, *62*(4), 896–908. https://doi.org/10.1044/2018_JSLHR-L-18-0093

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henry, L. A., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, *33*(1), 19–32. https://doi.org/10.1177/0265659016655378

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. https://doi.org/10.1080/13682820010019874

Hilpert, M. (2021). *Ten lectures on diachronic construction grammar*. Brill.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Howard, J. H., & Howard, D. V. (2013). Aging mind and brain: Is implicit learning spared in healthy aging? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00817

Hsu, H. J., & Bishop, D. V. M. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352–365. https://doi.org/10.1111/desc.12125

Huang, Y., & Ferreira, F. (2020). The Application of Signal Detection Theory to Acceptability Judgments. *Frontiers in Psychology*, *11*, 73. https://doi.org/10.3389/fpsyg.2020.00073

Hubert, V., Beaunieux, H., Chételat, G., Platel, H., Landeau, B., Danion, J.-M., Viader, F., Desgranges, B., & Francis Eustache. (2007). The dynamic network subserving the three phases of cognitive procedural learning. *Human Brain Mapping*, *28*(12), 1415–1429. https://doi.org/10.1002/hbm.20354

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-native Speakers: Theory and research* (Vol. 41). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.41

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*(4), 555–582. https://doi.org/10.1017/S0142716409990014

Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal*, *54*(4), (pp. 300-309). https://doi.org/10.2307/811114

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & et al. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing Limitations in Children With Specific Language Impairment: The Role of Executive Function. *Child Development*, *77*(6), 1822–1841. https://doi.org/10.1111/j.1467-8624.2006.00976.x

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language: A Meta-Analysis Into 25 Years of Research. *Cognitive Science*, *46*(9). https://doi.org/10.1111/cogs.13198

Isbilen, E. S., Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2022). Statistically based chunking of nonadjacent dependencies. *Journal of Experimental Psychology: General*, *151*(11), 2623–2640. https://doi.org/10.1037/xge0001207

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12848

Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment: STM, receptive vocabulary and fast mapping in children with SLI. *International Journal of Language & Communication Disorders*, *51*(1), 61–73. https://doi.org/10.1111/1460-6984.12185

Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019a). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders*, *79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2019b). The evaluation of word-learning abilities in people with developmental language disorder: A scoping review. *International Journal of Language & Communication Disorders*, *54*(5), 742–755. https://doi.org/10.1111/1460-6984.12490

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2020). Working, Declarative, and Procedural Memory in Children With Developmental Language Disorder. *Journal of*

Speech, Language, and Hearing Research, *63*(12), 4162–4178. https://doi.org/10.1044/2020_JSLHR-20-00135

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2021). Word learning and verbal working memory in children with developmental language disorder. *Autism & Developmental Language Impairments*, *6*, 239694152110041. https://doi.org/10.1177/23969415211004109

James, E., Gaskell, M. G., & Henderson, L. M. (2020). Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: It'll be alright on the night? *Journal of Child Psychology and Psychiatry*, *61*(10), 1104–1115. https://doi.org/10.1111/jcpp.13253

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America*. Basic Books.

Jensen, A. R. (1998). *The g factor: The science of mental ability* (1. publ). Praeger.

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-Year Follow-Up of Children With and Without Speech-Language Impairments: Family, Educational, Occupational, and Quality of Life Outcomes. *American Journal of Speech-Language Pathology*, *19*(1), 51–65. https://doi.org/10.1044/1058-0360(2009/08-0083)

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Johnson, K. (1996). *Language Teaching and Skill Learning*. Wiley.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *37*(2), 418–421. https://doi.org/10.1044/jshr.3702.418

Kail, R. V., & Ferrer, E. (2007). Processing Speed in Childhood and Adolescence: Longitudinal Models for Examining Developmental Change. *Child Development*, *78*(6), 1760–1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

Kamhi, A. G. (2019). Speech-language development as proceduralization and skill learning: Implications for assessment and intervention. *Journal of Communication Disorders*, *82*, 105918. https://doi.org/10.1016/j.jcomdis.2019.105918

Kamhi, A. G. (2020). Response to commentaries. *Journal of Communication Disorders*, *83*, 105952. https://doi.org/10.1016/j.jcomdis.2019.105952

Kan, P. F., & Windsor, J. (2010). Word Learning in Children With Primary Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

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. https://doi.org/10.1016/j.cognition.2010.05.011

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kidd, E., & Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development*, *87*(1), 184–193. https://doi.org/10.1111/cdev.12461

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology*, *200*, 104964. https://doi.org/10.1016/j.jecp.2020.104964

Kidd, E., Arciuli, J., Christiansen, M. H., & Smithson, M. (2023). The sources and consequences of individual differences in statistical learning for language development. *Cognitive Development*, *66*, 101335. https://doi.org/10.1016/j.cogdev.2023.101335

Kidd, E., & Donnelly, S. (2020). Individual Differences in First Language Acquisition. *Annual Review of Linguistics*, *6*(1), 319–340. https://doi.org/10.1146/annurev-linguistics-011619-030326

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

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. https://doi.org/10.1080/01690965.2010.493735

Kievit, R. A. (2020). Sensitive periods in cognitive development: A mutualistic perspective. *Current Opinion in Behavioral Sciences*, *36*, 144–149.https://doi.org/10.1016/j.cobeha.2020.10.007

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106–120. https://doi.org/10.1101/lm.1.2.106

Konopka, G., & Roberts, T. F. (2016). Insights into the Neural and Genetic Basis of Vocal Communication. *Cell*, *164*(6), 1269–1276. https://doi.org/10.1016/j.cell.2016.02.039

Kostadinova, V., & Gardner, M. H. (2024). Getting "good" data in a pandemic, part 1: Assessing the validity and quality of data collected remotely. *Linguistics Vanguard*, *9*(s4), 329–334. https://doi.org/10.1515/lingvan-2023-0170

Krassowski, E., & Plante, E. (1997). IQ variability in children with SLI: Implications for use of cognitive referencing in determining SLI. *Journal of Communication Disorders*, *30*(1), 1–9. https://doi.org/10.1016/0021-9924(95)00052-6

Krishnan, S., Cler, G. J., Smith, H. J., Willis, H. E., Asaridou, S. S., Healy, M. P., Papp, D., & Watkins, K. E. (2022). Quantitative MRI reveals differences in striatal myelin in children with DLD. *eLife*, *11*, e74242. https://doi.org/10.7554/eLife.74242

Krishnan, S., & Watkins, K. E. (2019). A challenge for the procedural deficit hypothesis: How should we measure sequential learning in childhood? *Developmental Science*, e12815. https://doi.org/10.1111/desc.12815

Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2016). Neurobiological Basis of Language Learning Difficulties. *Trends in Cognitive Sciences*, *20*(9), 701–714. https://doi.org/10.1016/j.tics.2016.06.012

Krzemien, M., Jemel, B., & Maillart, C. (2017). Analogical reasoning in children with specific language impairment: Evidence from a scene analogy task. *Clinical Linguistics & Phonetics*, *31*(7–9), 573–588. https://doi.org/10.1080/02699206.2017.1302509

Krzemien, M., Maillart, C., Parisse, C., & Leroy, S. (2019). Impact of processing load on analogical mapping with visual sequences in children with developmental language disorders (DLD). *International Journal of Language & Communication Disorders*, *54*(3), 417–429. https://doi.org/10.1111/1460-6984.12446

Krzemien, M., Thibaut, J.-P., & Maillart, C. (2020). How language and inhibition influence analogical reasoning in children with or without developmental language disorder? *Journal of Clinical and Experimental Neuropsychology*, *42*(1), 76–89. https://doi.org/10.1080/13803395.2019.1676881

Kuhn, M. (2008). Building Predictive Models in *R* Using the caret Package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kuiack, A. K., & Archibald, L. M. D. (2023). Identifying and describing developmental language disorder in children. *International Journal of Language & Communication Disorders*, 1460-6984.12984. https://doi.org/10.1111/1460-6984.12984

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–523. https://doi.org/10.1038/35097076

Lammertink, I., Boersma, P., Rispens, J., & Wijnen, F. (2020). Visual statistical learning in children with and without DLD and its relation to literacy in children with DLD. *Reading and Writing*, *33*(6), 1557–1589. https://doi.org/10.1007/s11145-020-10018-4

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical Learning in Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3474–3486. https://doi.org/10.1044/2017_JSLHR-L-16-0439

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020a). Children With Developmental Language Disorder Have an Auditory Verbal Statistical Learning Deficit: Evidence From an Online Measure. *Language Learning*, *70*(1), 137–178. https://doi.org/10.1111/lang.12373

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020b). Statistical Learning in the Visuomotor Domain and Its Relation to Grammatical Proficiency in Children with and without Developmental Language Disorder: A Conceptual Replication and Meta-Analysis. *Language Learning and Development*, *16*(4), 426–450. https://doi.org/10.1080/15475441.2020.1820340

Lancaster, H. S., & Camarata, S. (2019). Reconceptualizing developmental language disorder as a spectrum disorder: Issues and evidence. *International Journal of Language & Communication Disorders*, *54*(1), 79–94. https://doi.org/10.1111/1460-6984.12433

Lancaster, H. S., Milne, A., Gordon, K., Emerson, S. N., & Selin, C. (2024). *Analyzing the Impact of Explicit and Implicit Cognitive Constructs on NVIQ Test Performance*. [pre-print] https://doi.org/10.31219/osf.io/fusr6

Langacker, R. W. (1987). *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites* Standord, CA:Stanford University Press.

Lany, J. (2014). Judging Words by Their Covers and the Company They Keep: Probabilistic Cues Support Word Learning. *Child Development*, *85*(4), 1727–1739. https://doi.org/10.1111/cdev.12199

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling Developmental Language Difficulties From School Entry Into Adulthood: Literacy, Mental Health, and Employment Outcomes. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Leonard, L. B. (2014a). *Children with specific language impairment*. MIT press.

Leonard, L. B. (2014b). Specific Language Impairment Across Languages. *Child Development Perspectives*, *8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Deevy, P., Bredin-Oja, S. L., & Schroeder, M. L. (2024). Sources of Misinterpretation in the Input and Their Implications for Language Intervention With English-Speaking Children. *American Journal of Speech-Language Pathology*, *33*(2), 598–610. https://doi.org/10.1044/2023_AJSLP-23-00016

Leonard, L. B., Ellis Weismer, S., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of Processing, Working Memory, and Language Impairment in Children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 408–428. https://doi.org/10.1044/1092-4388(2007/029)

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J. B., & Krok, W. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Developmental Language Disorder I: The Benefits of Repeated Retrieval. *Journal of Speech, Language, and Hearing Research*, *62*(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Leonard, L. B., & Kueser, J. B. (2019). Five overarching factors central to grammatical learning and treatment in children with developmental language disorder. *International Journal of Language & Communication Disorders*, *54*(3), 347–361. https://doi.org/10.1111/1460-6984.12456

Leroy, S., Maillart, C., & Parisse, C. (2014). Analogical mapping across modalities in children with specific language impairment (SLI). *Research in Developmental Disabilities*, *35*(9), 2158–2171. https://doi.org/10.1016/j.ridd.2014.05.005

Leroy, S., Parisse, C., & Maillart, C. (2012). Analogical reasoning in children with specific language impairment. *Clinical Linguistics & Phonetics*, *26*(4), 380–396. https://doi.org/10.3109/02699206.2011.641059

Lewis, B. A., & Thompson, L. A. (1992). A Study of Developmental Speech and Language Disorders in Twins. *Journal of Speech, Language, and Hearing Research*, *35*(5), 1086–1094. https://doi.org/10.1044/jshr.3505.1086

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lezak, M. D., Howieson, D.B., Loring, D.W., Hannay, H.J., & Fischer, J.S. (Eds.). (2004). *Neuropsychological assessment* (4th ed). Oxford University Press.

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Special Issue: Synthesizing Research on Form-Focused Instruction: The Complementary Contributions of Narrative Review and Meta-Analysis.*, *36*(3), 385–408.

Li, S. (2016). The construct validity of language aptitude: : A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, *8*(3), 346–368. https://doi.org/10.1017/langcog.2016.16

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3). https://doi.org/10.1515/COGL.2009.022

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, *24*(1), 187–219. https://doi.org/10.1017/S0305000996002930

Lin, Y.-P., Shi, Y., Zhang, R., Xue, X., Rao, S., Yin, L., Lui, K. F. H., Pan, D. J., Maurer, U., Choy, K.-W., Paracchini, S., McBride, C., & So, H.-C. (2024). A genome-wide association study of Chinese and English language phenotypes in Hong Kong Chinese children. *Npj Science of Learning*, *9*(1), 26. https://doi.org/10.1038/s41539-024-00229-7

Llompart, M., & Dąbrowska, E. (2020). Explicit but Not Implicit Memory Predicts Ultimate Attainment in the Native Language. *Frontiers in Psychology*, *11*, 569586. https://doi.org/10.3389/fpsyg.2020.569586

Llompart, M., & Dąbrowska, E. (2023). "Foreign" language aptitude predicts individual differences in native grammatical proficiency. *Linguistics*, *0*(0). https://doi.org/10.1515/ling-2022-0009

Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 1040–1056. https://doi.org/10.1037/xhp0000383

Llompart, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1590–1610. https://doi.org/10.1037/xlm0000832

Llompart, M., & Reinisch, E. (2021). Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics*, *88*, 101080. https://doi.org/10.1016/j.wocn.2021.101080

Long, M. (2013). Maturational constraints of child and adult SLA. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude and ultimate L2 attainment* (pp. 3–41). John Benjamins.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lukács, Á., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, *28*(3), 472–483. https://doi.org/10.1037/neu0000052

Lukács, Á., & Kemény, F. (2015). Development of Different Forms of Skill Learning Throughout the Lifespan. *Cognitive Science*, *39*(2), 383–404. https://doi.org/10.1111/cogs.12143

Lukács, Á., Lukics, K. S., & Dobó, D. (2021). Online Statistical Learning in Developmental Language Disorder. *Frontiers in Human Neuroscience*, *15*, 715818. https://doi.org/10.3389/fnhum.2021.715818

Lum, J. A. G., & Bleses, D. (2012). Declarative and procedural memory in Danish speaking children with specific language impairment. *Journal of Communication Disorders*, *45*(1), 46–58. https://doi.org/10.1016/j.jcomdis.2011.09.001

Lum, J. A. G., & Clark, G. M. (2022). Implicit manual and oculomotor sequence learning in developmental language disorder. *Developmental Science*, *25*(2). https://doi.org/10.1111/desc.13156

Lum, J. A. G., 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. https://doi.org/10.1016/j.cortex.2013.10.011

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154. https://doi.org/10.1016/j.cortex.2011.06.001

Lum, J. A. G., 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. https://doi.org/10.3109/13682820902752285

Lundine, J. P., & McCauley, R. J. (2016). A Tutorial on Expository Discourse: Structure, Development, and Disorders in Children and Adolescents. *American Journal of Speech-Language Pathology*, *25*(3), 306–320. https://doi.org/10.1044/2016_AJSLP-14-0130

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Fromm, D. (2016). Child Language Data Exchange System Tools for Clinical Analysis. *Seminars in Speech and Language*, *37*(02), 063–073. https://doi.org/10.1055/s-0036-1580743

Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*(2), 339–366. https://doi.org/10.1017/S0305000900009302

Marengo, L. (2015). Representation, search, and the evolution of routines in problem solving. *Industrial and Corporate Change*, *24*(5), 951–980. https://doi.org/10.1093/icc/dtu023

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology*, *38*(4), 465–494. https://doi.org/10.1006/cogp.1999.0721

Mayer, M. (1969). *Frog, where are you?* Dial Books for Young Readers.

Mayes, A. K., Reilly, S., & Morgan, A. T. (2015). Neural correlates of childhood language disorder: A systematic review. *Developmental Medicine & Child Neurology*, *57*(8), 706–717. https://doi.org/10.1111/dmcn.12714

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. https://doi.org/10.1080/09297049.2012.734293

McCauley, S. M. (2020). Towards an integrated, single-system account of language development as skill learning. *Journal of Communication Disorders*, *83*, 105942. https://doi.org/10.1016/j.jcomdis.2019.105942

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

McGregor, K., Arbisi-Kelm, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*, *19*(1), 43–57. https://doi.org/10.3109/17549507.2016.1159337

McGregor, K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism & Developmental Language Impairments*, *5*, 239694151989931. https://doi.org/10.1177/2396941519899311

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding Deficits Impede Word Learning and Memory in Adults With Developmental Language Disorders. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why Words Are Hard for Adults With Developmental Language Impairments. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders*, *48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

McGregor, K. K., Van Horne, A. O., Curran, M., Cook, S. W., & Cole, R. (2021). The Challenge of Rich Vocabulary Instruction for Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *52*(2), 467–484. https://doi.org/10.1044/2020_LSHSS-20-00110

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Messer, D., & Dockrell, J. E. (2006). Children's Naming and Word-Finding Difficulties: Descriptions and Explanations. *Journal of Speech, Language, and Hearing Research*, *49*(2), 309–324. https://doi.org/10.1044/1092-4388(2006/025)

Miller, C. A., & Gilbert, E. (2008). Comparison of performance on two nonverbal intelligence tests by adolescents with and without language impairment. *Journal of Communication Disorders*, *41*(4), 358–371. https://doi.org/10.1016/j.jcomdis.2008.02.003

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. https://doi.org/10.1044/1092-4388(2001/034)

Miller, C. A., Leonard, L. B., Kail, R. V., Zhang, X., Tomblin, J. B., & Francis, D. J. (2006). Response Time in 14-Year-Olds With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *49*(4), 712–728. https://doi.org/10.1044/1092-4388(2006/052)

Miller, C. A., & Poll, G. H. (2009). Response time in adults with a history of language difficulties. *Journal of Communication Disorders*, *42*(5), 365–379. https://doi.org/10.1016/j.jcomdis.2009.04.001

Montgomery, J. W. (1995). Sentence Comprehension in Children With Specific Language Impairment: The Role of Phonological Working Memory. *Journal of Speech, Language, and Hearing Research*, *38*(1), 187–199. https://doi.org/10.1044/jshr.3801.187

Montgomery, J. W. (2000a). Relation of working memory to off-line and real-time sentence processing in children withspecific language impairment. *Applied Psycholinguistics*, *21*(1), 117–148. https://doi.org/10.1017/S0142716400001065

Montgomery, J. W. (2000b). Verbal Working Memory and Sentence Comprehension in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *43*(2), 293–308. https://doi.org/10.1044/jslhr.4302.293

Montgomery, J. W., & Evans, J. L. (2009). Complex Sentence Comprehension and Working Memory in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 269–288. https://doi.org/10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., Gillam, R. B., Evans, J. L., & Sergeev, A. V. (2017). "Whatdunit?" Sentence Comprehension Abilities of Children With SLI: Sensitivity to Word Order in Canonical and Noncanonical Structures. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2603–2618. https://doi.org/10.1044/2017_JSLHR-L-17-0025

Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., & Gillam, R. B. (2018). Structural Relationship Between Cognitive Processing and Syntactic Sentence Comprehension in Children With and Without Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *61*(12), 2950–2976. https://doi.org/10.1044/2018_JSLHR-L-17-0421

Montgomery, J. W., & Gillam, R. B. (2024). Introduction to the Forum: Intervention With Children With Developmental Language Disorder. *American Journal of Speech-Language Pathology*, *33*(2), 527–529. https://doi.org/10.1044/2024_AJSLP-24-00046

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2021). A New Memory Perspective on the Sentence Comprehension Deficits of School-Age Children With Developmental Language Disorder: Implications for Theory, Assessment, and Intervention. *Language, Speech, and Hearing Services in Schools*, *52*(2), 449–466. https://doi.org/10.1044/2021_LSHSS-20-00128

Montgomery, J. W., Gillam, R. B., & Plante, E. (2024). Enhancing Syntactic Knowledge in School-Age Children With Developmental Language Disorder: The Promise of Syntactic Priming. *American Journal of Speech-Language Pathology*, *33*(2), 580–597. https://doi.org/10.1044/2023_AJSLP-23-00079

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working Memory and Specific Language Impairment: An Update on the Relation and Perspectives on Assessment and Treatment. *American Journal of Speech-Language Pathology*, *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Mountford, H. S., Braden, R., Newbury, D. F., & Morgan, A. T. (2022). The Genetic and Molecular Basis of Developmental Language Disorder: A Review. *Children*, *9*(5), 586. https://doi.org/10.3390/children9050586

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology*, *75*(1), 25–36. https://doi.org/10.4097/kja.21209

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Newcomer , P., & Hammill, D., (1988). *Test of Language Development - Primary*. 2nd Edition Austin, Tx: Pro-Ed.

Newbury, D. F., Gibson, J. L., Conti-Ramsden, G., Pickles, A., Durkin, K., & Toseeb, U. (2019). Using Polygenic Profiles to Predict Variation in Language and Psychosocial Outcomes in Early and Middle Childhood. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3381–3396. https://doi.org/10.1044/2019_JSLHR-L-19-0001

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S. L., Baird, G., Jannoun, L., Slonims, V., Stott, C. M., Merricks, M. J., Bolton, P. F., Bailey, A. J., Monaco, A. P., & International Molecular Genetic Study of Autism Consortium. (2002). FOXP2 is not a major susceptibility gene for autism or specific language impairment. *American Journal of Human Genetics*, *70*(5), 1318–1327. https://doi.org/10.1086/339931

Newell, A. & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (Ed.), *Cognitive Skills and their Acquisition*. Hillsdale, NJ: Lawrence Erlbaum. 1-55

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition*, *35*(2), 159–182. https://doi.org/10.1016/0010-0277(90)90013-A

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in Neurosciences*, *30*(4), 135–141. https://doi.org/10.1016/j.tins.2007.02.003

Nicolson, R. I., & Fawcett, A. J. (2019). Development of Dyslexia: The Delayed Neural Commitment Framework. *Frontiers in Behavioral Neuroscience*, *13*, 112. https://doi.org/10.3389/fnbeh.2019.00112

Nippold, M. A., Erskine, B. J., & Freed, D. B. (1988). Proportional and Functional Analogical Reasoning in Normal and Language-Impaired Children. *Journal of Speech and Hearing Disorders*, *53*(4), 440–448. https://doi.org/10.1044/jshd.5304.440

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational Versus Expository Discourse: A Study of Syntactic Development in Children, Adolescents, and Adults. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1048–1064. https://doi.org/10.1044/1092-4388(2005/073)

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository Discourse in Adolescents With Language Impairments: Examining Syntactic Development. *American Journal of Speech-Language Pathology*, *17*(4), 356–366. https://doi.org/10.1044/1058-0360(2008/07-0049)

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic Development in Adolescents With a History of Language Impairments: A Follow-Up Investigation. *American Journal of Speech-Language Pathology*, *18*(3), 241–251. https://doi.org/10.1044/1058-0360(2008/08-0022)

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Nitido, H., & Plante, E. (2020). Diagnosis of Developmental Language Disorder in Research Studies. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2777–2788. https://doi.org/10.1044/2020_JSLHR-20-00091

Nitin, R., Shaw, D. M., Rocha, D. B., Walters, C. E., Chabris, C. F., Camarata, S. M., Gordon, R. L., & Below, J. E. (2022). Association of Developmental Language Disorder With Comorbid Developmental Conditions Using Algorithmic Phenotyping. *JAMA Network Open*, *5*(12), e2248060. https://doi.org/10.1001/jamanetworkopen.2022.48060

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & 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. https://doi.org/10.1111/jcpp.12573

Norbury, C. F., Vamvakas, G., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2017). Language growth in children with heterogeneous language disorders: A population study. *Journal of Child Psychology and Psychiatry*, *58*(10), 1092–1105. https://doi.org/10.1111/jcpp.12793

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical Learning in Specific Language Impairment and Autism Spectrum Disorder: A Meta-Analysis. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01245

Oliveira, C. M., Henderson, L. M., & Hayiou-Thomas, M. E. (2023). Limited Evidence of an Association Between Language, Literacy, and Procedural Learning in Typical and Atypical Development: A Meta-Analysis. *Cognitive Science*, *47*(7), e13310. https://doi.org/10.1111/cogs.13310

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A Multivariate, Longitudinal Genetic Investigation of Language, Cognition and Behavior Problems from Childhood Through Adolescence. *Twin Research and Human Genetics*, *10*(1), 96–105. https://doi.org/10.1375/twin.10.1.96

Osborne, J. W. (2006). *Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analyses*. https://doi.org/10.7275/Z7WE-DG46

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Benjamins.

Parsons, S., (2021). splithalf: robust estimates of split half reliability. Journal of Open Source Software, 6(60), 3041, https://doi.org/10.21105/joss.03041

Park, J., Miller, C. A., & Mainela-Arnold, E. (2015). Processing Speed Measures as Clinical Markers for Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *58*(3), 954–960. https://doi.org/10.1044/2015_JSLHR-L-14-0092

Park, J. S., Miller, C. A., Sanjeevan, T., Van Hell, J. G., Weiss, D. J., & Mainela-Arnold, E. (2020). Bilingualism and Processing Speed in Typically Developing Children and Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1479–1493. https://doi.org/10.1044/2020_JSLHR-19-00403

Pauls, L. J., & Archibald, L. M. D. (2016). Executive Functions in Children With Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1074–1086. https://doi.org/10.1044/2016_JSLHR-L-15-0174

Pawłowska, M. (2014). Evaluation of Three Proposed Markers for Language Impairment in English: A Meta-Analysis of Diagnostic Accuracy Studies. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2261–2273. https://doi.org/10.1044/2014_JSLHR-L-13-0189

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Pimsleur, P. (1966). *Pimsleur language aptitude battery (form S)*. New York, NY: Harcourt, Brace and World.

Pimsleur, P., Reed, D.J., & Stansfield, C.W. (2004). *Pimsleur language aptitude battery: Manual 2004 edition*. Bethesda, MD: Second Language Testing

Pinker, S. (1991). Rules of Language. *Science*, *253*(5019), 530–535. https://doi.org/10.1126/science.1857983

Pinker, S. (1994). *The language instinct: The new science of language and mind*. Allen Lane, Penguin Press.

Pinker, S. (1999). *Words and rules: The ingredients of language* (1st ed). New York: Basic Books.

Pinker, S., & Ullman, M. (2002). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences*, *6*(11), 472–474. https://doi.org/10.1016/S1364-6613(02)02013-2

Piske, T., & Steinlen, A. K. (Eds.). (2022). *Cognition and second language acquisition: Studies on pre-school, primary school and secondary school children*. Narr Francke Attempto.

Plante, E., Shenkman, K., & Clark, M. M. (1996). Classification of Adults for Family Studies of Developmental Language Disorders. *Journal of Speech, Language, and Hearing Research*, *39*(3), 661–667. https://doi.org/10.1044/jshr.3903.661

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Plug, M. B., Van Wijngaarden, V., De Wilde, H., Van Binsbergen, E., Stegeman, I., Van Den Boogaard, M.-J. H., & Smit, A. L. (2021). Clinical Characteristics and Genetic Etiology of Children With Developmental Language Disorder. *Frontiers in Pediatrics*, *9*, 651995. https://doi.org/10.3389/fped.2021.651995

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of Clinical Markers of Specific Language Impairment in Adults. *Journal of Speech, Language, and Hearing Research*, *53*(2), 414–429. https://doi.org/10.1044/1092-4388(2009/08-0016)

Poll, G. H., Miller, C. A., & Van Hell, J. G. (2016). Sentence Repetition Accuracy in Adults With Developmental Language Impairment: Interactions of Participant Capacities and Sentence Structures. *Journal of Speech, Language, and Hearing Research*, *59*(2), 302–316. https://doi.org/10.1044/2015_JSLHR-L-15-0020

Poll, G. H., Watkins, H. S., & Miller, C. A. (2014). Lexical Decay During Online Sentence Processing in Adults With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2253–2260. https://doi.org/10.1044/2014_JSLHR-L-13-0265

Pratt, A. S., Peña, E. D., & Bedore, L. M. (2021). Sentence repetition with bilinguals with and without DLD: Differential effects of memory, vocabulary, and exposure. *Bilingualism: Language and Cognition*, *24*(2), 305–318. https://doi.org/10.1017/S1366728920000498

Prela, L., Llompart, M., & Dąbrowska, E. (2022). Aptitude and experience as predictors of grammatical proficiency in adult Greek-English bilinguals. *Frontiers in Psychology*, *13*, 1062821. https://doi.org/10.3389/fpsyg.2022.1062821

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2019). Hearing Matters More Than Seeing: A Cross-Modality Study of Statistical Learning and Reading Ability. *Scientific Studies of Reading*, *23*(1), 101–115. https://doi.org/10.1080/10888438.2018.1485680

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for the Raven's Progressive Matrices and vocabulary scales*. Pearson, San Antonio, TX.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1996). *Implicit Learning and Tacit Knowledge*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

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), 888–896. https://doi.org/10.1037/0278-7393.17.5.888

Redgrave, P.; Prescott, T.J., & Gurney, K. (1999). *The basal ganglia: A vertebrate solution to the selection problem? 89*(4), 1009–1023. https://doi-org.ezproxyd.bham.ac.uk/10.1016/S0306-4522(98)00319-4

Redmond, S. M., Ash, A. C., Christopulos, T. T., & Pfaff, T. (2019). Diagnostic Accuracy of Sentence Recall and Past Tense Measures for Identifying Children's Language Impairments. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2438–2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416–451. https://doi.org/10.1111/1460-6984.12102

Ren, J., Wang, M., & Arciuli, J. (2023). A meta-analysis on the correlations between statistical learning, language, and reading outcomes. *Developmental Psychology*, *59*(9), 1626–1644. https://doi.org/10.1037/dev0001577

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University, Evanston, Illinois. R package version 2.3.9, https://CRAN.R-project.org/package=psych.

Rice, M. L. (2020a). Causal Pathways for Specific Language Impairment: Lessons From Studies of Twins. *Journal of Speech, Language and Hearing Research (Online)*, *63*(10), 3224–3235. https://www.proquest.com/docview/2475528652?accountid=8630&bdid=38212&_bd=rTXosWn4CWgY8us3LbVNszzdXJU%3D

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of Input Effects on Word Comprehension of Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *37*(1), 106–122. https://doi.org/10.1044/jshr.3701.106

Rice, M. L., Taylor, C. L., Zubrick, S. R., Hoffman, L., & Earnest, K. K. (2020b). Heritability of Specific Language Impairment and Nonspecific Language Impairment

at Ages 4 and 6 Years Across Phenotypes of Speech, Language, and Nonverbal Cognition. *Journal of Speech, Language, and Hearing Research*, *63*(3), 793–813. https://doi.org/10.1044/2019_JSLHR-19-00012

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. https://doi.org/10.1044/jshr.3906.1239

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. https://doi.org/10.1044/jshr.3804.850

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense Over Time: The Longitudinal Course of Tense Acquisition in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

Rice, M. L., Zubrick, S. R., Taylor, C. L., Gayán, J., & Bontempo, D. E. (2014). Late Language Emergence in 24-Month-Old Twins: Heritable and Increased Risk for Late Language Emergence in Twins. *Journal of Speech, Language, and Hearing Research*, *57*(3), 917–928. https://doi.org/10.1044/1092-4388(2013/12-0350)

Rice, M. L., Zubrick, S. R., Taylor, C. L., Hoffman, L., & Gayán, J. (2018). Longitudinal Study of Language and Speech of Twins at 4 and 6 Years: Twinning Effects Decrease, Zygosity Effects Disappear, and Heritability Increases. *Journal of Speech, Language, and Hearing Research*, *61*(1), 79–93. https://doi.org/10.1044/2017_JSLHR-L-16-0366

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms: Sentence repetition in children with SLI.

International Journal of Language & Communication Disorders, *47*(5), 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Riches, N. G., Faragher, B., & Conti-Ramsden, G. (2006). Verb schema use and input dependence in 5-year-old children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, *41*(2), 117–135. https://doi.org/10.1080/13682820500216501

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb Learning in Children With SLI: Frequency and Spacing Effects. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1397–1411. https://doi.org/10.1044/1092-4388(2005/097)

Riches, N., Letts, C., Awad, H., Ramsey, R., & Dąbrowska, E. (2022). Collocational knowledge in children: A comparison of English-speaking monolingual children, and children acquiring English as an Additional Language. *Journal of Child Language*, *49*(5), 1008–1023. https://doi.org/10.1017/S0305000921000490

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*(3), 249–273. https://doi.org/10.1016/j.jecp.2006.02.002

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, *29*(8), 1358–1369. https://doi.org/10.1177/0956797618774253

Rodd, J. M. (2023). *Moving Experimental Psychology Online: How to Maintain Data Quality When We Can't See Our Participants*. [pre-print] https://doi.org/10.31234/osf.io/2fhcb

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Rost, G. C., & McGregor, K. K. (2012). Miranda Rights Comprehension in Young Adults With Specific Language Impairment. *American Journal of Speech-Language Pathology*, *21*(2), 101–108. https://doi.org/10.1044/1058-0360(2011/10-0094)

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language*, *47*(1), 5–21. https://doi.org/10.1017/S0305000919000655

Rowe, M.L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *5*(83), 1762–1774.

Rudolph, J. M. (2017). Case History Risk Factors for Specific Language Impairment: A Systematic Review and Meta-Analysis. *American Journal of Speech-Language Pathology*, *26*(3), 991–1010. https://doi.org/10.1044/2016_AJSLP-15-0181

Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of Neurodevelopmental Disorders*, *10*(1), 35. https://doi.org/10.1186/s11689-018-9252-y

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural Leanring and Neostryrial dysfunction in man. *Brain*, *111*(4), 941–960. https://doi.org/10.1093/brain/111.4.941

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *39*, 164–171. https://doi.org/10.1016/j.lindif.2015.03.010

Schmid, H.-J. (2020). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment* (First edition). Oxford University Press.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning1. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Childhood Language Skills and Adult Literacy: A 29-Year Follow-up Study. *Pediatrics*, *125*(3), e459–e466. https://doi.org/10.1542/peds.2008-2111

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., Goyal, S., Rea, E., French, C. A., Puliyadi, R., Groszer, M., Fisher, S. E., Mundry, R., Winter, C., Hevers, W., Pääbo, S., Enard, W., & Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, *111*(39), 14253–14258. https://doi.org/10.1073/pnas.1414542111

Scott, C. M., & Windsor, J. (2000). General Language Performance Measures in Spoken and Written Narrative and Expository Discourse of School-Age Children With Language Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324–339. https://doi.org/10.1044/jslhr.4302.324

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Malden, MA: Blackwell.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385. https://doi.org/10.1017/S0142716400010845

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*(1), 53–67. https://doi.org/10.1017/S0142716400010572

Semel, E., Wiig, E., Secord, W. (2006) *Clinical evaluation of language fundamentals – Fourth edition, Australian: Examiner's Manual*. Sydney, Australia: Pearson Clinical and Talent Assessment.

Shipley, W.C., Gruber, C.P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2 Manual.* Los Angeles, CA: Western Psychological Services.

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. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177. https://doi.org/10.1016/j.jecp.2017.08.005

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268–288. https://doi.org/10.1016/0010-0285(75)90012-2

Skehan, P. (2002). 4. Theorising and updating aptitude. In P. Robinson (Ed.), *Language Learning & Language Teaching* (Vol. 2, pp. 69–93). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.06ske

Smalle, E. H. M., & Bogaerts, L. (2024). Sensitive periods in language development: Do children outperform adults on auditory word-form segmentation? *Cortex*, *179*, 35–49. https://doi.org/10.1016/j.cortex.2024.07.001

Smith, N. V., & Tsimpli, I.-M. (1995). *The mind of a savant: Language learning and modularity*. Blackwell.

Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, *28*(3), 347–369. https://doi.org/10.1007/s11145-014-9527-y

Smith-Spark, J. H., & Gordon, R. (2022). Automaticity and Executive Abilities in Developmental Dyslexia: A Theoretical Review. *Brain Sciences*, *12*(4), 446. https://doi.org/10.3390/brainsci12040446

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., & Hulme, C. (2020). Dyslexia and Developmental Language Disorder: Comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*, *61*(6), 672–680. https://doi.org/10.1111/jcpp.13140

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., & Wellcome Language and Reading Project Team. (2019). Developmental Outcomes for Children at High Risk of Dyslexia and Children With Developmental Language Disorder. *Child Development*, *90*(5). https://doi.org/10.1111/cdev.13216

Sparks, R.L, Dale, P.S., & Patton, J.M. (2023). Individual differences in L1 attainment and language aptitude predict L2 achievement in instructed language learners. *The Modern Language Journal*, *107*(2), 479–508. https://doi.org/10.1111/modl.12841

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. https://doi.org/10.1007/s11145-014-9533-0

Squire, L. R. (1982). Comparisons between forms of amnesia: Some deficits are unique to Korsakoff's syndrome. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(6), 560–571. https://doi.org/10.1037/0278-7393.8.6.560

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., & Wixted, J. T. (2011). The Cognitive Neuroscience of Human Memory Since H.M. *Annual Review of Neuroscience*, *34*(1), 259–288. https://doi.org/10.1146/annurev-neuro-061010-113720

Squire, L. R., & Zola-Morgan, S. (1988). Memory: Brain systems and behavior. *Trends in Neurosciences*, *11*(4), 170–175. https://doi.org/10.1016/0166-2236(88)90144-0

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1

Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The Predictive Value of IQ. *Merrill-Palmer Quarterly*, *47*(1), 1–41. https://doi.org/10.1353/mpq.2001.0005

St Clair, M., Horsham, J., Lloyd-Esenkaya, V., Jackson, E., Gibson, J., Leitão, S., & Botting, N. (2022). The Engage with Developmental Language Disorder (E-DLD) project:

Cohort profile. *International Journal of Language & Communication Disorders*, 1460-6984.12835. https://doi.org/10.1111/1460-6984.12835

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-Impaired Preschoolers: A Follow-Up into Adolescence. *Journal of Speech, Language, and Hearing Research*, *41*(2), 407–418. https://www.proquest.com/docview/85683858?accountid=8630&bdid=38212&_bd=d%2FyMbwboZwUkhcEg8JeWCpaUiLI%3D

Street, J. A., & Dąbrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua*, *120*(8), 2080–2094. https://doi.org/10.1016/j.lingua.2010.01.004

Street, J. A., & Dąbrowska, E. (2014). Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics*, *35*(1), 97–118. https://doi.org/10.1017/S0142716412000367

Suddarth, R., Plante, E., & Vance, R. (2012). Written Narrative Characteristics in Adults With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *55*(2), 409–420. https://doi.org/10.1044/1092-4388(2011/10-0295)

Taatgen, N. A., Huss, D., Dickison, D., & Anderson, J. R. (2008). The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General*, *137*(3), 548–565. https://doi.org/10.1037/0096-3445.137.3.548

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Sentence Repetition as a Clinical Marker of Developmental Language Disorder: Evidence From Arabic. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4876–4899. https://doi.org/10.1044/2021_JSLHR-21-00244

Tallal, P., & Piercy, M. (1973a). Defects of Non-Verbal Auditory Perception in Children with Developmental Aphasia. *Nature*, *241*(5390), 468–469. https://doi.org/10.1038/241468a0

Tallal, P., & Piercy, M. (1973b). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia*, *11*(4), 389–398. https://doi.org/10.1016/0028-3932(73)90025-0

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, *12*(1), 83–93. https://doi.org/10.1016/0028-3932(74)90030-X

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, *13*(1), 69–74. https://doi.org/10.1016/0028-3932(75)90049-4

Tarvainen, S., Launonen, K., & Stolt, S. (2021). Oral language comprehension interventions in school-age children and adolescents with developmental language disorder: A systematic scoping review. *Autism & Developmental Language Impairments*, *6*, 239694152110104. https://doi.org/10.1177/23969415211010423

Thorndike, R. L., Hagen, E. P., & France, N. K. (1968). *Cognitive Abilities Test* Second Edition: Administration Manual. nferNelson, Windsor .

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard Univ. Press.

Tomasello, M. (2023). Differences in the Social Motivations and Emotions of Humans and Other Great Apes. *Human Nature*, *34*(4), 588–604. https://doi.org/10.1007/s12110-023-09464-0

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., & Herrmann, E. (2010). Ape and Human Cognition: What's the Difference? *Current Directions in Psychological Science*, *19*(1), 3–8. https://doi.org/10.1177/0963721409359300

Tomblin, J. B. (2010). The EpiSLI Database: A Publicly Available Database on Speech and Language. *Language, Speech, and Hearing Services in Schools*, *41*(1), 108–117. https://doi.org/10.1044/0161-1461(2009/08-0057)

Tomblin, J. B., & Buckwalter, P. R. (1998). Heritability of Poor Language Achievement Among Twins. *Journal of Speech, Language, and Hearing Research*, *41*(1), 188–199. https://doi.org/10.1044/jslhr.4101.188

Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing Specific Language Impairment in Adults for the Purpose of Pedigree Analysis. *Journal of Speech, Language, and Hearing Research*, *35*(4), 832–843. https://doi.org/10.1044/jshr.3504.832

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. https://doi.org/10.1080/15475440701377477

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. https://doi.org/10.1044/jslhr.4006.1245

Torchiano, M. (2016). *Effsize: Efficient effect size computation*. R package, version .8.1, Zenodo. https://doi.org/10.5281/ZENODO.1480624

Trevarthen, C., & Aitken, K. J. (2001). Infant Intersubjectivity: Research, Theory, and Clinical Applications. *Journal of Child Psychology and Psychiatry*, *42*(1), 3–48. https://doi.org/10.1111/1469-7610.00701

Tucci, A., & Choi, E. (2023). Developmental Language Disorder and Writing: A Scoping Review From Childhood to Adulthood. *Journal of Speech, Language, and Hearing Research*, *66*(8), 2900–2920. https://doi.org/10.1044/2023_JSLHR-22-00583

Tucker-Drob, E. M. (2019). Cognitive Aging and Dementia: A Life-Span Perspective. *Annual Review of Developmental Psychology*, *1*(1), 177–196. https://doi.org/10.1146/annurev-devpsych-121318-085204

Ukrainetz, T. A. (2024). Evidence-Based Expository Intervention: A Tutorial for Speech-Language Pathologists. *American Journal of Speech-Language Pathology*, *33*(2), 654–675. https://doi.org/10.1044/2023_AJSLP-23-00036

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2016). The Declarative/Procedural Model. In *Neurobiology of Language* (pp. 953–968). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T., Clark, G. M., Pullman, M. Y., Lovelett, J. T., Pierpont, E. I., Jiang, X., & Turkeltaub, P. E. (2024). The neuroanatomy of developmental language disorder: A systematic review and meta-analysis. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-024-01843-6

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific Language Impairment is not Specific to Language: The Procedural Deficit Hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *51*, 205–222. https://doi.org/10.1016/j.neubiorev.2015.01.008

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. https://doi.org/10.1177/0022219413508981

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends in Cognitive Sciences*, *9*(2), 53–59. https://doi.org/10.1016/j.tics.2004.12.002

Van Bergen, E., Snowling, M. J., De Zeeuw, E. L., Van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2018). Why do children read more? The influence of reading ability on voluntary reading practices. *Journal of Child Psychology and Psychiatry*, *59*(11), 1205–1214. https://doi.org/10.1111/jcpp.12910

Van Herwegen, J. (2015). Williams syndrome and its cognitive profile: The importance of eye movements. *Psychology Research and Behavior Management*, 143. https://doi.org/10.2147/PRBM.S63474

Vang Christensen, R. (2019). Sentence Repetition: A Clinical Marker for Developmental Language Disorder in Danish. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4450–4463. https://doi.org/10.1044/2019_JSLHR-L-18-0327

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). **mice**: Multivariate Imputation by Chained Equations in *R*. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex Publishing Corporation.

VanPatten, B. (2004). Input processing in second language acquisition. In: B. VanPatten (Ed.), *Processing Instruction: Theory, Research, and Commentary*. (pp. 5-31_ Mahwah, NJ: Lawrence Erlbaum Associates.

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, *34*(9), 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

Vugs, B., Hendriks, M., Cuperus, J., Knoors, H., & Verhoeven, L. (2017). Developmental Associations Between Working Memory and Language in Children With Specific Language Impairment: A Longitudinal Study. *Journal of Speech, Language and Hearing Research (Online)*, *60*(11), 3284–3294. https://www.proquest.com/docview/1967813397?accountid=8630&bdid=36209&_bd=oPkt0g3yIaNr%2BmD9MVxwvrxXKeg%3D

Wang, D., Zheng, L., Lin, Y., Zhang, Y., & Sheng, L. (2022). Sentence Repetition as a Clinical Marker for Mandarin-Speaking Preschoolers With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *65*(4), 1543–1560. https://doi.org/10.1044/2021_JSLHR-21-00401

Ward, L., Polišenská, K., & Bannard, C. (2024). Sentence Repetition as a Diagnostic Tool for Developmental Language Disorder: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 1–31. https://doi.org/10.1044/2024_JSLHR-23-00490

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain*, *125*(3), 452–464. https://doi.org/10.1093/brain/awf058

Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*(1/2), 28. https://doi.org/10.2307/2332510

West, G., Melby-Lervåg, M., & Hulme, C. (2021). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. *Developmental Psychology*, *57*(5), 749–770. https://doi.org/10.1037/dev0001172

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

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817. https://doi.org/10.2307/1912934

Whitehouse, A. J. O., Line, E. A., Watt, H. J., & Bishop, D. V. M. (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. https://doi.org/10.1080/13682820802708080

Wiig, E. H., Semel, E. M., & Secord, W. (2013). *CELF-5: Clinical Evaluation of Language Fundamentals* (5th edition). Pearson.

Wilmot, A., Boyes, M. E., Sievers, R., Leitao, S., & Norbury, C. (2024). *Impact of developmental language disorders on mental health and well-being across the lifespan: Lived experience perspectives.* [pre-print] https://doi.org/10.31234/osf.io/87hyg

Wilson Van Voorhis, C. R., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. https://doi.org/10.20982/tqmp.03.2.p043

Winckel, E., & Dąbrowska, E. (2024). Language Analytic Ability, Print Exposure, Memory and Comprehension of Complex Syntax by Adult Native Speakers. *Journal of Cognition*, *7*(1), 7. https://doi.org/10.5334/joc.333

Windsor, J., & Hwang, M. (1999). Children's Auditory Lexical Decisions: A Limited Processing Capacity Account of Language Impairment. *Journal of Speech, Language, and Hearing Research*, *42*(4), 990–1002. https://doi.org/10.1044/jslhr.4204.990

Witherstone, H. (2024). Speed of processing in Developmental Language Disorder (DLD): The case of real-time grammatical processing. *International Journal of Language & Communication Disorders*, 1460-6984.13014. https://doi.org/10.1111/1460-6984.13014

Wright, R., Geertsema, S., Le Roux, M., Winckel, E., & Dąbrowska, E. (2024). The acquisition of constructions: Does modality matter? *Reading & Writing*, *15*(1). https://doi.org/10.4102/rw.v15i1.489

Wu, S., Zhao, J., De Villiers, J., Liu, X. L., Rolfhus, E., Sun, X., Li, X., Pan, H., Wang, H., Zhu, Q., Dong, Y., Zhang, Y., & Jiang, F. (2023). Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: First evidence for Mandarin-speaking children in a population-based study. *The Lancet Regional Health - Western Pacific*, *34*, 100713. https://doi.org/10.1016/j.lanwpc.2023.100713

Zapparrata, N. M., Brooks, P. J., & Ober, T. (2023). Developmental Language Disorder Is Associated With Slower Processing Across Domains: A Meta-Analysis of Time-Based Tasks. *Journal of Speech, Language, and Hearing Research*, *66*(1), 325–346. https://doi.org/10.1044/2022_JSLHR-22-00221

Ziegenfusz, S., Paynter, J., Flückiger, B., & Westerveld, M. F. (2022). A systematic review of the academic achievement of primary and secondary school-aged students with developmental language disorder. *Autism & Developmental Language Impairments*, *7*, 239694152210993. https://doi.org/10.1177/23969415221099397

Zoelch, C., Seitz, K., & Schumann-Hengsteler, R. (2005). From rag (bag) s to riches: Measuring the developing central executive. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young Children's Cognitive Development: Interrelationships among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind* (pp. 39–69). Abingdon: Routledge.

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*(3), 286–292. https://doi.org/10.1016/j.bandc.2004.07.003

Appendices

The following appendices correspond to study two (Blake, Dąbrowska, & Llompart, in preparation)

Appendix 1. Study 2: Pearson's correlation matrix for the LT group

	SP	**SP_RT**	**IQ**	**IQ_RT**	**PSEUDO**	**PSEUDO_RT**	**ISL**	**ISL_RT**	**AUTOC**	**AUTOD**	**GJT**	**GJT_RT**	**P&S**	**P&S_RT**	**VOCAB**	**VOCAB_RT**
SP	-															
SP_RT	0.00															
IQ	0.73	-0.04														
IQ_RT	0.08	0.69	0.16													
PSEUDOWORD	0.20	0.17	0.18	0.24												
PSEUDOWORD_RT	-0.12	0.45	-0.26	0.37	-0.20											
ISL	0.52	0.22	0.57	0.28	0.35	0.15										
ISL_RT	0.54	0.40	0.47	0.40	0.28	0.30	0.74									
AUTOCV	-0.38	-0.25	-0.45	-0.35	-0.37	-0.06	-0.50	-0.34								
AUTODIFF	0.16	-0.24	0.14	0.09	0.22	-0.04	0.04	0.07	-0.18							
GJT	0.41	-0.33	0.33	-0.29	-0.06	-0.20	0.35	0.04	-0.25	0.11						
GJT_RT	0.27	0.51	0.07	0.39	0.15	0.57	0.35	0.68	-0.10	0.13	-0.21					
P&S	0.57	0.08	0.54	0.34	-0.01	0.18	0.42	0.32	-0.46	0.31	0.42	0.18				
P&S_RT	-0.07	0.44	-0.17	0.51	-0.11	0.61	0.03	0.33	0.03	0.09	-0.54	0.72	0.10			
VOCAB	0.41	-0.32	0.30	-0.09	0.08	-0.05	0.30	0.00	-0.31	0.17	0.62	-0.25	0.48	-0.36		
VOCAB_RT	-0.10	0.72	-0.25	0.55	0.06	0.63	-0.01	0.25	-0.09	-0.12	-0.40	0.67	0.05	0.66	-0.37	-

SP: Sentence pairs accuracy; SP_RT: Sentence pairs RT; IQ: Shipley block accuracy; IQ_RT; Shipley block response time; Pseudo: Pseudoword learning task (composite measure of accuracy); Pseudo_RT: Pseudoword learning task (composite measure of response time); ISL: Visual statistical learning task, accuracy; ISL_RT: Visual statistical learning task; response time; AutoCV: measure of automatization based on the MToH task; AutoDiff: measure of automatization based on the MToH task; GJT: Grammaticality judgement task accuracy; GJT_RT: Grammaticality judgement task response time; P&S: Pictures & sentences accuracy; P&S_RT: Pictures and sentences response time; Vocab: Shipley vocabulary – accuracy; Vocab_RT: Shipley vocabulary response time.

Appendix 2. Study 2: Pearson's correlation matrix for the DLD group

	SP	**SP_RT**	**IQ**	**IQ_RT**	**PSEUDO**	**PSEUDO_RT**	**ISL**	**ISL_RT**	**AUTOC**	**AUTOD**	**GJT**	**GJT_RT**	**P&S**	**P&S_RT**	**VOCAB**	**VOCAB_RT**
SP	-															
SP_RT	-0.29															
IQ	0.20	0.02														
IQ_RT	-0.13	0.64	-0.08													
PSEUDOWORD	0.28	-0.06	0.42	0.07												
PSEUDOWORD_RT	-0.29	0.38	-0.41	0.44	-0.34											
ISL	0.42	0.10	0.54	0.07	0.39	-0.34										
ISL_RT	0.02	0.32	0.19	0.51	0.29	0.12	0.40									
AUTOCV	-0.24	0.32	-0.39	0.20	-0.18	0.25	-0.26	-0.28								
AUTODIFF	0.02	-0.37	-0.02	-0.44	0.11	-0.37	0.07	-0.28	0.14							
GJT	0.44	-0.41	0.30	-0.20	0.56	-0.37	0.27	0.12	-0.14	0.19						
GJT_RT	-0.39	0.83	0.01	0.61	-0.05	0.44	-0.13	0.31	0.31	-0.44	-0.40					
P&S	0.41	-0.27	0.37	-0.06	0.61	-0.42	0.44	0.05	-0.21	0.36	0.55	-0.40				
P&S_RT	-0.45	0.72	-0.10	0.53	-0.17	0.56	-0.21	0.09	0.26	-0.40	-0.55	0.77	-0.34			
VOCAB	0.38	-0.32	0.20	-0.25	0.42	-0.47	0.43	0.26	-0.34	0.24	0.64	-0.51	0.59	-0.59		
VOCAB_RT	-0.07	0.80	-0.21	0.71	-0.25	0.47	-0.05	0.22	0.30	-0.38	-0.43	0.73	-0.29	0.68	-0.46	-

SP: Sentence pairs accuracy; SP_RT: Sentence pairs RT; IQ: Shipley block accuracy; IQ_RT; Shipley block response time; Pseudo: Pseudoword learning task (composite measure of accuracy); Pseudo_RT: Pseudoword learning task (composite measure of response time); ISL: Visual statistical learning task, accuracy; ISL_RT: Visual statistical learning task; response time; AutoCV: measure of automatization based on the MToH task; AutoDiff: measure of automatization based on the MToH task; GJT: Grammaticality judgement task accuracy; GJT_RT: Grammaticality judgement task response time; P&S: Pictures & sentences accuracy; P&S_RT: Pictures and sentences response time; Vocab: Shipley vocabulary – accuracy; Vocab_RT: Shipley vocabulary response time.

Appendix 3. Study 2: Questions used to determine DLD status

1. Please advise information regarding your language difficulties. For example, specific problems that you experience?
2. From what age did you experience language difficulties?
3. Do you find it difficult to make yourself understood?
4. Do you have problems finding the right words to say what you think?
5. Do you have problems understanding what other people say to you?
6. Do you have a formal diagnosis for your language difficulties (e.g., Developmental Language Disorder, Specific Language Impairment),
7. If you have a formal diagnosis, please advise at what age you were diagnosed?
8. Have you had speech and language therapy in the past?
9. If you have not had speech and language therapy in the past, please advise details of any additional support you have received?
10. Do any other members of your family have language difficulties?
11. Do you have any other difficulties?

Appendix 4. Response time data

Percentages of response time data removed according to +/- 3 Median Absolute Deviations (on accurate trials only)

	Language typical	***DLD***
Sentence Pairs	8.57	7.63
Shipley Block Design	13.35	12.5
Visual statistical learning	7.56	6.91
Pseudoword training phase	7.91	6.80
Pseudoword test 1	4.98	7.41
Pseudoword test 2	8.53	8.36
Grammaticality Judgement task	6.61	5.74
Pictures & Sentences	3.62	8.33
Shipley Vocabulary	12.66	8.21

All figures reported as percentages.