



University of Zagreb

Faculty of Education and Rehabilitation Sciences
Postgraduate doctoral study Speech, Language and
Hearing Disorders

Antonija Blaži Ostojić

**DISCOURSE COMPREHENSION: THE
EFFECT OF PSYCHOLINGUISTIC WORD
FEATURES ON ANAPHORA
PROCESSING IN CHILDREN WITH
LANGUAGE DISORDER**

DOCTORAL THESIS

Zagreb, 2026



University of Zagreb

Faculty of Education and Rehabilitation Sciences
Postgraduate doctoral study Speech, Language and
Hearing Disorders

Antonija Blaži Ostojić

**DISCOURSE COMPREHENSION: THE
EFFECT OF PSYCHOLINGUISTIC WORD
FEATURES ON ANAPHORA
PROCESSING IN CHILDREN WITH
LANGUAGE DISORDER**

DOCTORAL THESIS

Mentor(s):

Jelena Kuvač Kraljević, PhD, Full professor

David Saldaña, PhD, Full Professor

Zagreb, 2026



Sveučilište u Zagrebu

Edukacijsko-rehabilitacijski fakultet

Poslijediplomski doktorski studij Poremećaji jezika govora i
slušanja

Antonija Blaži Ostojić

**RAZUMIJEVANJE DISKURSA: UTJECAJ
(PSIHO)LINGVISTIČKIH OBILJEŽJA
RIJEČI NA OBRADU ANAFORIČKIH
ODNOSA U DJECE S JEZIČNIM
POREMEĆAJEM**

DOKTORSKI RAD

Mentor(i):

Prof.dr.sc. Jelena Kuvač Kraljević

Prof.dr.sc. David Saldaña

Zagreb, 2026

Content:

1. INTRODUCTION	1
1.1 The concept of reading comprehension	2
1.2 Models of reading comprehension	3
1.2.1 The Simple View of Reading theory	4
1.2.2 Reading System Framework and Lexical Quality Hypothesis.....	27
1.3 Anaphora processing and reading comprehension.....	38
1.3.1 Comparison of anaphoric structure between Croatian and English	39
1.3.2 The process of anaphora resolution.....	39
1.4 Reading comprehension and reader's individual differences	41
1.4.1 Models of language-based disorders from the perspective of general phonological and non-phonological language skills.....	41
1.4.2 Models of language-based disorders from the perspective of reading skills and reading comprehension features	47
1.5 Assessment of reading comprehension and its underlying components	50
1.5.1 Eye-tracking in reading research.....	53
2. THE PRESENT STUDY	59
3. METHOD	62
3.1 Participants	62
3.2 Sampling procedure.....	63
3.3 Research materials, tasks and methods	69
3.4 Experimental procedure	75
3.4.1 Development of the experimental materials.....	76
3.5 Experimental procedure in online conditions (eye-tracking).....	81
3.6 Experimental procedure in offline conditions (comprehension accuracy).....	85
4. DATA ANALYSIS.....	86
4.1 Analysis 1: Real-Time Anaphora Processing during reading comprehension.....	87
4.2 Analysis 2: Reading Comprehension Accuracy	88
4.3 Analysis 3: Predictors of Reading Comprehension.....	88
5. RESULTS	89
5.1 Analysis 1: real-time anaphora processing during reading comprehension.....	89
5.1.1 Comparison of results between children with typical language development and those with language-based disorder	90
5.1.2 Summary of results for children with typical language development and those with language-based disorder across AOIs.....	111

5.1.3	Comparison of results between children with TLD and specific profiles of LBD	112
5.1.4	Summary of results for children with typical language development and specific profiles of language-based disorder across AOIs	135
5.2	Analysis 2: Offline comprehension.....	136
5.2.1	Reading comprehension performance of children with TLD and children with LBD	136
5.2.2	Reading comprehension performance in children with TLD and children with different LBD profiles.....	137
5.3	Analysis 3: Predictive components	139
6.	DISCUSSION.....	143
6.1	Lexical Quality and Language-based profiles as determinants of anaphora resolution in reading comprehension.....	144
6.1.1	Word frequency and imageability effects in TLD and LBD.....	145
6.1.2	Word frequency and imageability effects across different language-based disorder profiles	153
6.2	The effect of question type and specific language difficulties profiles on reading comprehension.....	159
6.3	The predictive role of (meta)linguistic skills in reading comprehension.....	163
7.	HYPOTHESIS VERIFICATION	167
8.	LIMITATIONS OF THE STUDY.....	171
9.	SCIENTIFIC CONTRIBUTIONS.....	172
10.	IMPLICATIONS FOR FUTURE RESEARCH	174
11.	CONCLUSION	175
12.	REFERENCES.....	178
13.	APPENDICES.....	230
14.	CURRICULUM VITAE	240

Zahvala

I would like to express my gratitude to my supervisors Prof. Jelena Kuvač Kraljević and Prof. David Saldaña, on constructive suggestions and guidance during this process. I am deeply grateful for the opportunity to spend time at the University of Seville whilst working on this thesis. I hope the work I have done repaid the trust you gave me.

Zahvalu upućujem i svim kolegicama i kolegama logopedima iz zagrebačkih škola, a posebno profesorici i profesoru Bonetti koji su mi značajno pomogli u prikupljanju uzorka za ovo istraživanje. Zahvaljujem i dragim kolegicama, doc.dr.sc. Ani Matić i dr.sc. Sari Košutar, na brznoj vezi Zagreb-Sevilla i transferu preostalih eye-tracking podataka, tijekom mog boravka u Seville. Zahvaljujem i dragoj Valentini na danima i noćima konstruktivnih metodoloških i statističkih rasprava.

Moja postignuća i pobjede su moji, no oni ne bi bili mogući bez ključnih ljudi koji su me svakodnevno podržavali, pomagali, bodrili, poticali, iskazivali mi ljubav na bezbroj načina: suprug, mama, tata, brat, baka, šogorica... Vi ste mi dali svu snagu, Vi ste me ‘trpili’, Vi mi niste dali da odustanem niti zbog čega- niti zbog koga, Vi ste me osnaživali, i u najtežim trenucima.

I zato VAM najveća HVALA! Vi ste moj stup, oslonac i ponos. Nadam se da sam i ja vaš.

Usljed izrade ovog istraživanja i rada, na svijet je došla moja najljepša djevojčica. To maleno čudo od prvog je trenutka imalo neopisivo razumijevanje za sve dane, sate i minute koje je mama provela čitajući, računajući, pišući. Iako malena, znala je točno kada mami treba dati ‘vjetar u leđa’, zagrliti i reći: ‘Mama, sve će biti u redu!’.

Stoga ovaj rad posvećujem mojoj kćerkici i mojoj divnoj obitelji.

‘‘Za njih sve, bez njih ništa.’’

The supervisors:

Jelena Kuvač Kraljević, PhD, Full professor

Jelena Kuvač Kraljević graduated in speech-language pathology at the University of Zagreb Faculty of Education and Rehabilitation Sciences and obtained her master's and doctoral degrees in linguistics at the University of Zagreb Faculty of Humanities and Social Sciences. She has been employed at the University of Zagreb Department of Speech and Language Pathology since 2000. Her research interests include language development and processing, as well as developmental and acquired language disorders.

She is head of the Postgraduate Doctoral Programme *Speech, Language and Hearing Disorders*; the Croatian representative on the Child Language Committee of the International Association of Communication Sciences and Disorders (IALP); a member of the editorial board of *First Language*; and Deputy Lead of Early-Stage Researcher (ERS) in the field of aphasiology.

She has published over 100 articles in journals, proceedings and books (H index of 14 in Scopus and 13 in Web of Science). She is co-author of the book *Methodology of Children's Language Research* (published in Croatian in 2007) and editor of the book *Recognition and Education of Children with Language Disabilities* (published in Croatian in 2015). She has developed or adapted seven language tests in Croatian to assess the language, reading and writing skills of children and adults with language disorders, including TROG-2:HR and the New Reynell Developmental Language Scale (NRDLS-HR). She has been principal investigator on six scientific projects, most of which focused on knowledge and technology transfer. She supervisor of seven completed dissertations.

Jelena Kuvač Kraljević was awarded the State Prize for Science in 2016, the Prize for Scientific Productivity in 2016, and the Prize for the Best Scientist of the Faculty of Education and Rehabilitation Sciences in 2014. In addition, she received two professional prizes: the Croatian Association of Speech-Language Pathologists Prize for combining science and practice in speech-language therapy in 2018, and the "Stjepan Cipek" Prize for improving teaching skills in higher education in 2020.

David Saldaña, PhD, Full Professor

David Saldaña's research focuses on oral and written language comprehension in atypical and neurodiverse populations. He directs the Diversity, Cognition, and Language Laboratory at the

University of Seville, where his team studies the cognitive and linguistic processes of autistic individuals, deaf and hard-of-hearing students, and people with intellectual disabilities, language impairments, hyperlexia, or low reading comprehension levels.

His work connects basic cognitive research with practical social applications. Using eye-tracking technologies in recently updated facilities, his research group analyzes real-time sentence processing and the impact of visual and digital supports.

Throughout his career, David Saldaña has managed numerous research grants and competitive funding. He has been the Principal Investigator (PI) for 6 national R&D projects (*Plan Nacional I+D+i*) and 2 regional projects funded by the *Junta de Andalucía*, focusing mainly on literacy and autism. His recent projects include *PID2021-122658NB-I00* (€108,658.00) and *PID2024-160482NB-I00* (€100,000.00) from the Spanish Ministry of Science, as well as the *CX24-00251* project from Fundación La Caixa (€68,920.00). Internationally, he coordinated the University of Seville's involvement in European Union Framework Program networks, such as the ITN *LanPercept* (7th Framework Program) and the ITN *e-LADDA* (Horizon 2020), managing a local budget of €429,558.13 to study language development in digital environments. He has also represented Spain on the national management committee of the COST Action *IS1406* on language disorders.

He directs the Autism Chair (*Cátedra de Autismo*) at the University of Seville, a joint initiative with the *Autismo Sevilla* association.

His academic output includes 69 papers and 37 book chapters, with an H index of 20 in Scopus and 19 in Web of Science. Regarding graduate mentorship, he has supervised 11 completed doctoral theses, including international co-tutelles with the Norwegian University of Science and Technology (NTNU) and the *Universidad Autónoma de Baja California*.

ABSTRACT

Reading comprehension is a multi-level and complex process, shaped by numerous cognitive-linguistic factors. Its complexity is particularly evident in populations of children in whom these factors are not equally developed. Children with language-based disorders (dyslexia, developmental language disorder, combined difficulties) constitute a heterogeneous group, highlighting the need to examine specific psycholinguistic and (meta)linguistic factors that may disrupt the processing of written information across different language profiles.

The aim of this doctoral dissertation was to examine the mechanisms that support written discourse comprehension, operationalised through the processing of anaphoric relations, in school-aged children with typical language development and with language-based disorders. The research was guided by three objectives: examining the role of psycholinguistic word properties (frequency and imageability) in the real-time processing of anaphoric relations; analysing comprehension accuracy with respect to question type and children's language profile; and identifying (meta)language and cognitive predictors of individual differences in reading comprehension. The theoretical framework comprises the Simple View of Reading (SVR; Gough & Tunmer, 1986), which distinguishes decoding and language comprehension as key components of reading competence, and the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007), which posits that the quality of orthographic, phonological, and semantic representations determines the efficiency of lexical access.

Participants were 70 Croatian-speaking children in Years 5 and 6 of primary school, classified in accordance with the Bishop and Snowling (2004) framework into a typical language development group (TLD; N = 33) and a language-based disorder group (LBD; N = 37). Following the classification proposed by Catts et al. (2005) and the SVR model, the LBD group was further divided into three subgroups: poor decoders (PD; N = 16), poor oral language group (POL; N = 10), and mixed difficulties group (MD; N = 11). Both online eye-tracking measures and offline question-answering measures were employed. Four passages (through four stories) with manipulated levels of word frequency and imageability were analysed across three areas of interest: the antecedent, the anaphor, and the words preceding the target areas.

The findings of the first research objective indicate that word frequency is the most robust predictor of processing effort across all phases and all areas of interest. Imageability played a temporally constrained and conditional role: in the earliest processing phase, high

imageability paradoxically increased processing demands when combined with low frequency, consistent with the distinction between conceptual richness and the automaticity of lexical access within the LQH framework. Children with language-based disorders showed longer processing durations, more fixations, and more regressive eye movements than their typically developing peers. The analysis across four subgroups revealed distinct profiles: the PD and MD groups showed sensitivity to word frequency already in early integrative processing; the POL group specifically failed to benefit from high frequency despite preserved processing speed, indicative of a deficit in the semantic-conceptual dimension of lexical quality; and the MD group was the only one to show frequency sensitivity in anaphor processing as well, reflecting the cumulative effect of co-occurring difficulties.

The findings of the second research objective indicate that all children, regardless of language profile, achieved lower accuracy on inferential than on literal questions, with no significant interaction between question type and group, suggesting that inferential questions are proportionally more demanding across all profiles. The TLD group consistently outperformed all clinical profiles, whilst within the LBD group the MD subgroup achieved the lowest scores, and the PD group outperformed MD but not the POL group, suggesting that combined difficulties across both SVR components have the most pronounced negative impact on reading comprehension.

The findings of the third research objective confirm the developmental premise of the SVR model: measures of language comprehension (vocabulary, grammatical understanding, and inferencing) were the strongest predictors of reading comprehension, whilst decoding and phonological processing were not statistically significant. The dominant role of language comprehension is particularly pronounced in transparent orthographies such as Croatian, in which decoding skills become automatised earlier. Group membership remained a significant predictor even after controlling for all measured variables, suggesting that language-based disorders encompass broader developmental differences not fully captured by isolated measures.

Based on the findings across all three research objectives, it is concluded that the type of language-based disorder potentially modulates which aspect of lexical quality most critically affects the establishment of referential relations during reading comprehension. Online eye-tracking measures reveal differences at the level of lexical integration and discourse processing that offline measures do not capture, which represents a methodological

contribution of this research. A comprehensive understanding of language-based difficulties in the domain of reading comprehension requires multidimensional assessment that addresses the quality of lexical representations, language comprehension, and higher-level discourse processes, with implications for assessment and targeted intervention.

Keywords: reading comprehension, anaphoric relations, eye-tracking, word frequency, imageability, language-based disorder, lexical quality, Simple View of Reading

SAŽETAK

Razumijevanje pročitano­g višerazinski je i kompleksan proces, uvjetovan brojnim kognitivnim i jezičnim čimbenicima. Njegova složenost osobito je vidljiva u populaciji djece kod kojih ti čimbenici nisu jednako razvijeni. Djeca s jezično-utemeljenim poremećajima (disleksija, razvojni jezični poremećaj, kombinirane teškoće) čine heterogenu skupinu, što naglašava potrebu za istraživanjem specifičnih psiholingvističkih i (meta)jezičnih čimbenika koji mogu narušiti obradu pisanih informacija u različitim jezičnim profilima.

Cilj ovog doktorskog rada bio je ispitati mehanizme koji podupiru razumijevanje pisanog diskursa, operacionalizirano kroz obradu anaforičkih odnosa, u djece školske dobi s tipičnim jezičnim razvojem i s jezično-utemeljenim poremećajima. Istraživanje je vođeno trima ciljevima: ispitivanjem uloge psiholingvističkih obilježja riječi (učestalosti i predočivosti) u obradi anaforičkih odnosa u stvarnom vremenu, analizom točnosti razumijevanja s obzirom na vrstu pitanja i jezični profil djeteta te identifikacijom (meta)jezičnih i kognitivnih prediktora individualnih razlika u razumijevanju pisanog teksta. Teorijski okvir čine Jednostavan model čitanja (Simple View of Reading- SVR; Gough i Tunmer, 1986), koji razlikuje dekodiranje i jezično razumijevanje kao ključne komponente čitalačke kompetencije te Hipoteza o kvaliteti leksičkih reprezentacija (Lexical quality hypothesis- LQH; Perfetti i Hart, 2002; Perfetti, 2007), koja pretpostavlja da kvaliteta ortografskih, fonoloških i semantičkih reprezentacija određuje učinkovitost leksičkog pristupa.

U istraživanju su sudjelovala djeca petih i šestih razreda (N = 70), govornici hrvatskoga jezika, razvrstana prema modelu Bishop i Snowling (2004) u skupinu tipičnog jezičnog razvoja (TLD; N = 33) i skupinu jezično-utemeljenih poremećaja (LBD; N = 37). Sukladno klasifikaciji Catts i suradnika (2005) i SVR-u, skupina LBD podijeljena je u tri podskupine: slabi dekoderi (PD; N = 16), skupina sa siromašnim usmenim jezičnim znanjem (POL; N = 10) i skupina s mješovitim teškoćama (MD; N = 11). Primijenjene su online mjere praćenja pokreta očiju i offline mjere odgovaranja na pitanja. Četiri priče s manipuliranim razinama učestalosti i predočivosti analizirane su u trima područjima interesa: antecedent, anafora i riječi koje prethode ciljnim područjima.

Rezultati prvog istraživačkog cilja pokazuju da je psiholingvističko obilježje učestalosti riječi najrobusniji prediktor procesnog napora kroz sve faze obrade i sva područja interesa.

Predočivost je imala vremenski ograničenu i uvjetovanu ulogu: u najranijoj fazi obrade visoka predočivost paradoksalno je povećavala procesni napor u kombinaciji s niskom učestalošću,

sukladno razlikovanju konceptualnog bogatstva i automatiziranosti leksičkog pristupa unutar LQH-a. Djeca s jezično-utemeljenim poremećajima pokazala su dulje trajanje obrade, više fiksacija i više regresivnih pokreta oka od vršnjaka tipičnog razvoja. Analiza na četiri podskupine otkrila je različite profile: PD i MD pokazali su osjetljivost na učestalost već u ranoj integrativnoj obradi, POL skupina nije profitirala od visoke učestalosti unatoč očuvanoj procesnoj brzini, upućujući na deficit u semantičko-konceptualnoj dimenziji leksičke kvalitete, a MD skupina jedina je pokazala osjetljivost na učestalost i u obradi anafore, odražavajući kumulativni učinak supojavnih teškoća.

Rezultati drugog istraživačkog cilja pokazuju da su sva djeca, neovisno o jezičnom profilu, postigla nižu točnost na inferencijalnim nego na literalnim pitanjima, bez značajne interakcije između vrste pitanja i grupe, što upućuje na to da su inferencijalna pitanja proporcionalno zahtjevnija za sve profile. TLD skupina dosljedno je nadmašila sve kliničke profile, dok je MD skupina unutar LBD-a postigla najniže rezultate, a PD je nadmašila MD, ali ne i POL skupinu, sugerirajući da kombinacija teškoća u objema SVR komponentama ima najizraženiji negativni učinak.

Rezultati trećeg istraživačkog cilja potvrđuju razvojnu pretpostavku SVR modela: mjere jezičnog razumijevanja (rječnik, gramatičko znanje i inferencijalno zaključivanje) bili su najsnažniji prediktori razumijevanja pisanog teksta, dok dekodiranje i fonološka obrada nisu bili statistički značajni. Dominantna uloga jezičnog razumijevanja posebno je izražena u transparentnim ortografijama poput hrvatskoga, u kojima se dekodiranje automatizira ranije. Pripadnost grupi ostala je značajan prediktor i nakon kontrole svih mjerenih varijabli, upućujući na to da jezično-utemeljeni poremećaji obuhvaćaju šire razvojne razlike koje izolirane mjere ne zahvaćaju.

Temeljem rezultata sva tri istraživačka cilja zaključuje se da podskupina jezično-utemeljenog poremećaja potencijalno modulira koji aspekt kvalitete leksičke reprezentacije najkritičnije utječe na uspostavljanje referencijalnih odnosa tijekom razumijevanja pročitano. Online mjere praćenja pokreta očiju otkrivaju razlike na razini leksičke integracije i diskursne obrade koje offline mjere ne bilježe, što predstavlja metodološki doprinos ovog istraživanja. Cjelovito razumijevanje jezično-utemeljenih teškoća zahtijeva multidimenzionalnu procjenu koja zahvaća kvalitetu leksičkih reprezentacija, jezično razumijevanje i višerazinske diskursne procese, s implikacijama za dijagnostiku i intervenciju.

Ključne riječi: razumijevanje pisanog teksta, anaforički odnosi, praćenje pokreta očiju, učestalost riječi, predočivost, jezično-utemeljeni poremećaj, leksička kvaliteta, jednostavan model čitanja

1. INTRODUCTION

In recent years, the complexity of reading comprehension has increasingly attracted the attention of researchers in the field of written language development and processing. It has become clear that no single component, skill, or dimension can fully uniquely explain the processes of written language comprehension (Cain & Oakhill, 2006; Scarborough et al., 2009; Perfetti & Stafura, 2014). Furthermore, the identification of children whose reading comprehension skills — referred to in the literature as *poor comprehenders* — are significantly lower than their reading (phonological) skills has raised questions about the true nature of reading comprehension processes, their interconnection with other language and cognitive factors, and their multidimensionality across different language profiles and linguistic contexts (Catts et al., 2005; Cain & Oakhill, 2006; Nation & Snowling, 2006; Perfetti & Stafura, 2014).

The researchers aimed to describe and explain the underlying processes of reading comprehension and to provide possible explanations for the interdependence of processing components and their causal roles (Gough & Tunmer, 1986; Perfetti, 2001; Knitsch & Rawson, 2005; LARRC & Chiu, 2018). Previous studies have also led to the development of valuable theoretical models that can further guide researchers in constructing and addressing specific research aims. To date, the most prominent models and hypotheses, such as the *Simple View of Reading* (Gough and Tunmer, 1986), the *Lexical quality hypothesis* (Perfetti & Hart, 2002; Perfetti, 2007), the *Situation model* (van Dijk & Kintsch, 1983), and the *Coherent mental model* (Johnson-Laird, 1980), have aimed to clarify the reading comprehension processes in the context of a specific language processing levels and have thus created opportunities to study written language comprehension through an interactive, multilevel approach.

Therefore, the general motivation of this doctoral study is to further clarify the nature of these processes by contributing to data systematicity, especially in languages orthographically different from English (e.g. Croatian) and to extend the previous models and approaches in the methodology of research on child language and reading comprehension processes. There is a clear need for further research in this field, specifically with the inclusion of clinical subgroups such as children with different profiles of language disorder, and reading abilities, and by combining various experimental methodologies (Bishop & Snowling, 2004; Catts, 2018, Joseph et al., 2021).

1.1 The concept of reading comprehension

Reading comprehension is widely acknowledged as one of the most complex language behaviours, involving the simultaneous processing and integration of meaning from printed text. It is widely recognised as the outcome of reading (Clarke et al., 2010; Snow, 2021), yet no single formal definition exists because of its inherent complexity. However, different theoretical approaches highlight complementary aspects of this construct.

Constructivist approach (Goodman, 1967; Rosenblatt, 2018) view comprehension as an active meaning-making process, shaped by the interaction between reader and text. The RAND Reading Study Group (Snow, 2002) further emphasises the sociocultural dimension, describing comprehension as a process influenced by the characteristics of the reader (e.g. motivation), the text, and the reading purpose, with engagement and prior experience playing a central role.

From a psycholinguistic approach (proponents of this approach are, for example, Keith Rayner, Jane Oakhill, Arthur C. Graesser), reading comprehension builds on general language skills, but differs from oral comprehension by relying on written input, which requires additional decoding and integration processes (Perfetti et al., 2001; Hulme & Snowling, 2021). Therefore, it is at least partially justified to consider the written discourse as the representation of some form of spoken language (Hulme & Snowling, 2021). Readers draw on linguistic and general knowledge, cognitive resources, and topic familiarity to build coherent representations of the text (Cain, 2009; Hogan et al., 2011). Similarly, Cain (2009) defines reading comprehension as a strategic, dynamic, and interactive process in which the reader builds a mental representation of situations or events described in the text by recognising individual words and their meanings, connecting them to existing syntactic knowledge, and integrating them into larger units.

Cognitive approach expands the view on reading comprehension by highlighting the role of working memory, executive functions and prior knowledge in constructing integrated mental models of the text (Gernsbacher, 1990; Kintsch, 1998). For example, Schema theory (Bartlett, 1932; Anderson et al., 1977) further explains how new information is assimilated into existing frameworks, reinforcing the dynamic and strategic nature of comprehension (Duke & Pearson, 2002; Willingham, 2017).

Taken together, approaches of reading comprehension reflect an evolving understanding: it is a dynamic, interactive process involving the decoding of written language integrating

language and cognitive resources, drawing on prior knowledge, and adapting strategies to context. What unites these approaches is the view of comprehension as an active and constructive act, with success shaped by both the reader and the text. The present work builds on the definition in line with the psycholinguistic approach, which frames comprehension primarily as the understanding of written language through the interplay of lexical, semantic-syntactic, and discourse-level processes.

1.2 Models of reading comprehension

From a psycholinguistic approach, reading comprehension is best explained by models that describe how key linguistic processes operate and interact.

Reading comprehension is now well established as extending beyond visual word recognition. It involves coordinated language and cognitive mechanisms operating across multiple levels — word, sentence, and discourse — to extract meaning and build a coherent mental representation (Perfetti, 2001; Rayner & Reichle, 2010). Contemporary perspectives emphasise that these processes are activated simultaneously and interactively, reflecting parallel, rather than serial processing (Rumelhart, 1977; Perfetti, 1985).

Theoretical models explain how comprehension unfolds across three domains: lexical (access), semantic-syntactic (integration), and discourse-level processes (mental representation construction). Lexical access models describe how the mental lexicon is organised and how readers retrieve phonological, orthographic, and semantic information from written words (Perfetti & Hart, 2002). This process underpins all higher stages of comprehension (Perfetti, 1985; Ramus, 2001; Perfetti, 2007), serving as a core mechanism that enables the integration of word-level information into sentence and discourse meaning. Semantic-syntactic integration models focus on how word meaning and grammatical roles are combined with sentence-level context to support interpretation (Perfetti & Stafura, 2014; Van Dyke, 2021). Finally, mental representation models explain how information from words and sentences is integrated with prior knowledge to construct situation model of the text (Johnson-Laird, 1980; van Dijk & Kintsch, 1983).

The following section discusses the key models (The Simple View of Reading; The Reading System Framework) and hypothesis (The Lexical Quality Hypothesis) on which this dissertation is based, organised around these three domains (lexical-, semantic-syntactic-, and discourse-level processes).

1.2.1 The Simple View of Reading theory

The nature of the reading comprehension process is described by one of the best-known models in the reading and comprehension research: *The Simple View of Reading* (SVR) by Gough and Tunmer (1986). By explaining the interdependence of the two core components — decoding and language (listening) comprehension — this theory presents a 'simple view' of otherwise highly complex processes and thus introduces the first dimension of reading comprehension processes.

According to the original SVR (Gough & Tunmer, 1986), reading comprehension is the outcome of these two core component skills. The following equation represents the above assumption: $RC = D \times (L)C$, where RC stands for reading comprehension, D for decoding skills, and (L)C for listening comprehension. However, in 2020, Hoover and Tunmer revised and extended their SVR model to better reflect the process of reading comprehension. They broadened the two main components of the RC in their equation by using terms such as *word recognition* and *language comprehension*. Word recognition reflects the product of automatised, unconscious, and efficient decoding processes. Accordingly, decoding plays a critical role in successful word recognition, as it enables the development of the ability to translate graphemes in print into their corresponding phonemes in language, based on knowledge of letter-sound correspondence rules (Gough & Tunmer, 1986; Foorman et al., 2016). In line with both the original and revised SVR models (Gough & Tunmer, 1986; Hoover & Tunmer, 2020), this dissertation conceptualises word recognition as the outcome of decoding processes that become increasingly efficient and automatised through experience. Consequently, decoding and word recognition are inherently linked and cannot be treated as fully independent components. In contrast, language comprehension encompasses a broader set of language processes necessary for understanding language, including lexical knowledge, grammatical competence, background knowledge, and inferential skills (Hoover & Tunmer, 2020). Finally, reading comprehension itself can be defined as the process of constructing a coherent mental model of a written text by generating meaning-based inferences during reading (Gough & Tunmer, 1986; Hoover & Tunmer, 2020).

Regardless of the terminology or the specific version of the model adopted, a weakness in even a single component is sufficient to compromise the entire process of reading comprehension. This means that if a reader's decoding skills (D) are below average (regardless of their language comprehension level), this weakness will result in poor reading comprehension. Similarly, if language comprehension (LC) is violated, the reader will have

difficulty with the final product — reading comprehension (RC; Gough & Tunmer, 1986).

This theory provides an overview of the fundamental components that play a significant role in successful reading comprehension and need to be considered when assessing reading comprehension. It also explains the general variance in reading comprehension, and different profiles of reading comprehension difficulties (RC = 0):

- Poor decoders/good comprehenders (D = 0)
- Good decoders/poor comprehenders (LC = 0)
- Poor decoders/poor comprehenders (D = 0; LC = 0).

Thus, the SVR model indicates the variability between readers through different patterns of component skills. Accordingly, it is possible to distinguish less skilled and more skilled decoders, less skilled and more skilled comprehenders and all possible combinations of these two basic skills.

However, the authors of the SVR claim that their assumption may be incorrect in several cases:

- If decoding and language comprehension are intact, but reading comprehension is impaired.
- If decoding and reading comprehension are intact, but language comprehension is impaired.
- If decoding and language comprehension are impaired, but reading comprehension is intact.

Although such cases would disconfirm the SVR assumption, the authors of the model believe they are almost impossible. This can be explained by previous knowledge about the nature and relationship of decoding skills and language comprehension. Even though these skills are interconnected and their great role in successful reading and comprehension should always be recognized, they must also be distinguished and observed separately when measuring or assessing an individual's reading comprehension (component) skills.

The validity of the SVR model has been consistently supported across languages and developmental stages. Studies show that decoding and language comprehension together explain a substantial proportion of variance in reading comprehension — for example, about 60% in English-speaking first graders (Foorman et al., 2015, 2018) and up to 96% in early reading comprehension skills of Norwegian-speaking second graders (Lervåg et al., 2017).

Similar findings have been confirmed across orthographies: approximately 40-60% in non-transparent languages such as English and up to 80% in transparent languages such as Dominican Spanish (Foorman & Petscher, 2018; Sánchez-Vincitore et al., 2022).

Despite its robustness and validity, the SVR accounts for only part of the variance in reading comprehension, indicating the need for model expansion. Researchers (Cain et al., 2004; Johnston & Kirby, 2006; Geva & Farnia, 2012; Oakhill & Cain, 2012; Tunmer & Chapman, 2012) have proposed including additional domain-general components such as vocabulary, inferencing, and working memory, while others have noted that non-linguistic factors such as motivation and home environment are not captured at all (Joshi & Aaron, 2000; Catts, 2018; Snow, 2018). Specific components of the revised SVR model (Gough & Tunmer, 1986; Hoover & Tunmer, 2020) will be discussed in detail in the following sections.

Word recognition

According to the SVR framework, word recognition (WR) is a broader, multidimensional construct determined by both speed and accuracy, that is, fluency in word reading, primarily reflecting decoding efficiency (Gough & Tunmer, 1986; Catts & Hogan, 2003; LARRC, 2018). Word recognition refers to the outcome of automatised, unconscious, and efficient decoding processes. From a revised theoretical perspective (Hoover & Tunmer, 2020), WR is therefore conceptualised as a product of decoding processes that have become efficient and stabilised through experience. Thus, decoding and word recognition are inherently inseparable: successful word recognition presupposes decoding, and repeated decoding supports the development of fluent word recognition. Its later success is predicted by several decoding subskills, such as letter and sound identification, rapid automatized naming, and phonological processing skills (Bowey, 2005; Caravolas et al., 2012; Ramus et al., 2013), which will be discussed individually in the following sections. It represents one of the primary processes of language processing at the *word level* (Perfetti, 1999; Perfetti, 2001; Perfetti & Stafura, 2013; Knoepke & Richter, 2018).

The general role of WR in reading comprehension can be illustrated by the following example:

Imagine two students reading the following sentence: “*The little dog ran across the garden.*”

Student A reads each word quickly and accurately. As he does not need to break the words down into sounds, he can immediately consider the meaning — he imagines a little dog

running across the garden. His other cognitive resources (e.g. memory, attention) are directed towards comprehension and connecting the text with his own experience.

Student B struggles with the words *little* and *garden*, having to piece them together sound by sound (*li...ttl...e, g...ar...d...en*). While doing so, he loses the thread of the sentence, and by the time he reaches the end, he no longer remembers the beginning. His attention and working memory are consumed by the mechanics of reading rather than by understanding.

Student A has strong word recognition skills, while Student B reads less efficiently expending too much effort on decoding process and, as a result, has fewer mental resources left for understanding the text (Garnett, 2011; Hoover, 2024).

Word recognition develops along a continuum from effortful, decoding-dependent processing to automatised, unconscious access. According to Frith's (1985) developmental model, children move from an early logographic phase, in which words are recognised by salient visual cues rather than decoding, through an alphabetic phase, marked by effortful, sequential decoding, into an orthographic phase, in which familiar words are recognised directly and automatically. By fifth and sixth grade (approximately 10 to 12 years of age), typically developing readers are expected to have reached this orthographic phase, so word recognition no longer competes with comprehension for cognitive resources (cf. Student A). Continued reliance on decoding at this age, as with Student B, thus marks a deviation from the expected trajectory rather than a normative stage still in progress — which is what makes this age group informative for examining reading (dis)ability.

Decoding

Central to word recognition is decoding, which enables readers to convert written graphemes into their corresponding phonemes based on knowledge of grapheme-phoneme correspondence rules. Through repeated and successful decoding experiences, word recognition becomes increasingly automatic, freeing cognitive resources for higher-level comprehension processes (Gough & Tunmer, 1986; Foorman et al., 2016), such as in the previous example.

Decoding is a foundational component of early reading development and is essential for establishing fast and accurate word recognition (Ehri, 2005). It provides the mechanism by which learners access written language and therefore plays a critical role in reading comprehension (Wang et al., 2019). Within the SVR (Gough & Tunmer, 1986), decoding is

conceptualised as the core yet dynamic contributor to comprehension, particularly in the early years.

Beginning readers rely on phonological decoding, applying grapheme-phoneme conversion rules to both familiar words and novel-letter strings. As proficiency develops, skilled readers recognise familiar words automatically with minimal conscious attention to their constituent letters, rapidly accessing the mental lexicon and efficiently integrating orthographic and phonological information. The Dual-Route Cascaded (DRC) model explains these processes through two complementary routes: a non-lexical (phonological) route requiring grapheme-phoneme conversion, and a lexical (orthographic) route enabling direct retrieval of stored word representations (Coltheart et al., 2001, 2005, 2006). When readers encounter unfamiliar or low-frequency words, phonological recoding is activated; failure to decode results in a breakdown of comprehension, consistent with the SVR prediction that if decoding equals zero, comprehension also collapses (Gough & Tunmer, 1986). Conversely, familiar words are retrieved automatically via the lexical route, freeing cognitive resources for higher-level comprehension processes.

The extent to which readers rely on each route is shaped by word characteristics (e.g., frequency, familiarity), orthographic transparency, and reading experience. Orthographic transparency significantly affects the rate and reliability with which decoding skills are learned (Kirby & Savage, 2008). In transparent orthographies such as Spanish, Italian, and Croatian, where grapheme-phoneme correspondences are highly consistent, readers learn decoding skills more quickly and rely on a more balanced combination of lexical and sublexical processes (Zoccolotti et al., 2005; Cuetos & Barbon, 2006; Davies et al., 2007; Davies et al., 2013). This promotes fluency, and stable word recognition development (Zaretsky et al., 2009). By contrast, English contains numerous irregular spelling patterns, requiring more frequent activation of the sublexical route and greater reliance on memorisation of irregular high-frequency sight words (Zaretsky et al., 2009). Through repeated accurate decoding, unfamiliar words become stored as lexical entries, allowing for faster retrieval and reduced cognitive load during comprehension (National Reading Panel, 2000; Perfetti & Hart, 2002).

Evidence for the role of decoding in reading comprehension is mixed. Some studies report strong correlations between decoding and comprehension in the early grades, with word decoding speed accounting for up to 33% of the variance in reading comprehension even

after controlling for prior comprehension ability (de Jong & van der Leij, 2002), and word reading efficiency explaining up to 42% of variance in silent reading comprehension in dyslexic readers (Katzir et al., 2006). However, these relationships appear to be moderated by sample characteristics (e.g. typical language development vs. language-based disorders), measurement type, and orthographic transparency of the language, suggesting that the decoding-comprehension link is not uniform across contexts and populations. Longitudinal research similarly varies: some studies show somewhat stable decoding-language comprehension relationships over time (Foorman et al., 2018; Cain & Oakhill, 2011), whereas others demonstrate that decoding plays a stronger role in early literacy before declining in influence as language comprehension skills develop (Deacon & Kirby, 2004; Francis et al., 2005; Cutting & Scarborough, 2006; Vellutino et al., 2007; Keenan et al., 2008; Abbott et al., 2010). Importantly, developmental change is not linear — decoding appears to grow steadily but slows with age, while language comprehension peaks between second and third grades (Aarnoutse et al., 2001; Foorman et al., 2018). These shifting patterns reflect evolving interdependencies between RC components (Castles et al., 2018; Lervag et al., 2017).

Nevertheless, some findings are consistent across languages and methodologies. Children with poor decoding, particularly those with dyslexia, show persistent comprehension deficits (Shankweiler et al., 1996; Catts et al., 2003; Leach et al., 2003). Weaker readers also display stronger decoding-comprehension links than typically developing readers (Oakhill et al., 2003; Keenan et al., 2008). Thus, decoding serves as a key, but age- and skill-dependent, predictor that becomes less influential once children develop automaticity and higher-order comprehension abilities.

To clarify the role of decoding, research increasingly examines its component skills, primarily reading accuracy and reading speed, as independent predictors of comprehension. Reading accuracy, often measured as the number of correctly read real or non-words (i.e. pseudo-words specifically, pronounceable letter strings that follow orthographic rules but carry no meaning, also referred to as nonwords in some literature) without time constraints (Rønberg & Petersen, 2015), is especially informative in the early years. Studies of children with stronger and weaker reading skills, including those with language disorders, demonstrate the relevance of accuracy to comprehension (Bishop & Snowling, 2004; Nation et al., 2004; Nation & Norbury, 2005; Nation et al., 2010; Ricketts, 2011). However, accuracy must be measured at developmentally appropriate stages. During the first three years of schooling,

instruction emphasises accurate grapheme-phoneme conversion, but beyond this period accuracy becomes less informative as most children reach ceiling levels. The Language and Reading Research Consortium (LARRC, 2015) recommends assessing accuracy from Grades 1-3, and incorporating speed thereafter. Using accuracy as the sole measure past Grade 2 may artificially reduce its correlation with comprehension, whereas using speed too early may inflate correlations due to general processing-speed influences (Foorman et al., 2018).

On the other hand, reading fluency, serving as a strong developmental indicator of automated decoding, is defined as accurate, rapid, and prosodically appropriate reading (National Reading Panel, 2000). Fluency measures, including oral reading fluency, silent reading fluency, and retell fluency, consistently predict comprehension (Turkyilmaz et al., 2013). Studies incorporating timed reading demonstrate that fluency predicts reading comprehension from Grades 1 to 4, with correlations between .36 and .47 (Burns et al., 2011; Padeliadu & Antoniou, 2014). As with accuracy as an individual variable, fluency has the strongest influence in the early primary years, while in later years, this connection becomes moderate and ultimately weaker (Benson, 2008; Padeliadu & Antoniou, 2014).

Furthermore, of all the variables that define fluency, reading speed contributes the most to it. It is commonly operationalised as the number of words or non-words read per minute. Skilled adult readers typically achieve rates of around 250 words per minute with high levels (80%-90%) of comprehension (Carver, 1982; Nation, 1997; Torgesen et al., 1999). Speed may also be measured using isolated word lists (commonly used to distinguish reading fluency as a separate construct from reading accuracy or to examine its unique role in reading comprehension; Protopapas et al., 2012) or connected text (Adlof et al., 2006; Høien-Tengesdal & Høien, 2012; Kershaw & Schatschneider, 2012; Silverman et al., 2013). Cross-linguistic evidence underscores the importance of including fluency in models of reading development. Transparent orthographies facilitate rapid learning of decoding skills, making fluency especially informative as a predictor of later reading outcomes (Ziegler et al., 2010; Caravolas et al., 2012). Meta-analytic evidence shows that fluency in early reading stages provides valuable insight into automatised word recognition and real-time text processing (Florit & Cain, 2011). Findings from Croatian (Kelić et al., 2021; Keresteš et al., 2024; Kuvač Kraljević et al., 2024), Finnish (Torppa et al., 2016), and Estonian (Lerkkanen et al., 2016) confirm that both accuracy and speed are important for profiling reading development, with speed becoming increasingly important as children progress through the stages of reading acquisition and transition from accuracy-based to fluency-based word decoding. In

non-transparent orthographies, accuracy is a more sensitive early indicator, but fluency becomes more informative as decoding stabilises (García & Cain, 2014). Experimental studies also show that increasing reading speed can improve comprehension (Durukan, 2020). Given this evidence, Cain et al. (2015) propose that reading fluency should be explicitly integrated into the SVR framework as part of the word-recognition component. Moreover, developmental findings suggest a threshold effect in which, once minimal fluency is achieved, further gains in word recognition no longer predict comprehension (Benson, 2008).

Decoding is a foundational component of word recognition and a key determinant of later reading success. Although it may appear automatic in skilled readers, decoding relies on the coordinated development of several underlying subskills, including rapid automatized naming, and phonological processing skills, such as phonological awareness and phonological working memory (National Reading Panel, 2000). The following sections therefore examine the core subskills that support decoding and enable the development of rapid and accurate word reading.

Rapid automatized naming (RAN)

Rapid automatized naming (RAN; Wagner et al., 1993) reflects a complex, temporally constrained cascade of processes, including visual attention, perceptual processing, access to lexical representations, integration of conceptual and semantic information, as well as motor planning and articulatory execution (Wolf et al., 2000). Within the SVR framework, rapid automatized naming (RAN) is best understood as a background skill that supports decoding efficiency, particularly by facilitating the speed and automatization with which decoded phonological forms are accessed and produced from long-term memory (Wagner et al., 1993; Catts et al., 2002; Norton & Wolf, 2012).

In a typical RAN task (Denckla & Rudel, 1974), children rapidly name repeated familiar symbols (letters, digits or colours) arranged in a grid. Studies show that alphanumeric RAN (letters/ digits) correlates more strongly with reading outcome than non-alphanumeric RAN (colours), particularly once children are literate (Li et al., 2009). Performance is commonly measured by speed, accuracy, and sometimes by separating articulation and pause times or examining consistency across trials. In this way, RAN tasks mimic the demands of relatively fast reading. This is supported by findings from studies claiming that RAN is even more strongly related to reading fluency (specifically speed) than to decoding (Georgiou et al.,

2006; Torppa et al., 2012). However, certain findings (Savage et al., 2007; Moll et al., 2009) indicate that RAN predicts both word and nonword reading to a similar extent, thus challenging the view that its role can be explained solely through orthographic processing.

Nevertheless, Araújo et al. (2015) conducted a large meta-analysis (137 studies; 28,826 participants) on the relationship between RAN and reading performance. They found a moderate to strong overall correlation between RAN and reading ($r = .43$). Specifically, RAN is related to all major reading outcomes at various processing levels — single words, connected text, nonwords, and reading comprehension, with the strongest coefficients found for real word and text reading. Importantly, Savage et al. (2007) emphasize that RAN is multi-componential: it shares variance with processing speed and phonological/decoding measures yet often contributes unique (small to moderate) variance to literacy after these are controlled. Examining the SVR model in the Finnish language context, a language with transparent orthography, Torppa et al. (2016) found RAN, alongside phonological awareness and letter knowledge, to be robust predictor of decoding and reading comprehension (see also Kirby et al., 2003; Lepola et al., 2005; Torppa et al., 2010). Similarly, Kelić et al. (2021) found RAN to be the strongest predictor of both reading accuracy and speed in Croatian, also a language with transparent orthography. In general, RAN remains a significant correlate regardless of language depth, with some moderation by transparency and grade (Wolf et al., 2000). However, in more transparent languages, accuracy plateaus early; speed (and thus RAN) becomes increasingly informative. Longitudinal and cross-linguistic studies in transparent orthographies show decoding/ fluency building comprehension early on, alongside language comprehension, which is consistent with the SVR extended with fluency (Torppa et al., 2016). Also, it seems that RAN has a dual role in reading comprehension. First, its primary role is indirect (via decoding and fluency), showing smaller correlation for reading comprehension than for word/text reading (Araújo et al., 2015). Thus, it supports the efficiency with which lower-level processes deliver input to comprehension. Second, when decoding/fluency and language are modelled, particularly in typical readers and later grades, some executive-control components of RAN (e.g. response stability, pause variability; Li et al., 2009) play a role in reading comprehension.

Phonological processing skills

Phonological awareness

For successful phonological processing (sub-lexical route), the conscious awareness of phonological units of various sizes — syllables, onsets, rimes, and phonemes — is essential alongside working memory processes (Wagner and Torgesen, 1987). This ability is broadly termed phonological awareness, with phonemic awareness referring more narrowly to awareness at the level of the individual phoneme, one of its constituent grain sizes. For the purposes of this dissertation, phonological awareness (PA) is used as an umbrella term that also encompasses phonemic awareness, and the two are not treated as separate constructs in the discussion that follows. Its role in acquiring reading skills and literacy development is considered even causal (Scarborough et al., 2001; Ramus et al., 2013). The development of decoding depends on phonological awareness, otherwise, the reader may struggle to apply letter-sound rules when decoding print (Adams, 1990; Goswami and Bryant, 1990).

Phonological awareness (PA) is typically measured through tasks that assess an individual's ability to recognise and manipulate (change or rearrange) phoneme structures in words and to reflect on how words are composed of different phonemic structures, most commonly targeting the phonemic level specifically. Common tasks used in research include phoneme identification, deletion, blending, segmentation, manipulation and spoonerisms. The choice of task depends on the participant's age, research questions, and language structure (Nation & Hulme, 1997; Pufpaff, 2009; Mott & Rutherford, 2012; Ramus et al., 2013; Kelić, 2017).

PA contributes to reading comprehension both developmentally and theoretically.

Developmentally, it is a strong predictor of early reading accuracy and speed. However, its influence changes as children progress from learning to read to reading for meaning, with effects moderated by language transparency: stronger and longer lasting in non-transparent orthographies than in transparent ones (Goswami, 2000; Carroll & Snowling, 2004; Ziegler et al., 2010; Kelić, 2017; Kelić et al., 2021; Milledge & Blythe, 2019). In Croatian children with adequate reading skills, after three years of reading instruction, reading is largely supported orthographically. However, the phonological route is still used when reading unfamiliar words or pseudowords (Kelić, 2017; Kelić et al., 2021). It is argued that phonological decoding route may be the primary route in highly transparent languages (Kirby et al., 2003), and thus PA remains a strong predictor of early reading outcomes. Interestingly, Keresteš et al. (2019), by tracking the development of reading and writing skills over four years in the domains of phonological awareness, working memory, and morphosyntactic awareness,

demonstrated that phonological awareness significantly contributed to explaining variance in reading comprehension, with no significant impact on word decoding speed.

Theoretically, PA supports reading outcomes because decoding relies on matching graphemes to phoneme, and difficulties with phoneme identification can disrupt this process (Goswami & Bryant, 1990). Hogan et al. (2005) found that impact of PA on word reading decreases by Grade 4 for typically developing children. However, Sánchez-Vincitore et al. (2022) examined the contribution of PA to word recognition within SVR model and found it significant impact on word recognition (WR) in Dominican students, with its effect slightly increasing from second to fourth grade. This suggests that PA and thus WR remain the main drivers of reading comprehension across grades, while the role of oral language comprehension becomes more important as children advance. To examine potential SVR predictors of reading comprehension in Finnish, Torppa et al. (2010) found that PA, alongside RAN and letter knowledge, strongly supports both decoding and reading comprehension in Grade 1, however, their impact on later reading outcomes is indirect. Instead, these early abilities shape later reading performance. Other research with Spanish readers shows that PA (particularly awareness of syllables) relates moderately to word reading, pseudoword reading, and comprehension (Míguez-Álvarez et al., 2022). As reading skills advance, phonology is used in a less obvious way: eye-tracking studies show children move from consciously sounding out graphemes (decoding) to quickly and automatically activating a word's sound when they see it (phonological recoding; Milledge & Blythe, 2019). Thus, early decoding components serve as a bridge between visual word recognition and higher-level comprehension, supporting the transition from controlled to automatized reading processes. This pattern is consistent with the Simple View of Reading, according to which proficient comprehension emerges from the interaction of decoding and language skills. The role of PA in reading outcomes is also evident in deficits in phonological awareness underlying decoding difficulties in dyslexia, which can cascade into comprehension problems (Tunmer & Greaney, 2010; Ramus et al., 2003a, 2013b).

(Phonological) working memory

Some components are not directly represented in the SVR (Gough & Tunmer, 1986); however, research shows their potential significant contribution to reading comprehension processes. One such component is phonological working memory (PWM), an important underlying component of decoding process. It allows individuals to store information about phonological units while simultaneously processing them (retaining, manipulating, and

retrieving), thus emphasizing its role in phonological processing (Baddeley & Hitch, 1974; Florit et al., 2009; Swanson et al., 2009).

In Baddeley and Hitch's (1974) widely cited model, working memory comprises the central executive, the visuospatial sketchpad, and the phonological loop. The central executive regulates both subsystems: the visuospatial sketchpad, which temporarily maintains and manipulates visual and spatial information, and the phonological loop, which temporarily stores and retrieves precise phonological word representations (Baddeley, 2000; Florit et al., 2009). PWM is often assessed through nonword repetition — a task linked to word recognition and learning, as it measures phonological processing independently of lexical knowledge — or through digit span (Snowling, 1981; Snowling et al., 1986; Snowling et al., 1991; Gathercole & Pickering, 2001; Coady & Evans, 2008). However, studies show that digit and pseudoword repetition are not equivalent: children with language-based disorders perform particularly poorly on pseudoword repetition compared to typically developing peers (Archibald & Gathercole, 2007). Digit span correlates more strongly with syllable repetition than with pseudoword repetition (Kelić et al., 2016), suggesting that pseudoword repetition depends not only on PWM but also on the phonological complexity of the stimuli.

Deficits in the PWM are considered central to difficulties in reading learning, particularly word recognition and decoding (van Dyke, 2021), and are characteristic of language-based disorders such as dyslexia (Snowling et al., 1986; Snowling et al., 1991; Swanson et al., 2009; Ramus & Ahissar, 2012; Ramus et al., 2013). Successful phoneme-grapheme mapping requires retaining phonological units in short-term memory; failure to do so hinders reading, vocabulary growth, and the linking of phonological and orthographic forms (Gathercole & Baddeley, 1993). Furthermore, the success of phonological awareness depends on both the phonological loop (Oakhill & Kyle, 2000) and the central executive (Alloway, 2009). Efficient WM further underpins phonological coding and decoding (Wagner et al., 1994). If all remaining limited working memory resources are redirected to decoding, the process of building coherent mental representations of text is disrupted (Gough et al., 1996; Verhoeven & Van Leeuwe, 2008; Tighe & Schatschneider, 2014).

A key debate in this field concerns whether working memory capacity or processing efficiency plays a greater role in reading comprehension, and whether it should be considered a unique contributor to the SVR model (Daneman & Merikle, 1996; Cain et al., 2004; Cain, 2006; Carretti et al., 2009; Hoover & Tunmer, 2020). Studies using behavioral and

neuroscientific methods, such as reading span (Daneman & Carpenter, 1980) and operation span tasks (Unsworth et al., 2005), have shown that working memory capacity predicts inference-making, a key element of comprehension (see page 24, Rai et al., 2011; Bohn-Gettler & Kendeou, 2014; Perez et al., 2015; Hamilton et al., 2016). Daneman and Merikle (1996) argued that comprehension requires not only temporary storage but also integration of new information with prior knowledge and the active construction of meaning. Similar findings have been reported in children with and without reading difficulties, with processing capacity for linguistic information identified as the main source of individual variation (Cain et al., 2004; Cain, 2006; Carretti et al., 2009). Other research points to the central executive (Andersson, 2010) and the semantic storage (Nouwens et al., 2017) as critical components. In contrast, Ding & Wang (2006), Zhang et al. (2012) and Nicolielo-Carrilho et al. (2018) highlight the role of phonological processing, suggesting that the phonological loop may be especially important in reading comprehension, with evidence across languages with non-transparent orthographies (e.g., English) and ideographic systems (e.g., Chinese). By temporarily storing and processing phonological information, it supports word decoding, meaning access, and integration with prior knowledge, all necessary for building coherent text representations.

However, the extent to which phonological loop capacity and phonological processing contribute to reading comprehension, particularly inference-making, remains debated (Ding & Wang, 2006; Wang et al., 2021). As Freed et al. (2017) note, individual differences are difficult to isolate because task performance reflects multiple processes, making it challenging to determine whether a specific working memory component uniquely predicts comprehension.

Language (linguistic/ listening) comprehension

Comprehending written discourse depends on the knowledge of the language being read — that is, on the reader's linguistic competence (vocabulary, grammar, and discourse-level knowledge), as distinct from knowledge of the orthographic system used to encode it. In the SVR model, this component is often referred to as *linguistic*, *listening*, and/or *language* comprehension. The term *linguistic* comprehension was first introduced to denote the understanding of spoken language (Gough & Tunmer, 1986). This term was later replaced by *listening* comprehension, referring specifically to understanding text read *aloud*. *Listening* comprehension (LC) relies on the same language processes that underlie reading (text) comprehension (RC), but without the additional burden of decoding written symbols (Gough

et al., 1996; Hoover & Gough, 1990). The processes used to understand LC and RC largely overlap. Brain imaging research further shows that listening to and reading the same text trigger similar neural activity, suggesting that meaning is represented consistently in the brain regardless of the input source (Deniz et al., 2019). However, some differences between spoken and written language do exist and must be acknowledged, especially in research context (e.g. for written language — pauses, revisions, shared context, formal language form, decontextualization; Kirby & Savage, 2008). As Hoover and Gough (1990) emphasize, a valid measure of linguistic comprehension should test a person's ability to understand spoken language, such as answering questions about a story that was heard. Similarly, reading comprehension should assess this same understanding, but through written rather than spoken text.

Furthermore, the term *language* comprehension (Cragg & Nation, 2006) is used more broadly to encompass the full range of processes involved in understanding language. Tunmer & Hoover (2019) point out that it is important to distinguish between the broader concept of language comprehension and the specific tasks used to test it. Listening and language comprehension refer to the same underlying construct and are therefore often used interchangeably. Studies show they are highly correlated in children from preschool to third grade, effectively functioning as a single oral-language factor (LARRC, 2018). However, the way they are measured can differ. For example, *listening* comprehension is commonly measured by retelling the read-aloud story, inference-making tasks, picture-pointing, etc., while *language* comprehension additionally covers measures (mainly standardized tests) of (receptive) vocabulary, grammar, and discourse skills used to understand language in any modality (Catts et al., 2006; Clarke et al., 2010; Verhoeven et al., 2012; Landi & Ryherd, 2017; LARRC, 2018), thus influencing the final data interpretation. For this reason, Tunmer & Chapman (2012), Hoover & Tunmer (2018), and Hoover & Tunmer (2020) explain that the term *language* comprehension is the most suitable label to use within the SVR model. Therefore, in this dissertation, the term language comprehension will be used when referring to the LC component.

In later grades, LC explains most of the variability in reading comprehension (RC; Catts, 2018). As children become better decoders and begin to *learn through reading* in later school years, the role of decoding skills in reading comprehension becomes less significant because language comprehension skills become more dominant (Verhoeven & van Leeuwe, 2012; Geva & Farnia, 2012; Cain, 2015; Lervåg et al., 2017). Language comprehension has a

predictive role in both the early and later developmental stages of reading comprehension (as does decoding), but it precedes decoding in later school years, around third or fourth grade (commonly after reading automatization, see Geva & Farnia, 2012; Cain, 2015; LARRC, 2015; Lervåg et al., 2017). It is important to note that these findings primarily reflect developmental changes in children learning to read languages with non-transparent orthography. However, in transparent orthographies this shift can occur even earlier. Müller & Brady (2001) found stronger relationship between LC and RC in beginner readers of transparent languages, compared to those of non-transparent languages. Florit & Cain (2011) presented a meta-analysis of the validity of SVR for beginner readers, comparing learners of transparent and non-transparent languages. They found that for English readers, the LC only moderately predicted RC, but LC was a stronger predictor of RC in transparent orthographies for beginner readers (first and second grade) than decoding (accuracy, not fluency). In cross-sectional study of Italian-speaking children, Tobbia & Bonifacci (2015) confirmed the primary contribution of the LC to RC in the primary grades. Evidence from Croatian, based on a study of second-grade Croatian-speaking primary school typical readers (Kuvač Kraljević et al., 2021) found the LC to be the sole predictor of reading comprehension, while both decoding and LC were significant predictors among children classified as poor readers. These findings suggest that the LC component may be even more important in the earlier educational years for readers of transparent languages and for skilled readers compared to those with reading difficulties.

Language comprehension is not unidimensional. According to the literature, different components of the receptive language, from both lower and higher levels of processing, inform and predict language comprehension (LARRC, 2018). Components considered part of language comprehension with the role in RC refer to word-, sentence-, and discourse-level of language processing (Perfetti, 2001; Rayner & Reichle, 2010), such as vocabulary knowledge, semantic integration, elaboration of syntactic information, narrative comprehension and background knowledge (Tanenhaus et al., 1995; Daneman & Merikle, 1996; Hagtvet, 2003). Studies show that preschool abilities in lower-level processing skills — vocabulary, grammar, and higher/discourse-level components — predict children's language comprehension in third grade, which in turn supports their reading comprehension (Foorman et al., 2015; LARRC & Chiu, 2018).

Vocabulary

Over the past decade, researchers have shown significant interest in studying the relationship between vocabulary and reading comprehension, as well as individual differences in both constructs, depending on language typology and age (Protopapas et al., 2013; Cain & Oakhill, 2014; Cho et al., 2019). Vocabulary was recognized as a key component of the broader construct of language comprehension only after the revisions to the SVR. Previously considered part of word meaning access or as a separate element, its role is now recognized as fundamental to constructing meaning from text. However, it is still often examined in isolation — a tendency unsurprising given the broad nature of the language comprehension construct.

Vocabulary knowledge strongly predicts comprehension, both concurrently and longitudinally, throughout early childhood (Muter et al., 2004; Florit et al., 2011; Lepola et al., 2012; Tompkins et al., 2013). Research consistently confirms that knowledge of word meanings underpins the ability to interpret and integrate information in narratives (Silva & Cain, 2015). Studies examining the SVR model within and beyond English, including French and Greek, demonstrate that vocabulary contributes uniquely to reading comprehension and operates as part of the broader language comprehension construct (Tunmer & Chapman, 2012; Protopapas et al., 2013; Braze et al., 2016; Massonnié et al., 2019). Overall, vocabulary does not function as an isolated skill but as a central mechanism through which language comprehension supports the understanding of written discourse. However, recent studies have questioned the idea that vocabulary functions solely as a component of language comprehension, suggesting it may also exert a more autonomous influence on reading and text processing (Ouellette, 2006; Tunmer & Chapman, 2011; Cho et al., 2019). Within the Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti, 2007) and the Reading System Framework (Perfetti & Stafura, 2014), vocabulary development enhances the precision of lexical representations and strengthens syntactic processing, thereby improving comprehension efficiency. Empirical findings show that vocabulary affects reading comprehension both directly — by facilitating faster and more accurate semantic access, and indirectly — through its role in decoding and broader language comprehension. This aligns with the SVR model (Gough & Tunmer, 1986), in which decoding and language comprehension jointly shape reading outcomes. A rich vocabulary supports this interplay by promoting accurate word recognition, efficient integration of lexical and syntactic information, and ultimately, deeper comprehension of written discourse. Notably, the strength

and nature of this relationship vary across languages, ages, linguistic backgrounds, and research methodologies (Ouellette, 2006; Tunmer & Chapman, 2011; Cho et al., 2019; Verhoeven et al., 2019; Dujardin et al., 2023).

When examining the role of vocabulary in reading comprehension, it is essential to recognize that vocabulary is multidimensional, encompassing both breadth and depth (Ouellette, 2006; Cain, 2010). Vocabulary breadth refers to the number of known word forms stored in the mental lexicon, whereas depth reflects the richness of lexical knowledge — the ability to access multiple meanings, semantic relations, and appropriate contextual uses (Swart et al., 2017; Peti-Stantić, 2022). Within Perfetti’s Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti, 2007), depth represents the quality of lexical representations crucial for comprehension. Breadth is commonly assessed through word recognition or meaning identification tasks, while depth is measured through more complex methods such as defining words or explaining semantic relationships (Tannenbaum et al., 2006; Proctor et al., 2012). These dimensions are strongly interrelated: expanding vocabulary size supports the development of deeper lexical understanding, and vice versa (Read, 2004). Although some authors argue that breadth and depth represent the same construct measured differently (Vermeer, 2001; McGregor et al., 2013), regression studies show that depth explains unique variance beyond vocabulary size (Qian, 1999a, 2000b), indicating that distinguishing between them remains theoretically and empirically justified.

Studies show that both vocabulary breadth and depth influence reading comprehension in children and adults (Roth et al., 2002; Nation & Snowling, 2004; Ouellette, 2006; Binder et al., 2017). Vocabulary breadth supports decoding and word recognition, while depth — reflecting the richness of semantic and contextual knowledge — uniquely predicts reading comprehension. Several hypotheses have been proposed to explain this relationship. The *Instrumental Hypothesis* (Anderson & Freebody, 1981) suggests that knowing the meanings of most words in a text enables comprehension, although later research indicates that this link is not fully explanatory (Elleman et al., 2009). The *Aptitude Hypothesis* (Anderson & Freebody, 1981) attributes the connection between vocabulary and comprehension to underlying cognitive abilities such as inference, metalinguistic awareness, and working memory, while the *Knowledge Hypothesis* emphasizes background and conceptual knowledge as mediators (Wagner & Meros, 2010).

Further studies have shown that comprehension skills are closely linked to lexical-semantic knowledge (Nation & Snowling, 1999; Cain et al., 2009; Weekes et al., 2008; Petersen & Elbro, 2009), supporting the *Access Hypothesis* (Mezynski, 1983), which highlights the importance of automatic access to word meaning. This aligns with Perfetti's Lexical Quality Hypothesis and Reading System Framework (Perfetti & Stafura, 2014). The *Reciprocal Hypothesis* (Stanovich, 1986) proposes a bidirectional relationship between vocabulary and comprehension, described by the Matthew Effect: children with stronger early vocabulary and comprehension skills continue to progress more rapidly, while those with weaker skills fall further behind. Although findings on this effect are mixed due to methodological differences (Aarnoutse & van Leeuwe, 2000; Cain & Oakhill, 2011; Kempe et al., 2011; Duff et al., 2015), evidence suggests that early language proficiency influences later vocabulary growth and reading success.

Environmental and developmental factors also play a role. Early language exposure strongly predicts vocabulary outcomes (Hart & Risley, 1995), while persistent language difficulties can lead to long-term deficits (Stothard et al., 1998). Although methodological inconsistencies remain, the accumulated evidence indicates that early differences in vocabulary and reading abilities interact dynamically, reinforcing each other over time. Clarifying these relationships — particularly the distinct roles of vocabulary breadth and depth — remains essential for refining theoretical models and informing educational and therapeutic practice.

Grammaticality and (Meta)linguistic Awareness

Syntactic and morphological skills and knowledge are often identified as additional factors contributing to the complex, multidimensional nature of language comprehension (LC) in the SVR model (Gough & Tunmer, 1986). Morpho-syntactic awareness, the ability to understand and manipulate sentence and word structure, has been shown to be a strong predictor of reading comprehension, including in transparent languages such as Croatian (Keresteš et al., 2019). According to Levelt's model (1999), sentence processing requires the activation of lemmas that carry semantic and grammatical information, their alignment within syntactic structures, and the integration of word order and inflectional features. Effective comprehension thus depends on the coordination of lexical access, syntactic relations, phonological and morphological processing to form coherent representations of meaning (Perfetti & Stafura, 2014).

Weaker grammatical and syntactic skills are consistently associated with reading comprehension difficulties (Florit & Cain, 2011; Hjetland et al., 2020). However, evidence regarding the extent of grammar's contribution remains mixed — some studies suggest a direct link (Tong et al., 2014), while others indicate indirect effects mediated by vocabulary and working memory (Cain, 2007). These inconsistencies partly reflect methodological differences in assessing grammatical knowledge. Studies using metalinguistic tasks (e.g., grammaticality judgements) tend to show stronger associations with comprehension than those using purely structural measures, such as standardised morphosyntax production tasks or sentence repetition paradigms, which assess the accuracy of grammatical form without requiring explicit metalinguistic awareness (Muter et al., 2004; Potocki et al., 2013).

Longitudinal research further suggests that the role of grammar evolves developmentally. In early schooling, grammar and vocabulary together predict reading comprehension (Muter et al., 2004), whereas in later years, higher-level skills such as inference-making and discourse integration become more influential (Oakhill & Cain, 2012). Models of reading development also differ: some support a unified language factor combining vocabulary and grammar (Adlof et al., 2006; Protopapas et al., 2012), while others distinguish separate structures in lower and upper grades (Tomblin & Zhang, 2006; Foorman et al., 2015; LARRC, 2015). Despite these differences, poor comprehenders consistently show weaker performance in grammatical correction, morphemic analysis, and syntactic complexity tasks (Tunmer et al., 1987; Catts et al., 2006; Adlof & Catts, 2015), reflecting an underlying connection between oral and written language. Overall, grammatical knowledge contributes to reading comprehension both directly — through syntactic and morphological processing — and indirectly — via vocabulary and broader linguistic mechanisms. Its impact appears to shift with age, language development, and task demands, emphasizing the need for refined approaches that examine specific grammatical features, such as cohesive structures, to better understand their role in reading comprehension (Silva & Cain, 2015).

Discourse-level processing components

Higher-level language processing commonly refers to sentence- and discourse- level processes (Perfetti & Stafura, 2013; Kim et al., 2015; Knoepke & Richter, 2018). Discourse processing requires efficient access to phonological, syntactic, and semantic word properties, as well as the integration of word meanings into coherent mental representations that reflect their structural relations (Perfetti & Stafura, 2014). To fully understand written text, readers must construct coherent, contextually appropriate mental, situation models of discourse

(*Situation model*, Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983; *Coherent mental model*, Johnson-Laird, 1980), enabling them to draw both explicit and implicit inferences from the information presented. These models integrate explicit textual information with the reader's prior knowledge, allowing for a richer representation than the text-base alone. To achieve coherence, readers must link adjacent and distant propositions, draw inferences, and monitor both plausibility and their own understanding (Graesser et al., 1994; Oakhill et al., 2003; Nation, 2005).

Discourse-level language skills, such as understanding narratives, making inferences, and monitoring comprehension, emerge early in development. Evidence shows that comprehension monitoring appears by around three years of age (Skarakis-Doyle & Dempsey, 2008), and even early word learning relies on inferential processing (Tomasello et al., 2007). Studies further indicate that these discourse-level abilities, together with vocabulary and grammar, each make unique contributions to language comprehension in preschool children (Lepola et al., 2012; Silva & Cain, 2014). From about first grade onward, higher-level language skills develop as a distinct ability, separate from vocabulary and grammar (LARRC, 2015). These skills make an independent contribution to reading comprehension, so even when vocabulary and grammar are strong, higher-level language processes still play a unique role in understanding text (Oakhill & Cain, 2012).

Linguistic markers of coherence, such as connectives and referential ties, play an important role in guiding comprehension. Studies show that poor readers have difficulty recognizing or using these cues and often rely heavily on their presence to understand connections between ideas (Geva & Ryan, 1985; Cain et al., 2005; Hall et al., 2014). Skilled readers, in contrast, infer coherence relations even when such signals are absent.

Kim et al. (2015) explored how different dimensions of discourse-level oral language, specifically language comprehension, oral retell, and oral production, relate to reading comprehension and written composition in Korean-speaking first graders. Their central aim was to clarify whether these discourse-level skills represent distinct constructs or a unified oral-language factor and to examine their predictive power for literacy outcomes. Using confirmatory factor analysis, they found that a bi-factor model best explained the data, revealing a general discourse-level oral language factor underlying all tasks, alongside distinct receptive (listening) and expressive (retell) dimensions. The study also found developmental sensitivity: discourse-level oral language was already established in the early

primary years, suggesting that comprehension processes at the discourse level emerge early and scaffold later literacy skills. The general discourse-level oral language skill strongly predicted reading comprehension, accounting for most of the explained variance (75%), while unique effects of listening and retelling were not significant. This suggests that reading comprehension depends on an integrated discourse-level competence that combines linguistic and cognitive processes, such as vocabulary, syntax, inference-making, and coherence building, rather than on isolated receptive or expressive skills. These findings extend the SVR model by showing that language comprehension involves a broader discourse-level construct encompassing both comprehension and production. Discourse-level oral language, as a multidimensional construct, provides the foundation for constructing coherent mental representations of text, consistent with models of text integration (Kintsch, 1988) and the Lexical Quality Hypothesis (Perfetti, 2007), which proposes that high-quality lexical representations contribute to more efficient integration of linguistic information during text comprehension. This integrative ability constitutes a critical component of the language comprehension factor within the SVR and is central to explaining individual differences in reading comprehension performance.

Reading comprehension

Reading comprehension involves integrating the meanings of all words within a sentence into coherent mental representations, considering their specific syntactic and semantic structures (Perfetti & Stafura, 2014), and requires the ability to interpret implied meaning accurately, while monitoring the comprehension process. Inference-making is one of the main components these processes are supported by (Perfetti et al., 2013; LARRC, 2018).

Inferencing

To understand language, it is often necessary to make inferences (Nation, 2005).

In written discourse i.e text, much of the meaning is implied rather than stated directly. Readers must make inferences — retrieving explicit information and integrating implicit cues, to build a coherent understanding of the text (Cain & Oakhill, 1999; van den Broek et al., 2005; Gough et al., 2018). This process involves linking textual information with prior knowledge to form a mental model of meaning (Kintsch, 1998; Perfetti & Stafura, 2014), a skill essential for effective comprehension of both written and spoken discourse. According to the Construction-Integration model (C-I; Kintsch, 1998), readers continuously connect

related parts of discourse to maintain referential coherence and construct a mental model. When encountering unclear references, they infer connections to earlier information.

Thus, inferencing serves two main purposes: linking information within the text and integrating it with the reader's background knowledge (Cain et al., 2004a, 2004b). Through these coherence inferences, readers form mental models that represent the text's structure and meaning.

For written discourse to be meaningful rather than a collection of disconnected sentences, readers must generate *co-referential* inferences that link ideas across the text (Perfetti et al., 2001). Researchers classify inferences in various ways — by their role in maintaining coherence, timing of activation, reasoning type, comprehension goals, or causal relationships (Kendeou, 2015; Gough et al., 2018). Despite differences in terminology, most agree that *coherence* inferences are essential for maintaining text integrity and meaning (Cain & Oakhill, 1999, 2001, 2014). These include *local* inferences, which connect adjacent sentences or fill informational gaps using prior knowledge, and *global* inferences, which integrate ideas across larger text units to derive broader meaning or themes (Gygax et al., 2004; Oakhill & Cain, 2018). At the local level, readers resolve referential ambiguity and link anaphors to their antecedents (e.g., recognizing who “she” refers to in a story), while at the global level, they synthesize information to infer abstract concepts such as character traits or moral lessons. Inferences are often divided into *bridging* (connective) inferences, which link explicit information and explain causal relations, and *elaborative* (predictive) inferences, which enrich the mental model but are not essential for coherence (Singer et al., 1992). Predictive or elaborative inferences — those that anticipate future events — occur less frequently, as they require greater cognitive effort and are context-dependent (Snow, 2002; Hemforth et al., 2014). Graesser et al. (1994) further distinguish between *online* inferences, generated during reading, and *offline* inferences, made afterwards when explicitly prompted. It is assumed that the ease and accuracy of inference-making depend not only on cognitive skill but also on text structure and content (Graesser et al., 2004; Lorch, 2015). Narrative texts, which focus on characters and their actions, naturally support inference generation because readers can rely on familiar, everyday situations and established story schemas (Aksu-Koç & Nicolopoulou, 2015). When information is missing, skilled readers draw upon story grammar knowledge to fill gaps and maintain coherence (Graesser et al., 2004). Moreover, the frequent use of pronouns and anaphoric forms in narratives facilitates

referential connections between sentences, enabling readers to track characters and events more efficiently (Graesser & McNamara, 2011; Walkington et al., 2019).

Inference generation is among the strongest predictors of reading comprehension development and achievement (Kendeou et al., 2009; Oakhill & Cain, 2012), from early childhood onwards. In children aged 4-6, inference skills predict narrative and language comprehension alongside vocabulary (Kendeou et al., 2009; Florit et al., 2011; Tompkins et al., 2013), and longitudinal findings confirm their lasting influence (Lepola et al., 2012). In older children, inference-making uniquely predicts both language and reading comprehension, even beyond vocabulary and grammar (Cain et al., 2004; Oakhill & Cain, 2012). Skilled readers integrate text information dynamically, using background knowledge to build coherent situation models, while poor comprehenders struggle to connect ideas, monitor understanding, and suppress irrelevant details (Graesser et al., 1994; Côté et al., 1998). Research consistently shows that local inferences — those linking adjacent sentences or referents — are easier and processed automatically during reading (Duffy & Rayner, 1990; McMaster et al., 2014), whereas global or elaborative inferences, which require integrating broader text segments or predicting events, are more cognitively demanding (Snow, 2002; Yang et al., 2007). Skilled comprehenders tend to generate meaningful causal or goal-related inferences, whereas less skilled readers often fail to make necessary connections, even when given access to the text (Cain & Oakhill, 1999). Their difficulty lies not in memory, but in strategy use and activating relevant knowledge during reading. Some propose that the key difference between good and poor comprehenders is in how they process literal and inferential information. Skilled readers accurately recall explicitly stated details (literal comprehension) and effectively draw inferences that explain or link ideas within the text, focusing on causal and goal-related relationships (Graesser et al., 1994; Côté et al., 1998). Poor comprehenders, in contrast, often generate irrelevant or superficial associations and struggle to integrate information or filter out unnecessary details. For example, Oakhill's studies (1982, 1984) showed that while both groups perform similarly on tasks requiring verbatim recall, skilled readers outperform poor comprehenders on literal and inferential questions, demonstrating stronger integration of text-based and background knowledge. Further research confirmed that less skilled readers have difficulty forming both text-connecting and gap-filling inferences, even when allowed to reread the text (Cain & Oakhill, 1999). These difficulties reflect limited spontaneous inference-making and weak comprehension monitoring rather than memory deficits.

Studies using offline measures (e.g., question-answering tasks) confirm that children generate both local and global inferences, although the latter require greater working memory and inferential effort (Albrecht & O'Brien, 1993). Online methods such as eye-tracking reveal that children form local inferences spontaneously but with delayed processing compared to adults (van der Schoot et al., 2008; Joseph et al., 2015, 2021). In contrast, global inference-making occurs only under low task demands or when coherence is disrupted.

These findings suggest that inference generation is gradual, cognitively demanding, and potentially dependent on text coherence, lexical complexity, and reader skill. Lexical and syntactic factors — such as word properties, sentence complexity, and discourse cohesion — may affect how efficiently readers form inferences (Dahl et al., 2021). Skilled readers, equipped with higher-quality lexical representations, retrieve and integrate meanings more efficiently, facilitating deeper comprehension (Perfetti et al., 2008). In contrast, for weaker readers limited lexical precision slows processing and disrupts coherence building.

This integrative perspective on the components supporting reading comprehension complements the Reading System Framework (Perfetti & Stafura, 2014) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti, 2007) by emphasizing the role of high-quality lexical representations within coherent discourse structures.

1.2.2 Reading System Framework and Lexical Quality Hypothesis

The Reading System Framework (RSF) proposed by Perfetti and Stafura (2014) builds on the SVR model by emphasizing the role of word-level processes and the quality of mental word representations — phonological (sound), orthographic (form), and semantic (meaning) — in reading comprehension. It highlights how these representations interact with broader linguistic and cognitive processes to construct meaning. According to the RSF, reading comprehension involves two main subprocesses: word identification and word-to-text integration, both supported by background knowledge (Figure 1.). Readers first decode words, then integrate their meanings into coherent sentences and, ultimately, into a unified text representation. Throughout this process, they draw on prior knowledge, reading goals, and contextual understanding to build a comprehensive situation model (Perfetti & Stafura, 2014; Bruggink et al., 2022).

By focusing on how lexical knowledge interacts with higher-order comprehension mechanisms, the RSF provides a more detailed account of how individual differences in word knowledge and processing quality contribute to overall reading success. It therefore serves as

a framework for examining how vocabulary, decoding, and integration processes jointly support the construction of meaning in text comprehension.

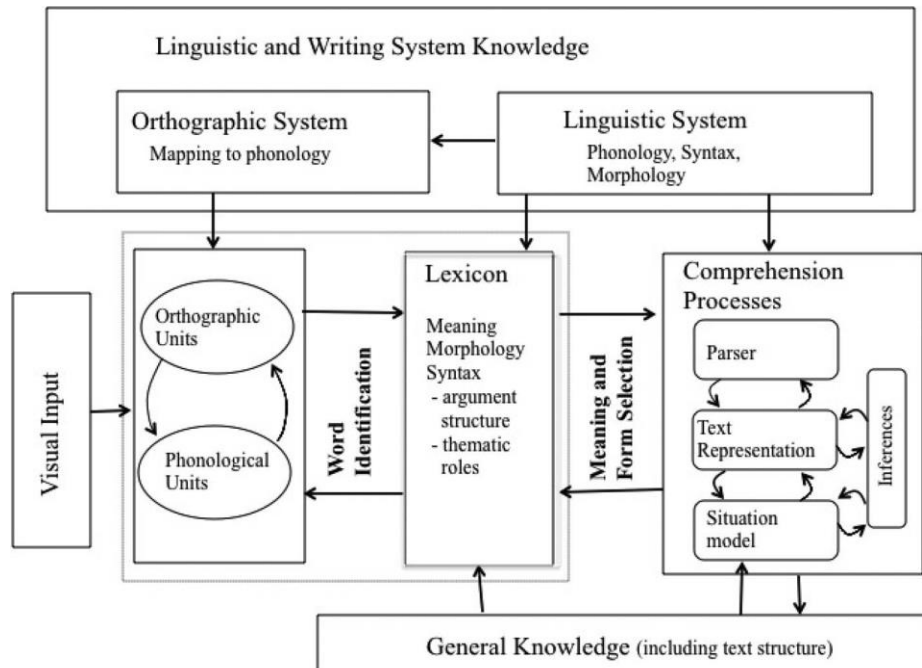


Figure 1. The Reading System Framework (Perfetti & Stafura, 2014)

In the RSF, reading comprehension begins with word identification, a process that largely overlaps with the word recognition-decoding component of the Simple View of Reading. In both models, decoding refers to the efficient recognition of written words through access to their orthographic and phonological representations. However, while the SVR conceptualises decoding as a core component of reading comprehension, the RSF further specifies: (1) how word identification enables access to lexical representations and (2) supports the activation of word meaning (Perfetti & Stafura, 2014). Skilled readers perform these steps accurately and automatically, allowing them to devote cognitive resources to higher-level comprehension (Wang et al., 2019). This process depends on the mental lexicon, which stores knowledge about a word’s spelling, pronunciation, and meaning, forming the foundation for comprehension.

The next process, word-to-text integration, involves combining individual word meanings into propositions — basic units of meaning — that are synthesized into sentence and text-level representations (Perfetti & Stafura, 2014). These representations enable the construction of coherent text-base and situation models. Research shows that skilled readers integrate word meanings more quickly and efficiently than less skilled readers, and that the quality of lexical representations — such as concreteness and lack of ambiguity — directly influences

comprehension success (Barry & Gerhand, 2003; Perfetti et al., 2008; Van der Schoot et al., 2009; Carpenter & Just, 2017; Perfetti & Helder, 2021; Norberg et al., 2022). Thus, the mental lexicon serves as the critical link between word recognition and comprehension, supporting the smooth integration of word meanings into coherent textual understanding.

Words as the basic unit of understanding

Within the RSF, word knowledge is central to comprehension, making it essential to understand how words are represented and accessed in the mental lexicon. This emphasis on lexical representation provides a useful link to psycholinguistic models that describe how word is defined and how lexical information is organized and retrieved. One such framework is Levelt's (1993) model of spoken language production. In Levelt's model, the concept, lemma, and lexeme represent three levels of lexical representation that together define a word. Although originally developed for speech production, these levels can also be used to examine how word-defining properties influence the processes a reader engages in while decoding, integrating, interpreting, and ultimately comprehending printed words.

In speech production, Levelt's model proceeds concept → lemma → lexeme, reflecting meaning-driven word retrieval. In reading, however, processing is typically form-driven, meaning that readers initially access orthographic and phonological information before fully activating higher-level semantic and conceptual representations. From this perspective, it is theoretically more accurate to describe the flow as: lexeme → lemma → concept.

Importantly, Levelt's levels should not be viewed as *strictly sequential* in reading.

Contemporary reading research supports interactive and cascaded processing, in which partial semantic activation can occur rapidly, but is nevertheless constrained by the quality of form-level representations (Perfetti & Roth, 1981; Perfetti, 2007; Wen et al., 2019). Therefore, processing typically begins at the level of the *lexeme*, which encodes the orthographic, phonological, and morphological information necessary for decoding and word-form recognition. Successful access to the lexeme enables the reader to retrieve the *lemma*, where syntactic and semantic properties are specified and contextually appropriate meanings are selected. Conceptual representations (*concept*), encompassing perceptual, semantic, and experiential knowledge, are subsequently engaged to support deeper meaning integration and the construction of a coherent mental representation of the text (Jescheniak & Levelt, 1994; Just & Carpenter, 2018).

Efficient comprehension relies on a well-organized mental lexicon that enables rapid access to these representations. Skilled readers readily select context-appropriate meanings, resolve ambiguities such as homographs, and integrate words into coherent sentences. However, the quality of lexical representations varies among individuals (Perfetti & Hart, 2002). High-quality representations, those that are frequent, familiar, and easily visualized, facilitate faster word identification and smoother comprehension (Perfetti, 2007). Conversely, weak or poorly specified representations slow retrieval, hindering decoding and overall text understanding.

This is the essence of the hypothesis proposed by Perfetti & Hart (2002), known as the Lexical Quality Hypothesis (LQH), whose basic postulates are described throughout this chapter. Specifically, the quality of mental representations refers to the level of specification, and clear conceptual, semantic, and syntactic distinction in each word component, with mental representations depicted as connection links (Perfetti & Hart, 2002; Perfetti, 2007). According to the LQH, both word identification and word-to-text integration processes are affected by lexical quality. The quality of the reader's mental representations determines how easily and successfully the reader identifies the word and creates a mental representation of the text.

Recent literature identified several possible reasons why this quality may be low (Perfetti & Hart, 2002; Murphy, 2016; Hsiao & Nation, 2018). Low-quality lexical representations may result from limited orthographic or phonological knowledge, making it difficult for readers to recognize and decode words accurately (Perfetti & Hart, 2002; Hsiao & Nation, 2018). Weak or superficial semantic knowledge can further reduce the stability of word meanings, impairing integration into broader contexts. Slow or inefficient access to these representations delays comprehension and disrupts text processing. Factors such as limited spoken and written exposure, developmental differences, or disorders like dyslexia and developmental language disorder also contribute to individual variability (Perfetti, 2007; Hersch & Andrews, 2012; Richter et al., 2013; Dyson et al., 2017). Overall, weaknesses in language processing, knowledge quality, and retrieval efficiency collectively undermine reading comprehension.

Perfetti (2017) defines lexical quality in terms of form, unification, and context-sensitive meaning, operationalized as the richness of word representations containing detailed and overlapping information about form, meaning, and exposure. Form includes phonological and orthographic features, meaning refers to the semantic representation of a word, while

exposure reflects the frequency and experience-based strengthening of lexical representations through repeated encounters. Together, these components support efficient word recognition and comprehension.

In line with this framework, Reichle and Perfetti (2003) emphasize these dimensions as core components of high-quality lexical representations that underlie fluent reading comprehension (Figure 2).

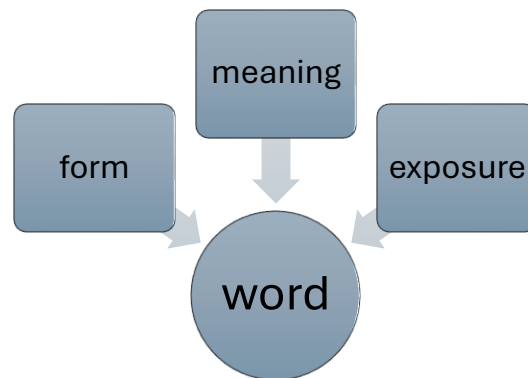


Figure 2. Word properties as three main components of the LQH theory (Reichle & Perfetti, 2003; Perfetti, 2017)

Word-form related measures

Word form is a core component of lexical quality, encompassing the phonological and orthographic structure of a word and shaping how efficiently it is identified during reading. Form-related properties such as word length, orthographic neighbourhood size, and neighbour frequency (Andrews, 1997; New et al., 2006; Brysbaert et al., 2016) influence both decoding and comprehension by determining the ease of accessing a word's representation in the mental lexicon (Lind et al., 2015; Soares et al., 2017). Word length is a well-documented form-related feature that directly affects recognition speed. Research shows a non-linear or quadratic effect: short words (3-5 letters) are recognized quickly, mid-length words (5-8 letters) show little variation, while long words (8-13 letters) significantly slow recognition (New et al., 2006; Yap & Balota, 2009), reflecting developmental shifts described in the Dual-Route Model (Coltheart, 2006). Young or less-skilled readers depend heavily on phonological decoding, while skilled readers retrieve familiar words directly from memory, bypassing phonological recoding (Martens & de Jong, 2006; Monster et al., 2022).

The orthographic neighbourhood — the number of words similar in spelling — also matters: words with many high-frequency neighbours are recognized faster because overlapping letter patterns facilitate lexical activation, and similar processes apply to phonological structure (Castles & Nation, 2006a, 2017b). Overall, orthographic and phonological form support

efficient word recognition, reducing cognitive demands on decoding and enabling resources to shift towards higher-level comprehension (Perfetti, 2017; Gerth & Festman, 2021).

Word-meaning related measures

Word-meaning measures capture how semantic properties shape lexical access and comprehension. Substantial evidence now shows that words with richer semantic networks — emotionally charged, highly imageable, or used in varied contexts — are recognised more efficiently (Buchanan et al., 2001; Pexman et al., 2008; Yap et al., 2012).

Word valence refers to a word's emotional tone, with both positive and negative words processed more quickly than neutral ones due to stronger semantic feedback (Balota & Chumbley, 1984; Yap & Seow, 2014). Semantic ambiguity, where words have multiple meanings, also facilitates faster recognition because several meanings activate broader semantic networks (Pexman & Lupker, 1999a, 2006b; Hino et al., 2002). Finally, words encountered in diverse contexts develop richer and more flexible representations, supporting more efficient comprehension (Hoffman et al., 2013).

Another central feature of word meaning is semantic richness, imageability, the ease with which a word evokes a mental image (Paivio et al., 1968). Imageability primarily affects the conceptual level of representation, as it reflects the richness of perceptual, semantic, and experiential information associated with a word's meaning (Levelt, 1992). Therefore, it is associated with semantic representation of the word (Balota et al., 2004). Highly imageable words activate their conceptual features more readily, leading to faster and more robust semantic and referential integration. Imageability has minimal influence on lexeme-level decoding and only indirectly influences lemma-level activation through the semantic richness it provides. In contemporary models of visual word recognition, lexical processing is understood as a cascaded and interactive rather than a strictly sequential system. Within the Dual-Route Cascaded framework (Coltheart et al., 2001), orthographic, phonological, and semantic information are activated in parallel and linked through bidirectional connections, allowing higher-level semantic representations to begin activating before the orthographic form reaches its recognition threshold. This architecture allows semantic properties of a word such as imageability, to exert an influence very early in lexical access. Highly imageable words, by virtue of their richer conceptual content, can accelerate recognition because semantic activation provides additional support to the emerging orthographic representation (Coltheart, 2004). Importantly, while semantic information is not primarily required for word

identification, it can facilitate processing during stages that rely on the initial establishment of meaning. Thus, imageability is expected to contribute most strongly during early conceptual activation, particularly in contexts where semantic support is beneficial for establishing a stable interpretation of the word.

High-imageability words are processed more efficiently, as they engage both verbal and visual systems. This dual representation is explained by Paivio's Dual Coding Theory (1971, 1990), which posits two processing routes: a linguistic-verbal one and a nonverbal-imagery one. Highly imageable words activate both, while low imageability words rely solely on linguistic coding, making them slower to process. Research shows that imageable words create stronger, more durable semantic networks, enabling faster retrieval and richer mental representations (Sadoski & Paivio, 2004; Peti-Stantić et al., 2018).

Studies across languages and developmental stages confirm that imageability supports reading and early word learning. Children recognize and recall highly imageable words more easily in context-free single word reading tasks, aiding vocabulary growth and reading acquisition (Masterson et al., 2008; Duff & Hulme, 2012). In adults, imageability predicts reading accuracy and lexical decision speed, particularly for low-frequency or irregular words (Pexman et al., 2012; Cortese & Schock, 2013; González-Nosti et al., 2014). These findings suggest that rich, concrete semantics enhance word recognition and comprehension efficiency (Brysbaert et al., 2018).

However, the evidence is not uniformly consistent, partly due to methodological limitations, small stimulus sets, and the well-known issue that psycholinguistic properties often correlate with one another (Juhasz & Rayner, 2003; Yates, 2005; Yap et al., 2015). This complexity presents challenges for disentangling the specific contribution of imageability from related factors such as familiarity, age of acquisition (AoA), or concreteness.

In fact, conceptual-semantic properties, such as imageability, have received comparatively little attention in research on at-risk or struggling readers (Keenan & Betjemann, 2007), despite their potential importance for children with language-based disorders. Words low in imageability or with weak semantic grounding are particularly difficult for such readers, who may struggle to generate mental images or activate stable semantic representations (Nation & Snowling, 1998; Erdeljac et al., 2014).

Concreteness, closely related to imageability, has traditionally been examined as aligned construct. Schwanenflugel and Stowe (1989) showed that abstract words were named more

slowly than concrete words in neutral contexts, but this concreteness effect disappeared in meaningful contexts. This pattern suggests that concreteness — and potentially by extension imageability — exerts its strongest influence early in word recognition, particularly when contextual support is minimal and semantic activation must rely on the intrinsic richness of the word's concept. Juhasz and Rayner (2003) provided a clearer identification of these early semantic effects. They systematically manipulated several intercorrelated lexical variables: frequency (adult and cumulative), familiarity, AoA, and concreteness — using carefully selected stimuli with minimized intercorrelations, each embedded in controlled sentence contexts. By applying participant-wise regressions and appropriate crossed-effects statistical procedures, they demonstrated that all variables influenced eye-movement behaviour during reading, but at different stages of processing. Notably, concreteness exerted measurable effects on early and late processing depending on the frequency metric used. Thus, they provided the first evidence that concreteness impacts fixation durations in sentence reading, highlighting the role of semantic richness during real-time lexical processing.

Surprisingly, few studies have manipulated imageability directly to test its causal influence on word learning or reading. The only known attempt, by Duff and Hulme (2012), used semantic cues to make nonwords more imageable, but found no additional benefit beyond phonological information. This makes it difficult to draw firm conclusions about the conditions under which imageability enhances reading processes, particularly in meaningful text contexts and in populations with weaker lexical-semantic systems. Furthermore, more recent studies indicate that conceptual-semantic features of words (e.g. imageability, concreteness) influence reading processes in more nuanced and word-specific ways than previously assumed. Research using more advanced statistical approaches, such as linear mixed-effects models and item-based IRT analyses, shows that semantic richness can facilitate the reading of both regular and irregular words, across languages of varying orthographic depth (e.g. English, Ricketts et al., 2016; Spanish, Davies et al., 2013). These models allow word-specific semantic properties, such as imageability, to predict word-specific reading performance, revealing fine-grained semantic effects that traditional methods often overlook. Such findings suggest that the influence of imageability emerges selectively during stages of processing that rely on conceptual activation, while its visibility depends on methodological sensitivity and the quality of a reader's semantic system. Consequently, semantic variables like imageability may exert measurable effects on early referential

encoding during text reading — particularly in populations whose lexical — semantic representations are less robust, such as children with language-based disorders.

Word-exposure related measures

Word-exposure related measures refer to how often a word is encountered, processed, and used, shaping its accessibility in the mental lexicon. Words that appear frequently in speech or print become more robustly represented, enabling faster and more accurate recognition, retrieval, and comprehension (Jurafsky, 2003; Perfetti, 2007). This relationship — known as the *word frequency effect* — is one of the most consistent findings in psycholinguistics, showing that frequent words have lower activation thresholds and require less cognitive effort to process, leading to faster and more accurate recognition, processing, storage, retrieval, naming, reading, production, and comprehension (Hennessey & Kirsner, 1999; Bley-Vroman, 2002; Jurafsky, 2003; Benki, 2003; Balota et al., 2004; Desrochers & Thompson, 2009; Ferrand et al., 2011; Joseph et al., 2013; Joseph & Liversedge, 2013; Kuperman & van Dyke, 2013; Brysbaert et al., 2018; Tiffin-Richards & Schroeder, 2015; Whitford & Joanisse, 2018; Gerth & Festman, 2021).

While imageability is conceptually and semantically grounded, the classification of word frequency is less straightforward. Some scholars, such as Kittredge et al. (2008), argue that frequency simultaneously pertains to concepts, lemmas, and lexemes, whereas Jescheniak & Levelt (1994) emphasize that frequency is primarily a feature of the lexeme level. Thus, frequency likely facilitates processing throughout the entire lexical pathway — from decoding to lexical access, to semantic integration, freeing cognitive resources for higher-level comprehension processes (Perfetti & Hart, 2002). Within models of visual word recognition, word frequency is understood as a fundamental determinant of processing efficiency because it influences activation across all levels of lexical representation. In the Dual-Route Cascaded Model (Coltheart et al., 2001), orthographic, phonological, and semantic codes are activated in parallel, and frequently encountered words reach recognition thresholds more rapidly due to their stronger and more stable orthographic and phonological representations. High-frequency words possess more stable and easily accessible orthographic and phonological forms (lexeme), faster access to lemma-level information (syntactic and semantic features), and maintain more strongly connected conceptual networks (Levelt, 1992). As activation flows bidirectionally across the system, frequency exerts a broad facilitative effect: it accelerates orthographic decoding, speeds lexical access, and supports efficient semantic integration (Coltheart, 2004). This system-wide influence reflects

the fact that repeated exposure strengthens mappings at each component of the lexical pathway, ultimately reducing cognitive load and freeing resources for higher-level comprehension (Perfetti & Hart, 2002). Consequently, frequency effects are expected to appear early and consistently across measures of word recognition, and to exert particularly strong effects for readers who rely more heavily on form-based decoding. Empirical evidence confirms that high-frequency words are recognized faster and with fewer errors than low-frequency ones (Inhoff & Rayner, 1986; Rayner & Duffy, 1986). Word frequency accelerates language processing and enhances its accuracy (Falkauskas & Kuperman, 2015), with high-frequency words being recognized more quickly than low-frequency ones (Brysbaert et al., 2018), reflecting more efficient lexical access. In naming and reading tasks, frequency also predicts fluency across age groups, reducing both phonological and semantic errors (Masterson et al., 2008; Balota et al., 2012). In contrast, low-frequency words require more decoding effort, slowing comprehension and increasing cognitive load. In eye-tracking studies using reading paradigm, longer first fixations and gaze durations were observed in conditions with low-frequency words, compared to high-frequency ones (Hyönä and Olson, 1995).

Although both children and adults exhibit frequency effects in language processing, the literature suggests that these effects manifest differently across age groups. Several studies (including those using eye-tracking methods) have demonstrated that frequency effects are more pronounced in children, who process low-frequency words more slowly than adults (Blythe et al., 2009; Huestegge et al., 2009; Hyönä & Olson, 1995; Joseph et al., 2013; Luke et al., 2015; Tiffin-Richards & Schroeder, 2015; Vorstius et al., 2014). One possible explanation, as proposed by the authors, is that children's lexical structures and processing mechanisms are not yet as developed as those of adults. Tiffin-Richards and Schroeder (2015) conducted an eye-tracking study in which they observed a frequency effect during silent reading. They found that both children and adults spent more time and made more fixations on less frequent words, however this frequency effect was generally larger for children than for adults. Young readers rely more on phonological decoding, and because their lexical networks are still developing, they process low-frequency words more slowly than adults (Blythe et al., 2009; Davies et al., 2017). As reading experience increases, word recognition becomes more automatic, and frequency effects diminish — high-frequency words are recognized equally well across proficiency levels, while low-frequency words remain challenging (Kuperman & van Dyke, 2013). For children with developmental

language disorder (DLD) or dyslexia, limited exposure reduces the quality of lexical representations, leading to slower retrieval and weaker comprehension (Nation, 2013, 2017). However, a strong frequency effect was not consistent in all studies, or across all the observed measures. For example, results from Gerth and Festman's study (2021) revealed no significant frequency effect in first-fixation duration or gaze duration, but only in total reading time.

From a cognitive perspective, repeated encounters strengthen a word's representation through statistical learning, reinforcing connections between form and meaning (Frishkoff et al. 2010; Baayen et al., 2011; Brysbaert & Cortese, 2011; Herdağdelen & Marelli, 2017). Two perspectives explain how frequency enhances lexical quality: the repetition-based view, which emphasizes the sheer number of exposures, and the contextual-quality view, which highlights the richness and variability of encounters (Hsiao & Nation, 2018). Words learned across diverse contexts develop stronger, more flexible representations, facilitating efficient integration into discourse. Perfetti and Hart (2002) and Perfetti (2007), through their Lexical Quality Hypothesis, provide further insight into why frequency affects lexical processing. They argue that repeated exposure to a word strengthens its mental representation by reinforcing and stabilizing the connections between all relevant components — phonological, orthographic, and semantic. The stronger and more stable these connections become due to frequent word exposure, the more accessible and easier the target word is to recognize and process.

Frequency also interacts with other psycholinguistic properties, such as word length, imageability and age of acquisition, influencing how words are stored and retrieved. Frequent, highly imageable words — those that evoke clear mental imagery — are processed most efficiently (Desrochers & Thompson, 2009; Duff & Hulme, 2012). Conversely, rare and abstract words, which are less imageable and encountered infrequently, impose greater processing demands, particularly for struggling readers. Studies in aphasia and other clinical populations show that word frequency reduces semantic and phonological errors, indicating its pervasive role in lexical access at all processing stages (Kittredge et al., 2008). The relationship between word frequency and imageability remains an active area of research, particularly in psycholinguistics. Current findings suggest two opposing trends: a proportional and an inversely proportional relationship between these two lexical attributes. In their studies, Desrochers & Thompson (2009), along with Gonthier et al. (2009) for French and Kuvač Kraljević & Olujić (2018) for Croatian, found evidence of a link between

frequency and imageability. They highlighted that, for less frequent words, when frequency is subjectively measured, it serves as a predictor of imageability. These findings support a proportional relationship between the two features, whereas research by Ferrand et al. (2008) suggests a decline in imageability as subjective frequency increases, supporting an inverse relationship.

How do psycholinguistic word properties (features of lexical complexity) such as imageability or frequency affect a reader's ability to draw inferences and maintain discourse cohesion? To what extent do these properties influence the construction of coherent meaning during reading comprehension? These questions are particularly relevant when considering how discourse, whether written or spoken, achieves cohesion and coherence through devices such as anaphora.

1.3 Anaphora processing and reading comprehension

Discourse, whether spoken or written, consists of interconnected units of meaning — propositions — that must be linked and interpreted to form a coherent understanding of the text (Sukthanker et al., 2020). Comprehension depends on the reader's ability to identify and resolve references, integrate information across sentences, and establish relationships between propositions to build a coherent mental model (Bowyer-Crane & Snowling, 2005; Perfetti & Stafura, 2014). Cohesion in discourse is maintained through linguistic devices such as anaphora, which connect ideas, track referents, and prevent redundancy, particularly in narratives (Halliday & Hasan, 1976).

Anaphora has been defined in various ways, from grammatical perspectives that describe it as a structural reference to a previously mentioned element (Crystal, 2008) to pragmatic views that emphasize its contextual and communicative function (Levinson, 1987). The most influential definition by Halliday and Hasan (1976) presents anaphora as a key mechanism of cohesion, linking two discourse elements whose interpretation depends on each other. In this relation, the anaphora refers to its antecedent, as in “*The mother bought Ana a hat. She put it on her head,*” where *she* refers to the *mother*.

Through such referential links, anaphora ensures semantic continuity and coherence, enabling readers to integrate information across sentences and maintain a unified understanding of discourse (Safir, 2004; Buring, 2005).

1.3.1 Comparison of anaphoric structure between Croatian and English

Croatian, a morphologically rich language with transparent orthography, is classified as a consistent null subject language (Huang, 2000; Barbosa, 2011), meaning the subject is often omitted when contextually inferable (*pro-drop* language). This results to frequent use of null anaphora, where pronouns are omitted — unlike in English, which requires explicit subjects. Croatian employs various anaphoric devices, including nouns, pronouns, adverbs, ellipses, repetition, and synonym substitution (Silić, 1984; Kordić, 1996; Katičić, 2002; Batinić Angster, 2020), though systematic classifications are less developed than in English (Košutar, 2020, 2022).

Due to its morphological flexibility, Croatian allows greater syntactic freedom, while English relies on explicit pronoun use to maintain coherence. In both languages, anaphora resolution — the process of linking an anaphor to its antecedent — is key to understanding how readers integrate information across discourse (Mitkov, 2002; Crystal, 2008). Successful resolution enables readers to maintain referential continuity, draw inferences, and construct coherent mental models of the text (Kintsch, 1998; Stanovich, 2000).

However, the mechanisms underlying anaphora resolution remain complex and underexplored, particularly in languages with transparent orthography such as Croatian. Differences in linguistic structure, morphology, and orthographic depth pose methodological challenges for cross-linguistic comparisons, highlighting the need for further research into how readers process and resolve anaphoric references for successful comprehension.

1.3.2 The process of anaphora resolution

Anaphora resolution is a multi-step process involving identification of the anaphor, the location of potential antecedents, and the selection of the correct one (Pedruco Nobre, 2011). Its efficiency depends on various linguistic and cognitive factors, among which psycholinguistic properties of the antecedent (such as word imageability or frequency) may play a key role. These properties may influence how easily readers retrieve and link antecedents with their anaphoric references during text comprehension.

Studies show that high-frequency nouns are resolved faster than low-frequency ones, indicating that frequent words are more readily accessible in the mental lexicon (Egusquiza et al., 2016). However, findings for pronouns are less consistent, with some studies showing no clear frequency effect, likely due to methodological differences (Van Gompel & Majid, 2004; Lago et al., 2014). In contrast, highly imageable words, which evoke clearer mental

representations, appear to facilitate anaphora resolution by making antecedents more salient and easier to retrieve (Gerwien et al., 2024).

Despite strong evidence from adult populations, few studies have examined these effects in children, providing limited data especially for pro-drop languages such as Croatian, where null anaphora require greater reliance on discourse and lexical cues. Existing research shows that children's ability to resolve anaphoric references develops gradually and is shaped by both linguistic and cognitive factors (Palmović et al., 2018). Less-skilled comprehenders have difficulty establishing referential links — especially when antecedents are distant or less salient — indicating challenges in maintaining coherence and constructing mental representations of text (Ehrlich & Remond, 1997).

Cross-linguistic findings suggest that children rely more on contextual and visual information than adults when interpreting pronouns (Hickmann, 2002; Palmović et al., 2018), while adults use more consistent, structure-based strategies (Papadopoulou et al., 2015). Eye-tracking studies confirm that children process pronouns more slowly and show longer gaze durations when resolving ambiguous references, indicating increased processing effort and incomplete integration of referential information (Megherbi et al., 2019; Eilers, 2021). However, children with stronger comprehension skills demonstrate greater anticipation of referents and more efficient use of discourse cues during reading (Engelen et al., 2014).

Research in Croatian indicates that adult-like preferences for null subject pronouns emerge after the age of six, while full referential cohesion in narratives develops later (Kraš, 2008; Palmović, 2014; Koren, 2019; Trtanj, 2015; Gabaj & Kuvač Kraljević, 2019). Children with poor reading comprehension tend to process anaphoric sentences more slowly and have difficulty with linguistic cues that require higher-level lexical access (Kim et al., 2017).

Overall, the findings suggest that children's success in anaphora resolution depends on the interaction between lexical quality, working memory, and inferencing ability, which together form the foundation for effective reading comprehension (Kintsch, 1998; Perfetti & Stafura, 2014; Joseph et al., 2015; Elbro et al., 2017). Understanding how psycholinguistic properties — such as word frequency and imageability — facilitate or hinder anaphora resolution could therefore provide crucial insights into the mechanisms that support coherent text understanding in developing readers.

1.4 Reading comprehension and reader's individual differences

Given the complexity and multidimensionality of reading comprehension processes, difficulties in understanding text are unsurprising, as readers vary in the efficiency of the underlying processes that support comprehension. Recent research has increasingly emphasized individual differences in these component skills to better explain variability in language and reading performance (Spencer & Wagner, 2018).

This study explores individual differences in reading comprehension and its component processes through the perspective of language-based disorders, primarily developmental language disorder (DLD) and dyslexia, as well as their overlap. Scholars remain divided on their classification, trying to explain why these disorders exhibit both behavioural overlap and frequent comorbidity, while still appearing to involve partially distinct weaknesses. Some argue that when decoding difficulties arise from broader language deficits, DLD and dyslexia should not be considered separate diagnostic categories (Tunmer & Greaney, 2010; Silliman & Berninger, 2011). Others regard them as overlapping but distinct conditions, each with unique cognitive and language profiles (Catts et al., 2005; Bishop et al., 2017). This ongoing debate emphasizes the need for models that integrate both language and reading-related perspectives to better explain the shared and distinct mechanisms underlying language-based disorders and their impact on reading comprehension.

1.4.1 Models of language-based disorders from the perspective of general phonological and non-phonological language skills

Developmental language disorder (DLD; Bishop et al., 2016, 2017), formerly known as specific language impairment (SLI), is characterised by persistent difficulties in acquiring, understanding, and producing language, affecting vocabulary, grammar, and overall communication (McGregor, 2020). These difficulties often hinder literacy development and academic success (Snowling et al., 2016). DLD affects approximately 7-8% of children (Norbury et al., 2016) and is assumed to have a polygenic aetiology rather than a single identifiable biological cause (Bishop, 2006).

Dyslexia, by contrast, is defined as a specific deficit in word reading that cannot be explained by low intelligence or sensory impairment (Lyon et al., 2003; APA, 2013). Its core features are phonological and orthographic processing difficulties, particularly affecting decoding and word recognition (Snowling, 2019). Nevertheless, a substantial proportion of individuals with dyslexia present with early oral language weaknesses, suggesting that dyslexia may form part of a broader continuum of language-based disorder rather than a purely isolated reading

disorder (Catts & Kamhi, 1999; Moats, 2008). Consistent with this view, high rates of comorbidity have been reported, with more than half of children with dyslexia meeting criteria for DLD, and vice versa (McArthur et al., 2000).

To clarify the relationship between DLD and dyslexia, Bishop and Snowling (2004) proposed a two-dimensional model of language disorder, organised along two orthogonal axes: phonological and non-phonological language skills (Figure 3). This framework was designed to account for both shared and distinct phenotypic profiles, allowing the identification of theoretically and neurobiologically coherent subgroups. Crucially, the model challenges the assumption that behavioural overlap necessarily reflects a common underlying cause, arguing instead that superficially similar profiles may arise from different causal mechanisms.

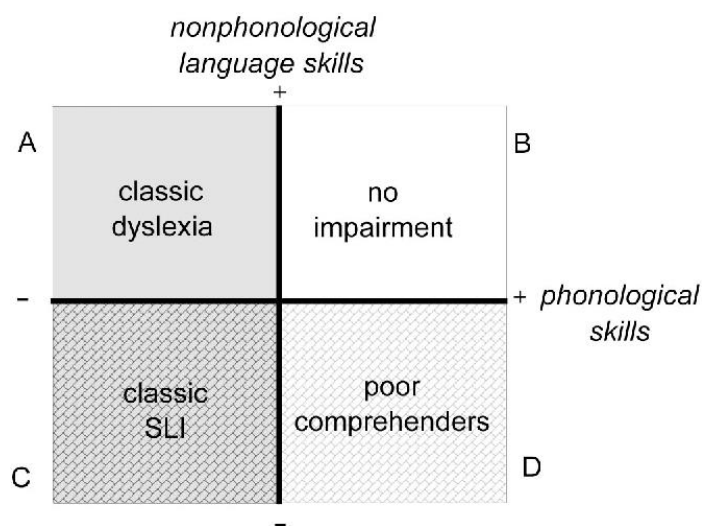


Figure 3. Two-dimensional model of language disorder proposed by Bishop & Snowling (2004)

Empirical evidence supports this distinction (Nation & Snowling, 1998; Ramus, 2001; Catts et al., 2005; Ramus et al., 2013). Comparative studies indicate that dyslexia is primarily associated with phonological deficits, whereas DLD involves broader language impairments encompassing semantics, syntax, and morphology, often alongside phonological weaknesses of varying severity (Gathercole & Baddeley, 1990; Snowling, 2000; Tallal, 2003; Catts et al., 2005; Bishop et al., 2009). Norbury et al. (2016) report that approximately 7.6% of children in Grade 1 meet criteria for developmental language disorder (DLD), whereas dyslexia is estimated to affect around 10% of school-aged children (Moats & Dakin, 2008; British Dyslexia Association [BDA]; International Dyslexia Association [IDA]). Within this framework, Bishop and Snowling also identified a subgroup of poor comprehenders — children who show intact decoding but impaired non-phonological language processing

(Nation, 2005; Cain & Oakhill, 2006; Clarke et al., 2010). This group, representing approximately 5-10% of school-aged children, exhibits difficulties in lexical-semantic access, morphosyntactic processing, and discourse-level integration, despite adequate word recognition skills (Catts et al., 2006; Henderson et al., 2013; Keenan et al., 2014). Studies suggest weaknesses in semantic access (Nation et al., 2001, 2010; Perfetti & Hart, 2001) and morphosyntactic processing (Nation et al., 2004; Adlof & Catts, 2015), though findings remain inconsistent (Oakhill et al., 2003). These children also struggle with higher-level metalinguistic processes, such as inference-making, narrative organization, and comprehension monitoring (Cain et al., 2001; Oakhill & Cain, 2012). Their profile raises important questions regarding the boundaries between reading-specific and language-based disorders.

Catts and colleagues (2005) further addressed why DLD and dyslexia frequently co-occur and share behavioural features yet still appear to involve partially dissociable weaknesses (Figure 4). They contrasted three competing hypotheses that differ in how they conceptualise the role of phonological versus broader oral language skills in the two disorders. Those hypotheses are the Phonological Deficit hypothesis, the Partial Distinction hypothesis, and the Fully Distinct hypothesis.

A central assumption of the Phonological Deficit Severity hypothesis (Kamhi & Catts, 1986) is that most children with DLD should also present with dyslexia, and phonological deficits should be at least as pronounced in DLD as in dyslexia. On this view, dyslexia reflects a milder phonological deficit primarily affecting word-level reading, whereas DLD reflects a more severe disorder extending to oral language.

The Partial Distinction hypothesis aligns with a two-dimensional framework by Bishop & Snowling (2004) also assuming a shared phonological weakness across both disorders. However, it posits that DLD involves additional, independent deficits in broader language domains such as semantics, syntax, and discourse. While overlap is still expected, particularly in word reading, this hypothesis allows DLD to be differentiated by the presence of more pervasive oral language difficulties beyond phonology.

The Fully Distinct hypothesis (Catts et al., 2005) proposes the strongest dissociation, viewing dyslexia and DLD as separate developmental disorders with different core deficits. Dyslexia is attributed to phonological processing disorder underlying word-reading difficulties, whereas DLD is rooted in broader oral language deficits. Any co-occurrence is interpreted as

comorbidity rather than a shared aetiology, leading to the prediction that many children with DLD will not have dyslexia, and many children with dyslexia will not show language problems in early years. Catts et al. (2005) offer particularly compelling evidence due to their population-based longitudinal design ($N > 500$), which separated the identification of DLD (in kindergarten, prior to reading instruction) from the later identification of dyslexia (Grades 2, 4, and 8), reducing the risk that oral-language weaknesses reflect secondary effects of poor reading experience. The results indicate clear but limited overlap between DLD and dyslexia, weaker than predicted by severity-based or strong shared-phonology accounts. When dyslexia was defined using IQ-achievement discrepancy criteria, only about 15-20% of children with dyslexia had met criteria for DLD in kindergarten, while a substantial proportion of children with early DLD did not go on to develop dyslexia, regardless of diagnostic operationalisation. This pattern reflects selective overlap rather than rare comorbidity and is more consistent with a comorbidity model than with a single-deficit continuum. Authors also showed that persistent deficits in phonological awareness and nonword repetition characterised children with dyslexia-only and comorbid DLD + dyslexia, whereas children with DLD-only showed mild and inconsistent phonological weaknesses. Factor-analytic results from the same cohort further demonstrated a separation between phonological and semantic-syntactic skills, supporting partial independence between phonological and non-phonological language deficits and allowing for both dissociation and co-occurrence.

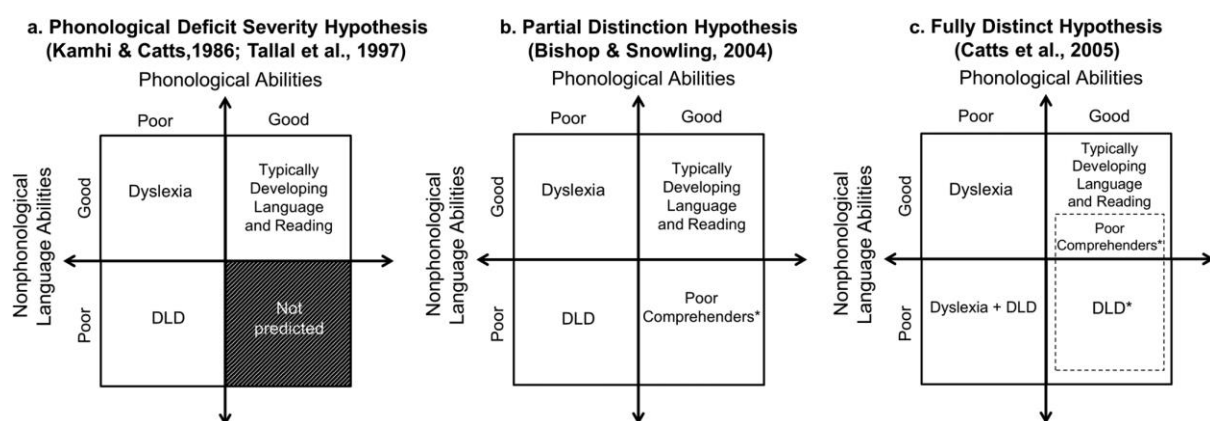


Figure 4. An Overview of the Three Hypothesis of the relationship between DLD (SLI) and dyslexia, according to performance along non-phonological and phonological language skills (Catts et al., 2005; Adlof & Hogan, 2018)

Ramus et al. (2013; Figure 5) and Adlof & Hogan (2018) further reviewed and analysed three major theoretical models based on the previously explained hypotheses (Figure 4), aiming to explain the relationship between dyslexia and DLD (Figure 5).

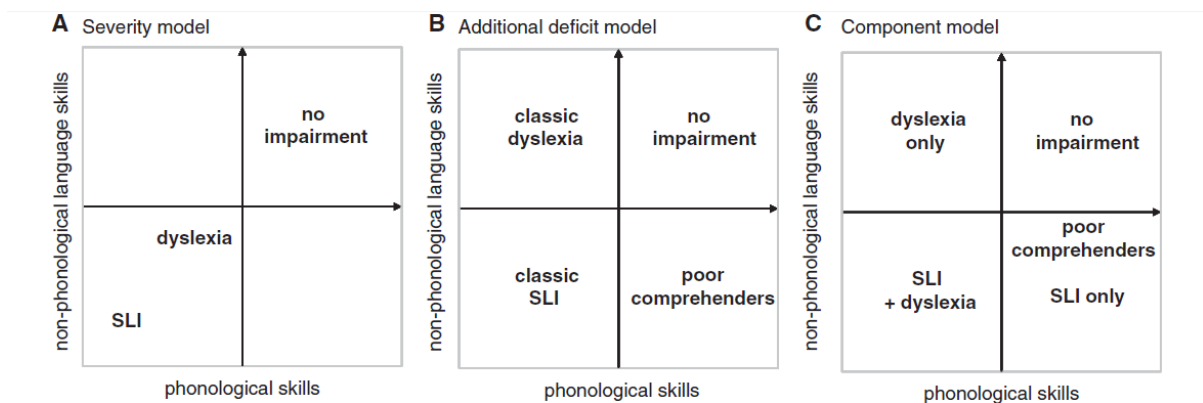


Figure 5. An Overview of the Three Models based on the three competing hypotheses of the relationship between DLD and dyslexia, according to performance along non-phonological and phonological language skills (Ramus et al., 2013). (A) Severity model, (B) additional deficit model and (C) component model.

The *Severity Model*, based on the Phonological Deficit Severity Hypothesis (Kamhi & Catts, 1986) conceptualises DLD and dyslexia as manifestations of the same underlying disorder that differ primarily in severity. According to this view, both disorders arise from the same underlying phonological deficit, with DLD representing a more severe manifestation that additionally affects oral language development (Kamhi & Catts, 1986; Fitch & Tallal, 1997; Tallal et al., 2003; Adlof & Hogan, 2018). However, this model has been challenged by findings demonstrating that non-phonological language impairments, such as deficits in grammar and vocabulary, can occur independently of phonological weaknesses (Adlof & Hogan, 2018).

The *Additional Deficit Model*, derived from Bishop and Snowling's (2004) Partial Distinction Hypothesis, proposes that dyslexia and DLD share phonological difficulties, but that DLD is characterised by additional non-phonological language deficits (Adlof & Hogan, 2018, Figure 5.). This model therefore treats DLD as a qualitatively distinct condition, defined by broader language disorder beyond phonology. It highlights the differential causal pathways underlying these disorders, and the complexity and heterogeneity of DLD, distinguishing it from dyslexia. Accordingly, it suggests that the diagnostic classification should be based on the primary, dominant domain of language difficulties. Studies confirm this distinction: individuals with dyslexia primarily exhibit phonological deficits (Vellutino et al., 2004), whereas those with DLD also struggle with syntax, semantics, and morphology (Nation & Snowling, 1998; Catts et al., 2005; Adlof & Hogan, 2018; Alonzo et al., 2020).

Finally, the *Component Model* (Ramus, 2003a, 2004b; Catts et al., 2005; van der Lely, 2005, Ramus et al., 2013; Adlof & Hogan, 2019), grounded in the Fully Distinct Hypothesis (Catts

et al., 2005; Figure 5.) represents the strongest theoretical separation between the two disorders. This model posits that dyslexia and DLD are independent, modular conditions with distinct aetiological bases. Their frequent co-occurrence is attributed not to a shared core deficit (e.g. phonological), but to overlapping genetic and environmental risk factors. Importantly, this model allows phonological and non-phonological language skills to vary independently, offering a principled account of the wide range of observed language and reading profiles.

Ramus et al. (2013), comparing children with DLD, dyslexia, comorbid DLD and dyslexia, and typical development, found results most consistent with the Fully Distinct Hypothesis (Catts et al., 2005) and the Component Model, which regards DLD and dyslexia as distinct but potentially overlapping disorders. Their factor analysis identified three independent sources of variance: non-phonological language skills, phonological skills, and phonological representations. This contradicted the Severity Model and refined Bishop & Snowling's (2004) two-dimensional framework by emphasizing phonological components. However, non-phonological skills such as grammar and lexicon were only mildly affected, and poor comprehenders were not considered.

Adlof & Hogan (2018) expanded this work within the Fully Distinct Hypothesis. They highlight that estimates of overlap between dyslexia and DLD vary widely across studies partly because diagnostic criteria for dyslexia differ (IQ discrepancy vs. IQ cut-off vs. low achievement definitions), and because clinically referred samples tend to over-represent comorbid and more severe cases. Also, their work helped to frame the poor comprehender profile in relation to these models. Therefore, their study yielded several key conclusions: (a) phonological deficits consistently mark dyslexia (with or without DLD), (b) language deficits outside phonology can be present and meaningful in dyslexia, but they do not imply that dyslexia collapses into DLD, and (c) children with adequate decoding but weak comprehension (poor comprehenders) often show language weaknesses in vocabulary/syntax/discourse, and a non-trivial proportion meet diagnostic criteria for DLD, supporting the broader point that comprehension problems can arise from non-phonological language constraints even when word recognition is intact. Similarly, Kelso et al. (2007, 2020) confirmed that a subgroup of children with DLD fits the poor comprehender profile: strong phonological skill but weak oral and text-level comprehension, affecting around 5-10% of school-age children.

In this respect, the Component Model aligns with the Additional Deficit Model but distances itself from the Severity Model, which places DLD and dyslexia on a continuum of the same disorder. However, Lancaster and Camarata (2019) propose viewing DLD as a continuum of language difficulties, emphasizing shared core features and individual variability rather than distinct subtypes. Their work suggests this perspective may explain within-group diversity and raises the question of whether poor comprehenders fall within the broader DLD spectrum (Bishop & Snowling, 2004). Nevertheless, these models highlight the heterogeneity of language-based disorders and the need to examine both shared and distinct language mechanisms underlying reading comprehension difficulties.

Taken together, the available evidence most strongly supports the Fully Distinct hypothesis, according to which dyslexia and DLD represent separable developmental, language-based disorders that may co-occur but do not arise from a single underlying deficit. This framework provides a principled basis for examining how phonological and non-phonological language skills contribute independently and interactively to reading comprehension outcomes and therefore serves as one of the main theoretical perspectives for the present dissertation. For this reason, dyslexia and DLD are jointly referred to as *language-based disorders* in the present dissertation, a neutral umbrella term intended to capture their language-related nature without implying shared etiology or diagnostic equivalence.

1.4.2 Models of language-based disorders from the perspective of reading skills and reading comprehension features

When examining language-based disorders in the context of reading comprehension, the Simple View of Reading (Gough & Tunmer, 1986) provides a key theoretical framework that integrates both word recognition and language comprehension components (Hogan et al., 2014; Castles et al., 2018). This model clarifies how deficits in either or both components contribute to reading difficulties and guides targeted interventions towards specific underlying weaknesses.

As individuals with language-based disorders may show distinct patterns across these components, researchers have proposed the Reading Component Model (Catts & Kamhi, 1999; Joshi & Aaron, 2000; Savage, 2001; Tunmer & Chapman, 2007; Tunmer & Greaney, 2009), which classifies poor readers based on individual differences in decoding and language comprehension rather than IQ-achievement discrepancies (Catts et al., 2003). Grounded in the SVR, this model provides a more precise means of identifying and addressing diverse reading profiles within language-based disorders.

Applying the Reading Component Model (Gough & Tunmer, 1986; Catts & Kamhi, 1999; Joshi & Aaron, 2000; Tunmer & Chapman, 2007; Tunmer & Greaney, 2009) to children with reading disorder typically identifies three subgroups:

1. **Poor Decoders:** These children have average language comprehension skills but significant difficulties decoding in written text. Phonological deficits, such as difficulties with phonological awareness, phonological encoding, and the acquisition of orthographic representations in lexical memory (Lyon et al., 2003; Catts & Kamhi, 2005), underlie their weaker decoding skills. This profile aligns with the phonological deficit hypothesis (Stanovich, 2000) and is often associated with dyslexia.
2. **Poor Comprehenders:** Children in this group struggle primarily with language comprehension, as shown by weaker vocabulary, semantic processing (Nation et al., 2004), morphological and syntactic skills (Nation et al., 2004), and text comprehension strategies, including inferential reasoning and auditory comprehension (Cain et al., 2001; Cain et al., 2003; Oakhill et al., 2003). Despite these challenges, their decoding skills (phonological language skills and phonological processing) are generally within the average range (Cain et al., 2000).
3. **Combined/Mixed Group:** Children in this group have difficulties in both components, with systematic phonological deficits and widespread language difficulties, which consequently affect their understanding of written language (Snowling et al., 2020).

Catts et al. (2003, 2005) analysed reading profiles in children and found substantial heterogeneity among poor readers. About 35% were poor decoders (dyslexia) and another 35% had mixed reading difficulties, with proportions remaining relatively stable from grade 2 to grade 8. Longitudinal data revealed that poor comprehenders initially lagged in phonological awareness but caught up in decoding over time, while their comprehension deficits persisted. By grade 8, mixed difficulties (33%) and poor comprehension (30%) were the most prevalent profiles, underscoring the enduring nature of language-related comprehension problems. Similar developmental trends were confirmed in later longitudinal studies (Nation et al., 2010; Elwér et al., 2013).

As most research is based on English, findings may differ in transparent orthographies, where decoding develops earlier and comprehension-related differences become more prominent. This underscores the need to examine reading profiles across different linguistic contexts.

Research in transparent orthographies also relies on the SVR model (Gough & Tunmer, 1986), which has been validated across alphabetic languages (Foorman & Petscher, 2018; Sánchez-Vincitore et al., 2022). Torppa et al. (2010), studying Finnish, identified five reading profiles, highlighting variability across decoding and comprehension abilities: a) poor readers (struggling with both — word recognition and reading comprehension); b) slow decoders (difficulties with fluent word recognition but gradually achieving average reading comprehension); c) poor comprehenders (average word recognition skills but underperforming in reading comprehension); d) average readers (adequate in both decoding and comprehension); e) good readers (above average in both components). Kuvač Kraljević et al., (2024) examined Croatian-speaking second graders and identified five subtypes of poor readers, with the mixed difficulties group being the most prevalent. Distinct profiles included poor decoders (accuracy deficits), slow readers (reduced speed but accurate decoding), and poor comprehenders, mirroring findings from Torppa et al. (2010). However, the authors stress the need for caution when generalizing results due to methodological variability across studies in sample selection and comprehension measures.

This study followed the two-dimensional framework proposed by Bishop and Snowling (2004) to distinguish typical language development from language-based disorders. Children with DLD and/or dyslexia were grouped according to the Fully Distinct Hypothesis (Catts et al., 2005) and the Simple View of Reading (Reading Component Model, Gough & Tunmer, 1986), ensuring each child was classified by his/her specific language and reading profile. The same theoretical basis was used to identify typically developing peers, allowing for a consistent and comparable grouping across the full sample, as detailed in the Method section. Furthermore, previously described developmental patterns motivate the age range selected for the present study. By Grades 5 and 6 (approximately 10 to 12 years of age), word recognition is expected to have reached the automatised, orthographic stage described earlier (Frith, 1985; see Section 1.3.1), language comprehension is expected to have become the dominant predictor of reading comprehension (Geva & Farnia, 2012; Cain, 2015; Lervåg et al., 2017; see Section 1.3.1), and children's eye-movement patterns during reading are expected to approximate those of adult readers (Rayner, 1986; see Section 1.7.1). Against this backdrop of expected automaticity, any continued reliance on decoding or atypical eye-movement behaviour becomes diagnostically informative, rather than being confounded with the decoding effort that is developmentally appropriate for younger, beginning readers. This is precisely why Grades 5 and 6 were selected for the present study: they offer a developmental

window in which residual difficulties are more likely to reflect a genuine reading or language-based disorder than a normative stage of reading acquisition still in progress.

1.5 Assessment of reading comprehension and its underlying components

Research on language-based disorders and reading comprehension difficulties reveals considerable inconsistency in how children are identified and classified, in both research and clinical practice. Keenan et al. (2014) highlight three main sources of variability: selection criteria for atypical language profiles, definitions of reading comprehension deviation, and the measures used for assessment. Participant sampling approaches typically include: (1) large-scale population studies categorizing children into SVR-based profiles (Foorman et al., 2018), (2) studies with pre-diagnosed groups further classified by language performance (Kelso et al., 2007), and (3) studies focusing specifically on poor comprehenders (Keenan et al., 2014).

The criteria for defining poor readers vary significantly. Some studies use fixed cut-off scores for comprehension tests, while others compare comprehension and decoding results to define subgroups (Keenan et al., 2014). Methods for assessing decoding also vary, with some relying solely on real-word accuracy and others including nonword reading and speed measures. For example, Yuill & Oakhill (1991) classified poor comprehenders as children whose comprehension age was at least six months behind their decoding age, whereas Nation & Snowling (1998) applied a stricter one-year-below criterion. Kelso et al. (2020) demonstrated how prevalence rates fluctuate depending on thresholds — from 5% under the strictest definitions to over 10% when broader percentile criteria are applied (Nation et al., 2010; Elwér et al., 2015). Given such variability, researchers stress the need for greater standardization in participant selection and diagnostic criteria to ensure consistency and comparability across studies. Accurate identification of language-based disorder profiles and reading subgroups should be guided by the SVR (Gough & Tunmer, 1986), emphasizing the distinction and interaction between decoding, language comprehension, and reading comprehension (Kintsch, 2012; Van den Broek et al., 2012).

Several authors have discussed the challenges and methodological issues in assessing reading comprehension (Cain & Oakhill, 2006; Fletcher, 2006; Westerveld, 2009; Keenan et al., 2014). Blaži Ostojić (2023) further elaborated on these challenges, emphasizing the complexity of measurement, theoretical clarity, misleading test labels, undetected comprehension difficulties, developmental and language-specific factors, and the need for

comprehensive, multi-level assessment. As reading comprehension extends beyond word-level decoding, it requires evaluation across multiple language domains, including vocabulary, syntax, memory, and discourse-level integration (Westerveld, 2009). Keenan et al. (2008) highlighted that different text types engage distinct processes — short texts depend more on word recognition, while discourse-level comprehension relies on contextual and inferential language skills. Consequently, assessing reading comprehension requires a careful alignment between the theoretical construct, processing level, and modality (written or oral) being measured (Keenan et al., 2014). For this reason, researchers emphasize the importance of including both language modalities and multiple language domains in assessment frameworks to provide a more accurate representation of children's reading and language profiles (Gillon et al., 2004; Catts et al., 2005).

Another major issue in assessing reading comprehension is construct validity — whether a test genuinely measures the intended construct and accurately captures all relevant profiles of comprehension difficulties (Keenan et al., 2008). Tunmer and Chapman (2012) stress the need to clearly distinguish reading comprehension components, as high correlations between word reading and comprehension ($r = .70$; de Jong & van der Leij, 2002) may obscure their unique contributions. Moreover, developmental factors such as age and education significantly influence the interaction and relative importance of these components, with their roles evolving across different stages of reading development (Tilstra et al., 2009; Cain et al., 2017). Variability in reading comprehension assessment largely results from differences in task format, question type, reading mode, and text characteristics (Kim et al., 2021). Open-ended questions, though cognitively demanding, offer richer insight into inferencing, working memory, and related processes than multiple-choice or cloze tasks, which often overemphasize word recognition (Cain & Oakhill, 2006; Keenan et al., 2008). Reading mode also plays a role, with evidence suggesting that reading aloud may hinder comprehension for some children (Westerveld, 2009). Furthermore, language-specific factors complicate assessment: most tools are developed for English, whose non-transparent orthography differs from transparent languages such as Croatian, where decoding is acquired more rapidly (Seymour et al., 2003; Ziegler et al., 2010). Given the lack of standardized Croatian measures and limited focus on higher-level comprehension (Zaretsky et al., 2009; Kelić et al., 2021; Kuvač Kraljević et al., 2024), developing psycholinguistically appropriate tools for Croatian is essential for valid assessment of reading comprehension.

Based on the analysis of studies, Blaži Ostojić (2023) provides an overview of formal and informal, as well as norms- and criterion-referenced approaches to reading comprehension assessment (Table 1.). Formal and informal assessments differ in their level of standardisation and structure, whereas norm-referenced and criterion-referenced assessments differ in the basis for interpreting performance, namely comparison with a normative sample versus evaluation against predefined criteria or learning objectives. Formal assessments commonly include standardized tests that provide norm-referenced data and enable comparisons across populations, although none have yet been developed for reading comprehension in Croatian. These instruments vary in focus, measured components, and administration methods, all of which contribute to interpretive variability. Criterion-referenced assessments are frequently experimental and allow researchers to manipulate linguistic stimuli to test specific hypotheses (Murray & Copens, 2013). Such assessments, however, require careful consideration of validity, reliability, and representativeness. They are typically divided into offline (post-reading comprehension tasks) and online (real-time processing) methods. Contemporary research emphasizes combining both approaches — offline measures to capture overall comprehension and online methods such as eye-tracking or neuroimaging to reveal the underlying cognitive processes during reading — as this dual approach offers a more comprehensive understanding of how readers process and construct meaning from text (Matić, 2020).

Table 1. Overview of Reading comprehension assessment methods (Blaži Ostojić, 2023)

<i>Norm-Referenced Tests</i>	<i>Criterion-Based Methods</i>
WECHSLER Individual Achievement Test (WIAT; Wechsler, 1992)	Offline methods: Cloze tasks, sentence verification, true/false or multiple-choice questions, open-ended comprehension questions (Cain & Oakhill, 2006)
Gates–MacGinitie Reading Test (MacGinitie et al., 2000)	Online methods: Eye-tracking, neuroimaging (fMRI, EEG; Rayner & Reichle, 2010; Ambridge & Rowland, 2013; Ozernov-Palchik et al., 2021), and pupillometry (Kahneman, 1973)
Diagnostic Assessment of Reading Comprehension (DARC; August et al., 2002)	
TORCH: Test of Reading Comprehension (Mossenson et al., 1987)	

Long established as the foundation of reading comprehension research, offline methods evaluate the outcome of comprehension rather than the underlying processes. These include picture selection, action performance, retelling, and accuracy tasks (Cain et al., 2001; Nation et al., 2007). Picture selection is useful for assessing inferential reasoning and syntactic understanding, action performance for assessing syntactic awareness, while retelling, though cognitively demanding especially for children with reading disorders, effectively captures discourse-level comprehension, and accuracy tasks are the most commonly used (Kintsch & van Dijk, 1978; Reed & Vaughn, 2012; Cao & Kim, 2021). However, the validity of these methods depends on participants' age, cognitive load, and assessment design, with struggling readers often requiring more support (Reed & Vaughn, 2012). Due to such limitations, researchers increasingly advocate online methods like eye-tracking, which offer more direct insight into real-time comprehension processes (Sabatini et al., 2013a; 2019b).

1.5.1 Eye-tracking in reading research

Techniques such as eye-tracking, EEG, and fMRI capture the temporal and neural dynamics of language processing (Ambridge & Rowland, 2013; Ozernov-Palchik et al., 2021). Online methods, particularly eye-movement tracking, allow the continuous observation of how readers process text without additional task interference (Rayner et al., 2006; Rayner & Reichle, 2010). The use of these specific measures is justified by the fact that they provide information about pre-lexical reading processes (such as lexical recognition) and post-lexical processes (e.g., syntactic, semantic, and pragmatic integration) (de-la-Peña, 2024). Unlike those tasks that require explicit decisions, eye-tracking relies on the assumption that where and for how long readers fixate reflects the cognitive processing of that specific region (Inhoff & Radach, 1998; Rayner, 1998). Researchers use these methods to examine the efficiency of text processing in readers, with shorter processing times generally indicating greater reading efficiency. They are also used to predict and explain performance in implicit reading comprehension tasks, such as how individuals read specific words (Kuperman et al., 2018), how they utilize working memory or make inferences (Eilers et al., 2018), how the type of task influences their understanding (Kaakinen & Hyönä, 2005), and whether the type of material they read (paper/screen) affects their comprehension (Jian, 2021). Additionally,

they are used to study the relationship with other skills, such as vocabulary, rapid automatized naming (RAN), and processing speed (de-la-Peña, 2024). Eye-tracking method can capture developmental changes in eye movement behaviour. By fourth or fifth grade, children show adult-like fixation durations and saccade lengths when reading age-appropriate texts (Rayner, 1986). Poor readers and readers with dyslexia, however, maintain longer fixations, shorter saccades, and more regressions, reflecting increased processing effort (Ashby et al., 2005; Chace et al., 2005). Thus, eye-tracking offers a major methodological advantage in reading research because it captures processing in a natural, time-sensitive, task-free reading environment, whereas measures such as lexical decision offer only a single reaction-time outcome (Rayner et al., 2006).

Recent studies (e.g., Luke et al., 2015; Ahn et al., 2020; Joseph et al., 2021; Mézière et al., 2023; de-la-Peña, 2024) have further refined key eye-tracking measures for investigating reading comprehension across linguistic and cognitive dimensions. Some conventional measures (e.g. dwell time, first fixation duration, number of fixations/ revisits, etc.) are frequently used to compare reading behaviours across different comprehension levels, text difficulties, and language proficiency (Ahn et al., 2020) and different language processing phases (Just & Carpenter, 1980; Staub & Rayner, 2007; Holmqvist et al., 2011). These measures are sensitive to different phases of processing during reading. Specifically, early measures reflect initial word access and recognition, capturing rapid visual and form-based access to its orthographic and phonological representation. In contrast, late measures reflect the additional processing to fit the word into the surrounding sentence or broader discourse, including semantic-syntactic integration and interpretive processes during reading. Furthermore, global measures (e.g., average fixation duration, saccade length) provide general indicators of reading efficiency at the sentence or text level, showing longer and more frequent fixations in less skilled readers such as children or individuals with dyslexia (Hyönä & Olson, 1995; Blythe & Joseph, 2011). However, they offer limited insight into specific cognitive mechanisms underlying comprehension differences (Mézière et al., 2023). In contrast, local measures focus on individual word processing, distinguishing between early measures, linked to initial lexical access, and late measures, associated with higher-order integration (Reichle et al., 2003; Clifton et al., 2007; Vasishth et al., 2013). Psycholinguistic word properties strongly shape these measures: higher word frequency, imageability, predictability, and age of acquisition reliably reduce fixation rates and times, and regressions (Ehrlich & Rayner, 1981; Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Marschark &

Cornoldi, 1991; Rayner & Well, 1996; Schilling et al., 1998; Juhasz & Rayner, 2003, 2006). These effects are also potentially influenced by task complexity, suggesting an interaction between cognitive and contextual factors in reading, and by individual differences (e.g. reading abilities; Jones et al., 2008; Kaakinen & Hyönä, 2010). Computational models such as E-Z Reader and SWIFT (Pollatsek, Reichle, & Rayner, 2006; Engbert et al., 2002, 2005) explain these patterns by assuming that lexical processing drives the eyes through text. Higher-order comprehension processes typically emerge only when the input is unexpected or anomalous, prompting longer fixations and regressions for re-analysis (Frazier & Rayner, 1982; Rayner et al., 2004).

Two measures are particularly central in eye-tracking research within the reading paradigm: First Fixation Duration (FFD) and First Pass Duration (Liversedge et al., 1998; Reichle et al., 2003; Staub, 2015). FFD reflects the earliest point at which lexical variables may influence processing and is defined as the duration (in milliseconds) of the first fixation within an Area of Interest (AoI) during the initial pass through the text, prior to any regressions. First Pass Duration captures the time required for the initial processing of a word before the gaze exits the AoI for the first time and corresponds to the sum of all fixation durations within the AoI during the first pass. In addition, Dwell Time (DT) presents the sum of all fixation durations during a dwell (from entry to exit) in an AoI (Holmqvist & Nystrom, 2011). It should be noted that, in the BeGaze software used in this study, DT includes not only fixation durations but also saccadic movements within the AoI, and thus corresponds to a measure of total gaze duration rather than a strictly fixation-based measure (SensoMotoric Instruments [SMI], 2014). Accordingly, the term *dwell time* is used throughout to refer to this broader, BeGaze-derived measure of total gaze duration.

For single words, key measures in addition to those mentioned above include: single fixation duration which refers to the duration of the first-pass fixation on a word that is fixated only once, total fixation time which refers to the sum of all fixations on a word, including those occurring after regressions, and reflects more sustained or integrative processing. Finally, fixation probability denotes the likelihood that the reader's gaze is directed toward a given area of interest (AOI), whereas regression probability reflects the probability that the gaze returns to a previously read region following fixation within the AOI. For larger regions, researchers use first-pass time, total reading time, and go-past time, which reflect both early parsing and later re-analysis (Rayner et al., 2006).

Both first fixation duration (FFD) and dwell time respond systematically to visual, lexical, and contextual properties: fixation times and dwell times increase when words are low in frequency or less predictable, and decrease when processing is facilitated. This reflects increased processing demands and less efficient lexical access, as well as the integration of complex or weakly specified representations, such as during anaphora resolution (Dambacher et al., 2013). Word length, word frequency, and predictability have been found to reliably influence first pass duration (Kliegl et al., 2004). Researchers frequently complement these measures with the probability of fixating a word and the frequency of regressions out of the word (regression out probability- RoP), both of which provide additional information about processing difficulty. Several studies have found shorter first fixation durations for high-frequency words in skilled, adult readers (Rayner, 1998; Rayner et al., 2011; Staub, 2015). Even in conditions with or without control of word length, frequency effect was found on first fixation duration and dwell time (Just & Carpenter, 1980; Rayner & Duffy, 1986). Shorter processing times typically indicate more efficient reading, while longer fixations or regressions suggest comprehension difficulties or higher text complexity (Rayner et al., 2006).

Robust frequency effect is found to be stronger in children than in adults, particularly when child-appropriate frequency norms are used (Joseph et al., 2013). During foveal processing, low-frequency words elicit longer and more numerous fixations than high-frequency words. This suggests that decisions about when to move the eyes are more tightly governed by cognitive processing demands in younger readers. Although most studies contrast only high- and low-frequency items, evidence indicates that the largest developmental differences are likely to emerge within the low-frequency range (Kuperman & van Dyke, 2013).

Research shows that short, within-word regressions often reflect minor oculomotor adjustments or brief difficulties in stabilising lexical processing, rather than genuine problems with comprehension. In contrast, longer regressions signal breakdowns in understanding, with skilled readers returning precisely to the source of difficulty and weaker readers engaging in broader, less targeted backtracking (Murray & Kennedy, 1988; Rayner, 1998). Frequent or highly imageable words typically produce shorter fixations, reflecting faster lexical processing, with these effects especially pronounced in children with dyslexia (Schilling et al., 1998; Jones et al., 2008; Kaakinen & Hyönä, 2010). In more complex texts, increased regressions indicate comprehension breakdowns (Hessel et al., 2021). Cook's (2014) experiment observed readers sensitivity to inconsistencies between anaphors and their

antecedents. She demonstrated that, although initial fixations on the anaphor did not differ across conditions, inconsistencies — whether involving high or low semantic overlap — triggered longer reading times in the spillover region, increased rereading of the anaphor, and more regressions compared with consistent pairs. Similarly, Rayner et al. (2009) found longer fixation durations and higher probability of regressions out of the area where the inconsistency occurred. Ashby et al. (2005) compared eye-movements of highly skilled and average readers during silent reading, observing frequency and predictability effects. They used measures such as first fixation duration, gaze duration, regression out probability (among others) to examine differential effects of target word-frequency across different processing phases. High-frequency words generally received shorter fixations and gaze durations in both groups, while readers rarely regressed out of the target region. The average readers regressed from the target word more frequently, but frequency did not have a significant effect on the percentage of regressions-out in either the overall analysis or the interaction with reading group. While these findings consistently link regression behaviour to processing difficulty and comprehension breakdowns, the precise functional role of regressions in reading remains a matter of ongoing debate, particularly regarding the extent to which they reflect targeted linguistic reanalysis versus more general processing disruptions.

Classic findings by Frazier and Rayner (1982) show that, when readers realise their initial syntactic analysis is incorrect, regressions are directed not to the start of the sentence but precisely to the point where the misanalysis likely occurred. Likewise, Meseguer et al. (2002) demonstrated that, during referential processing, readers selectively return to the antecedent region when resolving a mismatch. These results indicate that regressions reflect targeted reanalysis, not random backtracking. Late detection of ambiguity or inconsistency may prompt a selective return to the antecedent, particularly among readers who actively monitor textual cohesion. Wotschack and Kliegl (2011) showed that readers strategically adjust their regression behaviour when they anticipate higher comprehension demands, and Weiss et al. (2018) extended this finding by demonstrating that such adjustments occur after first-pass processing: more challenging comprehension expectations led to increased regressions to earlier sentence regions, while early fixation measures remained unaffected. Research on semantic anomaly detection provides strong evidence that readers often identify inconsistencies after completing initial lexical processing. Rayner et al. (2004) demonstrated that semantic anomalies do not always disrupt early eye movement measures such as first-fixation or gaze duration. Instead, skilled readers frequently show delayed anomaly effects,

emerging as increased regressions and longer second-pass reading once the broader sentence or discourse context has been evaluated. This pattern indicates that readers first commit to a rapid lexical interpretation and only subsequently verify whether that interpretation fits the evolving discourse model. Although older primary-school children can detect inconsistencies within and across sentences, many struggle to repair them, and weaker comprehenders show difficulties in updating their situation model and distinguishing relevant from irrelevant information (van der Schoot et al., 2008a, 2012b). In contrast, older adults compensate for slower processing by investing more effort in building a coherent discourse representation: they make fewer but longer fixations, skip words more often, and produce more regressions to check skipped material, reflecting a strategic adaptation to maintain comprehension (Laubrock et al., 2006; Rayner et al., 2006). A substantial body of research on predictive processing shows that readers do not passively wait for input but use lexical information to generate forward-looking expectations about upcoming content. Studies by Altmann and Kamide (1999) and Kamide et al. (2003) show that readers use lexical information to form rapid predictions about upcoming content. According to their findings, as soon as a word is encountered, readers activate a set of likely continuations and restrict the domain of reference even before the critical noun or anaphor appears. In other words, richer and more highly accessed lexical items (e.g., high-frequency or highly imageable words) generate stronger predictive commitments about how subsequent sentence elements should fit into the developing discourse. This pattern suggests that prediction facilitates early processing but increases the need for later coherence checking.

Combining early and late measures therefore enables a comprehensive understanding of both lexical and post-lexical processes (Joseph et al., 2021). Moreover, eye-tracking data have also been used successfully to predict reading comprehension outcomes, with studies achieving classification accuracies of up to 89% through machine learning (Copeland et al., 2014) and explaining up to 46% of variance in standardized comprehension scores (Mézière et al., 2023). Despite these promising results, the predictive link between eye-movement patterns and comprehension level remains insufficiently understood (Rayner et al., 2006). It is also important to note that eye-tracking results obtained from discourse-level stimuli may differ from those based on isolated single-word presentations. Isolated words reflect only their intrinsic lexical properties, whereas words embedded in coherent discourse are influenced by contextual support and repeated exposure. Consequently, a low frequency word that recurs within a passage is quickly accommodated and begins to behave like a high frequency item,

producing shorter fixations and reduced frequency effects (Desai et al., 2020). Another important point is that, beyond developmental factors, there are substantial intra- and inter-individual differences in readers' eye-movement behaviour, many of which remain not yet fully understood (Schroeder et al., 2015). Future research should focus on improving methodological consistency by predefining measures, analyses, and interpretation criteria to enhance reliability and diagnostic validity (Joseph et al., 2021).

2. THE PRESENT STUDY

This doctoral study aims to contribute to the existing findings on reading comprehension and language-based disorders by addressing several critical gaps in current literature. It examines how psycholinguistic word properties (imageability and frequency) influence comprehension during anaphora resolution in both online i.e. real-time and offline i.e. unconstrained time, using eye-tracking and question-answering methods. This study integrates several theoretical frameworks: the Two-dimensional model (Bishop & Snowling, 2004) as a basis for differentiating typical and atypical language profiles, the Fully Distinct Hypothesis (Catts et al., 2006) to further classify subgroups of language-based disorders (in line with the SVR model: poor decoders, poor oral language, and mixed language difficulties group), the Simple View of Reading (Gough & Tunmer, 1986) to analyze components of reading comprehension, and the Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti, 2007) to emphasize the impact of lexical representation quality on reading comprehension.

Importantly, previous research has largely focused on English and other widely studied languages. In contrast, this study examines Croatian, a typologically distinct and understudied language in psycholinguistic reading research, particularly in real-time paradigms, thereby offering insights about reading comprehension processes and behavioural features of Croatian-speaking children with typical language development and language-based disorders that are not fully captured in existing literature.

There were three general goals of the present study:

1. to examine how psycholinguistic word properties — specifically frequency and imageability — affect the processing of anaphoric relations (nominal antecedents and null anaphora) during reading comprehension in children with typical language development and language-based disorder profiles.
2. to examine the role of question type (literal vs. inferential) and language-based disorder profile (poor decoders, poor oral language group, and mixed language difficulties group) in

shaping reading comprehension outcomes, to better understand how different reader profiles manage integration of textual information.

3. to explore how individual differences in (meta)linguistic skills — phonological skills and awareness, semantic knowledge, inferential reasoning, and phonological working memory - contribute to reading comprehension performance (accuracy on literal and inferential questions) in Croatian-speaking, school-aged children regarding to the child's language profile.

The following research questions and hypotheses are posed:

1. What is the effect of specific (psycho)linguistic word properties (frequency and imageability) on the processing of anaphoric relations during text comprehension in children with language-based disorders and children with typical language development?

H1: Psycholinguistic word properties — specifically frequency and imageability — will influence the processing of anaphoric relations in both children with language-based disorders and children with typical language development.

H1.1: Both groups of children will exhibit longer fixation durations on low-frequency words compared to high-frequency words.

H1.2: Both groups of children will exhibit longer fixation durations on low-imageability words compared to high-imageability words.

H2: Psycholinguistic word properties — frequency and imageability — will exert a greater influence on anaphoric relations (i.e. anaphora processing) during text comprehension in children with language-based disorders compared to children with typical language development.

H2.1: The effect of word imageability will be more pronounced in children with poor reading comprehension, such that the difference in processing low- versus high-imageability words (in terms of longer gaze durations in the region of interest, increased number of fixations, longer total fixation durations, and a greater number of regressions) will increase as reading comprehension scores decrease.

H2.2: The effect of word frequency will be more pronounced in children with poor decoding skills, such that the difference in processing low- versus high-frequency words (in terms of longer gaze durations in the region of interest, increased number of fixations, longer total

fixation durations, and a greater number of regressions) will increase as reading comprehension scores decrease.

2. What is the effect of question type (literal vs. inferential) and type of language difficulty (poor decoders vs. children with poor reading comprehension) on text comprehension?

H3: Both children with typical language development and children with language-based disorders will demonstrate lower accuracy on inferential questions compared to literal questions.

H4: The type of question (literal vs. inferential) will differentially affect reading comprehension depending on the language-based disorder profile, such that:

H4.1: Poor decoders will show difficulties in accurately answering both literal and inferential questions.

H4.2: Children with poor reading comprehension will show greater difficulty accurately answering inferential questions compared to literal ones.

3. What is the predictive role of cognitive (working memory) and various (meta)linguistic measures in discourse comprehension, over and above the child's clinical language profile?

H5: Phonological working memory (phonological loop), (meta)phonological skills, semantic skills, and inferential reasoning ability will significantly predict performance in discourse comprehension.

H5.1: There is a significant positive correlation between phonological working memory and discourse comprehension.

H5.2: There is a significant positive correlation between (meta)linguistic measures (phonological and semantic) and discourse comprehension.

H5.3: There is a significant positive correlation between inferential reasoning skills and discourse comprehension.

H5.4: Each independent variable (phonological working memory and (meta)linguistic skills) will serve as a significant individual predictor of discourse comprehension success.

3. METHOD

3.1 Participants

Participants with typical language development

Children with typical language development were recruited from four primary schools in Zagreb (Sesvete, Iver, Sv. Klara, and Retkovec primary school; 5th and 6th grades), and all participants were monolingual speakers of Croatian. These children were enrolled in regular schools and had no history of language-based disorders, speech-motor impairments (e.g., childhood apraxia of speech), fluency disorders (e.g., stuttering or cluttering), neurodevelopmental communication disorders (e.g., autism spectrum disorder), attention-deficit/ hyperactivity disorder, motor impairments, hearing impairments, visual impairments, or previous participation in speech-language therapy. An additional inclusion criterion was performance within the average range on a nonverbal cognitive abilities test — Raven's Progressive Matrices (standard score ≥ 85).

Participants with language-based disorders (atypical language development)

Participants with language-based disorders were primarily recruited from primary schools in Zagreb (Sesvete, Iver, Sv. Klara, and Retkovec primary schools) and the Speech and Language Rehabilitation Center Blaži. Two additional participants were recruited from the SUVAG Polyclinic. All participants were 5th- and 6th- grade primary school students.

To distinguish disorder from typical development in terms of language abilities and decoding skills, only participants with a prior diagnosis of dyslexia and/or developmental language disorder (DLD) were included, using the Two-dimensional model proposed by Bishop and Snowling (2004). In Croatian clinical practice, diagnostic classification relies exclusively on the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10/MKB-10; World Health Organization, 2012). These prior diagnoses were therefore established through clinical assessment and coded as specific developmental disorder of scholastic skills (F81.0) for dyslexia and specific developmental disorder of speech and language (F80.1/F80.2) for developmental language disorder. In addition to the clinical ICD-10 diagnosis, participants' language profiles were further verified against the consensus diagnostic criteria for developmental language disorder proposed by Bishop et al. (2016, 2017), which provided an internationally recognised, non-exclusionary operationalisation of the construct beyond the categorical ICD-10 code.

The diagnostic procedure included comprehensive evaluations of language abilities (both comprehension and production), reading and writing skills and assessments of phonological awareness. All participants had normal hearing and normal or corrected-to-normal vision, and those with known neurological disorders, including epilepsy, were excluded from the sample.

Furthermore, all participants underwent an additional assessment of their (meta)linguistic skills during the pre-assessment phase, which also served to validate their diagnostic profiles. Based on this behavioral and cohort-based analysis, and in line with the Fully Distinct Hypothesis (Catts et al., 2005), participants were classified into subgroups, according to their language profiles: dyslexia, developmental language disorder - DLD, or comorbid dyslexia and DLD.

In addition to the speech-language assessment, the diagnostic protocol included psychological assessment. To assess nonverbal cognitive abilities, the Raven's Progressive Matrices test was administered during the pre-assessment phase, and only those children whose scores fell within the typical range (standard score ≥ 85) were included in the study. Children diagnosed with attention-deficit/ hyperactivity disorder (ADHD), fluency disorders (e.g., stuttering or cluttering), neurodevelopmental communication disorders (e.g., autism spectrum disorder), and similar conditions were excluded from participation.

All participants attended regular primary education programs, and their reading difficulties were not attributable to unfavorable socioeconomic or cultural factors. All participants were monolingual Croatian speakers, ensuring linguistic homogeneity within the sample.

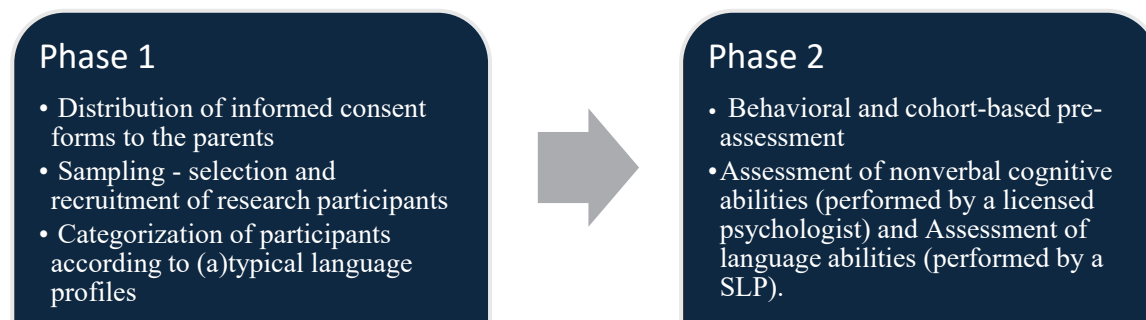
3.2 Sampling procedure

Ethical approval for this study was obtained from the Ethical Board of the Faculty of Education and Rehabilitation Sciences, University of Zagreb (Class: 602-04/19-42/24; Reference No.: 251-74/19-01/2); the Ministry of Science, Education and Youth, Republic of Croatia (Class: 602-02/20-03/00116; Reference No.: 533-05-20-0005), and from the Ethical Board of the SUVAG Polyclinic (Class: 510-08/20-20; Reference No.: 262-01-2020-43). Informed consent was provided to the heads of all institutions (e.g. primary schools) where the research was conducted.

Parents and guardians of potential participants were informed about the opportunity to participate in the study and then received a printed informed consent form with a detailed explanation of the study. Those who agreed to the terms signed the consent form. They could request additional information from the researcher at any time by email or telephone. In

addition, the children were informed verbally, in age-appropriate language, about the purpose, objectives, and procedures of the study, with clear emphasis on the voluntary nature of participation and the right to withdraw at any time without consequences.

The recruitment process was conducted in several phases (discussed below) to ensure the validity, precision, and theoretical consistency of participant classification in line with the aims of the study.



Phase 1: Initial recruitment

In the first phase, a total of 84 participants were recruited, as presented in the Table 2 below:

Table 2. Initial participants groups and sample sizes

<i>Participant Group</i>	<i>Number of Participants</i>
<i>Children with language-based disorders (DLD and/or dyslexia)</i>	44
<i>Children with typical language development</i>	40
<i>Total</i>	84

Phase 1 served as the initial clinical and educational selection based on external identification of language status, in accordance with the Two-dimensional theoretical model proposed by Bishop and Snowling (2004), which defines the criteria for typical language development and language-based disorders. In this phase, children with established DLD and/ or dyslexia (initial N = 44) were referred by licensed speech-language pathologists from institutions such as the Speech and Language Rehabilitation Center Blaži, SUVAG Polyclinic and primary schools. Subsequently, through additional screening and assessment, and within the framework of the Fully Distinct Hypothesis (Catts et al., 2005), aligning with the SVR model (Gough & Tunmer, 1986; Hoover & Tunmer, 2020), participants were first classified into the language-based disorder group (LBD), and then further into the respective subgroups, based on their specific language profiles. Simultaneously, children with typical language development were referred

by classroom teachers (initial N = 40), from four mainstream schools (5th and 6th grades) in Zagreb, with the support of school-based speech-language pathologists. Although this sample was primarily purposive, a random selection was made at a later stage of the selection process: 40 children from the screened classrooms were randomly selected to form the initial group of typically developing peers. Participants were matched by age and gender, so that each child with language-based disorder was paired with a typically developing peer of the same age and gender.

This dual recruitment strategy enabled the inclusion of both atypically and typically developing participants within a continuum of language abilities, consistent with the Fully Distinct Hypothesis (Catts et al., 2005) and supporting the broader conceptualization of language-based disorders advocated by Snowling and Hulme (2012). Exclusion criteria for all groups included the presence of motor, cognitive, visual, or auditory impairments.

Phase 2: Pre-assessment and reclassification

Raven's Standard Progressive Matrices (Raven et al., 1999) were used to assess nonverbal cognitive functioning (g factor - reasoning, perception, and relational insight) and served as an inclusion criterion. Participants who scored below 70 were excluded from further participation (N = 2). All psychological data were analysed by a qualified psychologist.

All initially recruited participants underwent an additional assessment, consisting of a series of standardized tasks and tasks taken from other research on (meta)language abilities, categorized into three functional domains relevant to reading comprehension (Gough & Tunmer, 1986; Catts et al., 2005): phonological and non-phonological measures (e.g., phonological awareness, vocabulary depth, pseudoword repetition), along with phonological working memory (Baddeley, 2003). These tasks are discussed in more detail in the following section *Materials and Procedure*.

The aim of this pre-assessment phase was to validate the initial diagnostic categories and to enable a more fine-grained, performance-based classification of language profiles in accordance with the Fully Distinct Hypothesis (Catts et al., 2005) and furthermore within the framework of the Simple View of Reading (Gough & Tunmer, 1986). This step was critical because previous research has shown that clinical labels alone are often insufficient to capture the full variability of language and reading skills in school aged children (Bishop & Snowling, 2004; Ramus, 2013).

The analysis began with a descriptive examination of the control group data to establish reference values for interpreting individual participant scores in the experimental groups. Participants were then categorized into a group of children with language-based disorders (LBD) and a group of children with typical language development (TLD) based on their performance on language variables. The classification criterion required participants to score at least 1.25 standard deviations below the mean of the control group ($M - 1.25 SD$) on two or more language measures. For variables where higher scores indicated poorer performance (e.g., number of errors), the threshold was set at $M + 1.25 SD$. These thresholds corresponded approximately to the 10th percentile and were used to identify participants with more pronounced language difficulties, in line with previous research (e.g., Tomblin et al., 1997; Bishop et al., 2016). A detailed rationale for the classification thresholds can be found in the *Data Analysis* chapter. These participants were classified as clinically significant cases and analyzed further in the subsequent phases of the study.

After excluding two participants (one from the typical language development group and one from the language-based disorder group) due to their scores on the Raven's Progressive Matrices, of the 43 participants with an initial diagnosis of DLD/ dyslexia, 5 were reclassified as typical. Of the 39 participants initially categorized as typical, 11 met the criteria for language-based disorders. As a result, 49 participants were categorized as having language-based disorders and 33 as typical language development. However, within the language-based disorders group, 9 children showed inconsistent ('unexplained') profiles (e.g., isolated impairments in spoonerisms and vocabulary). These cases were excluded to ensure analytical clarity. Subsequently, three additional participants were excluded after analyzing the eye-tracking data - scan paths (e.g., repeated gaze aversion, distractibility). Descriptive data for the final sample are shown in Table 3.

Table 3. Age and Gender of Participants in the Final Sample

<i>Participants</i>	<i>N</i>	<i>Min Age</i>	<i>Max Age</i>	<i>Mean Age (M)</i>	<i>SD</i>	<i>Male</i>	<i>Female</i>
<i>All participants</i>	70	11.00	13.33	11.97	0.53	38	32
<i>Typical language development</i>	33	11.00	13.00	11.86	0.44	15	18
<i>Language-based disorder</i>	37	11.00	13.33	12.01	0.57	23	14

Final sample profiles

Classification of the participant profiles based on pre-assessment data

To further differentiate the participants within the language-based disorder group, the individual performance profiles were mapped to the reading components defined by the SVR (i.e., decoding and language comprehension). A performance was categorised as clinically significant if a participant had values below the threshold for at least two variables. Based on this criterion, 70 participants were classified into four SVR-based profiles (Table 4).

Table 4. Classification Matrix of Participants According to the Simple View of Reading (SVR) Framework Based on Performance in (Meta)Language Measures

Group	SVR Profile	Phonological skills (decoding)	Non-phonological skills (language comprehension)	Phonological loop	Interpretation of the performance	Number of participants
Typical language development (TLD)	TLD	Average	Average	Average	Adequate decoding and language comprehension components	33
Language-based disorders (LBD)	Poor Oral Language Group (POL)	Average	Poor (below threshold on ≥ 2 out of 6 variables)	Poor	Specific difficulties in language comprehension (e.g., DLD)	10
	Poor Decoders (PD)	Poor (below threshold on ≥ 2 out of 5 variables)	Average	Poor	Specific decoding difficulties (e.g., dyslexia)	16
	Mixed Language Difficulties (MD)	Poor (below threshold on ≥ 1 out of 5 variables)	Poor (below threshold on ≥ 1 out of 5 variables)	Poor	Specific difficulties in both components - decoding and language comprehension	11

To facilitate the classification process, a performance matrix was constructed to map individual measures and task outcomes to the language profiles defined by SVR, in accordance with the Fully Distinct Hypothesis (Catts et al., 2005). This approach allowed for a theoretically sound and empirically robust classification of Croatian-speaking children - a population for which such profile-based analyses had previously been possible only to a limited extent. Consequently, the final sample comprised 70 participants, who were systematically categorised and grouped into four theoretically motivated subgroups, allowing for a differentiated analysis of reading comprehension and anaphoric processing across different language profiles. Rather than reiterating the specific thresholds and diagnostic criteria in the text, the characteristics of each subgroup are presented in Table 4. This matrix summarises the inclusion criteria, skill profiles, and the group sizes, thus providing a concise and transparent overview.

Classification Thresholds and Rationale

In the absence of standardised norms for all administered measures and to classify participants according to the presence of language difficulties, a normative group of typically developing children was used to establish internal thresholds. These thresholds served as empirical benchmarks and were used to assign binary classification scores (i.e., presence or absence of difficulties) for all variables.

A threshold criterion was applied based on the mean and standard deviation of the reference (control) group, most often defined as $M - (1.25 \times SD)$. This approach identified participants whose performance was significantly below the average level of language function, regardless of whether the individual variables followed a normal distribution within the whole sample. Importantly, the thresholds were based solely on the distribution of typically developing children to avoid bias, as including scores from the clinical group would lower the mean and increase the variance, making it harder to detect genuine language difficulties.

The applied threshold of ± 1.25 SD corresponded approximately to the 10th percentile (Appendix No. 2), which aligns with previous research indicating that about 10% of children in the general population are affected by developmental language disorder (Tomblin et al., 1997; Bishop et al., 2016). This approach allowed for a functional and practical identification of disorder, rather than a purely statistical one. Although some deviations from normality

were observed, the thresholds were retained as a classification criterion rather than a statistical test. This allowed the same standard to be applied to all variables and ensured consistent identification of participants with more pronounced difficulties.

Such an approach is well-established in the literature on language disorder classification and is commonly used when fully normed measures are not available for all relevant language domains (Tomblin et al., 1997; Keenan & Meenan, 2014; Leonard, 2014; Bishop et al., 2016; Bawayan & Brown, 2022; Selin et al., 2022). It ensures methodological transparency and reproducibility of the analysis, while also allowing for the inclusion of participants who objectively deviate from the expected levels of language competence.

Theoretical and methodological rationale for multi-phase sampling approach

This structured, theory-driven sampling approach was designed to align clinical diagnosis with standardized performance measures while capturing the heterogeneity of language-related skills beyond binary diagnostic categories. It enables meaningful subgrouping based on SVR profiles, thereby strengthening internal validity. The focus on Croatian-speaking children aged 11-13 ensures a relatively homogeneous developmental stage at which participants can reliably engage in discourse-level reading tasks.

Overall, this methodology provides a robust framework for examining individual differences in discourse-level reading comprehension in middle childhood within a Croatian linguistic and educational context.

3.3 Research materials, tasks and methods

Description of variables

The following dependent variables were analysed in the study:

- Children's performance on standardized tests (language measures) and experimentally constructed tasks (accuracy on literal and inferential comprehension questions);
- Raw data on eye movements, including fixation duration, regressions, and measures related to the area of interest (AOI);
- Interaction between (psycho)linguistic word properties and anaphoric resolution.

The independent variables included:

- Group membership: two main or independent groups (children with language-based disorders and children with typical language development), and four subgroups - three

with language-based disorders (poor oral language group, poor decoders, and mixed language difficulties group) and one control group;

- Phonological and non-phonological (meta)linguistic skills;
- Phonological working memory (phonological loop);
- Lexical complexity (based on individual psycholinguistic word properties).

Cohort and behavioral assessment

Cohort and behavioural assessment included a battery of phonological and non-phonological tasks, as well as phonological working memory task, as presented in the Table 5.

Table 5. Overview of (Meta)linguistic and Phonological Working Memory Measures Used for Language Profile Classification

Phonological Skills	Non-Phonological Skills	Phonological Working Memory (Phonological Loop)
Word and pseudoword reading-accuracy and time (Kelić, 2017)	PPVT-III-HR - receptive vocabulary (Dunn et al., 2010)	Pseudoword repetition (Kelić, 2017)
Phoneme manipulation - phonological awareness (Kelić, 2017)	TROG-2:HR - grammatical comprehension (Bishop et al., 2014)	
	Verbal semantic fluency task (Bousfield, 1953; Riva et al., 2001)	
	Inferential reasoning task (NRDLS-HR; Edwards et al., 2019)	
	Grammaticality judgment task (NRDLS-HR; Edwards et al., 2019)	
	Vocabulary depth - metalinguistic knowledge (WISC-IV-HR; Wechsler, 2009)	

For standardized tests (PPVT-III-HR, TROG-2:HR, NRDLS subtasks), both raw scores and percentile ranks were calculated according to the test manuals. For non-standardized, non-norm referenced measures (e.g., word and pseudoword reading, phoneme manipulation, semantic fluency), performance was evaluated using raw scores (e.g., number of correct responses, errors, or words produced) and, where appropriate, task-specific criteria were applied (e.g., reading time, fluency cut-off values based on normative studies). For each variable, the arithmetic mean (M) and standard deviation (SD) were computed, and these

values served as the basis for establishing thresholds that differentiate typical performance from scores suggestive of potential difficulties.

Phonological (Meta)Linguistic tasks

The assessment of phonological (meta)linguistic skills consisted of two tasks: a word and pseudoword reading task, and a phonological awareness task. Word recognition, defined as the ability to recognize words quickly and accurately as wholes, involves visual identification and accessing their meaning in the mental lexicon. Children with difficulties in this area - such as those with dyslexia - often struggle with automatic word reading, which can slow reading and hinder text comprehension (Catts et al., 2005; Ramus et al., 2013). In contrast, phonological processing refers to the ability to use and process phonological information, i.e., the sounds of language, during reading and speaking. Difficulties in phonological processing manifest as problems in decoding new or unfamiliar words (e.g., pseudoword reading), as the child may not be able to correctly map phonemes to graphemes or retain phonological information during reading, which impairs comprehension (Ramus et al., 2013).

Word and Pseudoword Reading. Children were asked to read aloud two lists without a time limit, one of real words (N=104) and one of pseudowords (N=44) developed and validated by Kelić (2017). The word and pseudoword lists were presented in printed form on separate sheets, using Calibri font (19 pt), with items ordered progressively by grapheme length. Participants were instructed to read each list as quickly and accurately as possible. For the pseudoword list, an additional instruction clarified that some items were not real words with meanings but should still be read exactly as written. For example, items included familiar words such as *konj* and unfamiliar pseudowords such as *mijeza*. These lists were compiled based on syllable count, syllable structure (CV, CCV, etc.), and grapheme frequency, respecting the phonological structure of the Croatian language.

During reading, the examiner recorded both the number of reading errors (reading accuracy) and the time taken to complete each list (reading speed), as these measures are key indicators of phonological processing and reading automatization.

This procedure allowed a distinction between word recognition - rapid, accurate retrieval of familiar words from the mental lexicon - and phonological processing, which involves decoding unfamiliar letter strings by mapping graphemes to phonemes. Whereas deficits in word recognition typically hinder fluency and comprehension in children with dyslexia (Catts et al., 2005), weaknesses in phonological processing are revealed through difficulties in

pseudoword reading and reflect broader challenges in phonological coding (Ramus et al., 2013).

Phonological Awareness (Phoneme Manipulation - Spoonerisms). Phonological Awareness was assessed using a phoneme manipulation task in which children were instructed to swap the initial phonemes of paired words to form two new words (e.g. *šapa - kuma* → *kapa - šuma*; Kelić, 2017). After two practice trials, instructions were repeated if the participant did not understand the task, and up to two additional practice trials were administered until the child understood that, after hearing two words, they needed to switch their first sounds. Experimental items varied in phonological complexity (for details see Kelić, 2017). Responses were scored for accuracy, reflecting the ability to isolate and manipulate phonemes as an index of higher-order phonological awareness (Ramus et al., 2013).

The rationale for including these tasks is based on evidence that children with language-based disorders may perform poorly in all these subtasks, depending on their specific profile of difficulties (Ramus et al., 2013).

Non-phonological (Meta)Linguistic tasks

In line with the Simple View of Reading (Gough & Tunmer, 1986) and the classification criteria proposed by Catts et al. (2005), these tasks were included to assess oral language and metalinguistic knowledge at different levels of processing. Their purpose was to identify children with specific oral (receptive) language disorder - poor oral language group (who share characteristics with profiles often described as poor comprehenders in the literature) and those with broader mixed language difficulties (e.g. mixed group), thereby distinguishing them from children whose reading problems were primarily phonological.

Receptive vocabulary. Receptive vocabulary was assessed using the Croatian version of the Peabody Picture Vocabulary Test (PPVT-III-HR; Dunn et al., 2010). The test includes four practice items and 204 lexical items organized into 17 sets of 12 tasks. For each item, the examiner pronounced a target word, and the child was asked to select the corresponding picture from four (black-and-white) illustrations, one representing the target meaning and the others serving as distractors that could be semantically, visually, phonologically, or entirely unrelated (e.g., hearing *blizzard* and choosing among pictures of snow, rain, or an unrelated object). The testing stops after eight consecutive failed items. The procedure and instructions were followed as specified in manual. The total number of correctly identified items was used as the main outcome measure, providing an index of vocabulary breadth and efficiency in

semantic mapping. Children with DLD often achieve lower scores (Betz et al., 2013; Škoro, 2024), particularly when distractors are semantically close, reflecting limitations in forming precise lexical representations.

Grammatical comprehension. Grammatical comprehension was assessed using the Croatian version of the Test for Reception of Grammar (TROG-2:HR; Bishop et al., 2014), which assesses comprehension (Croatian) grammar in terms of inflectional morphology, function words, and word order. The TROG-2:HR consists of 80 items, each requiring the child to match a spoken sentence with one of four (coloured) pictures, with distractors differing in grammatical or lexical elements (multiple-choice with images, e.g. distinguishing *the boy is chasing the girl* from *the girl is chasing the boy*, or identifying contrasts in function words such as *on* versus *under*). The procedure and instructions were followed according to the manual. Items are grouped into blocks of four per grammatical contrast, all of which must be answered correctly. These blocks increase in difficulty, and testing stops after five consecutive failed blocks. Accuracy, calculated as the number of correctly solved blocks, was used as an index of morphosyntactic understanding. This measure typically reveals marked weaknesses in children with DLD, especially in the comprehension of inflectional morphology and word order (Bishop & Snowling, 2004; Botting et al., 2005; Galeković, 2023), whereas children with dyslexia often perform within normal limits, suggesting that their difficulties lie elsewhere in the reading process (Bishop & Snowling, 2004; Catts et al., 2005).

Verbal semantic fluency. Lexical-semantic retrieval was assessed using a semantic fluency task, in which participants were asked to name as many words as possible belonging to the category *animals*, within one minute (e.g. *dog, cat, elephant*; Bousfield, 1953; Troyer, 2000; Abwender et al., 2001), a task previously used with Croatian-speaking children to assess reading-related language abilities (Lenček, 1994). The score was the total number of correct, non-repeated words generated, with performance below 12 words per minute considered impaired (Riva et al., 2001; Sauzéon et al., 2004). This task is considered a measure of semantic (categorical) verbal fluency specifically, indexing the efficiency and organisation of lexical-semantic retrieval from long-term memory, including the ability to generate exemplars within semantic subclusters (clustering) and to shift flexibly between subcategories (switching; Troyer, 2000). Semantic fluency draws primarily on organised conceptual knowledge rather than decoding or phonological processing. It was therefore included among the non-phonological (meta)linguistic tasks, contributing to the distinction

between children with primarily phonological reading difficulties (dyslexia) and those with broader oral language weaknesses (poor oral language/DLD or mixed group).

Inferencing. Inferencing was assessed using the Inferencing subtask from the Croatian version of the New Reynell Developmental Language Scales (NRDLS, Edwards et al., 2019). Participants were presented with a coloured picture depicting a sequence of familiar, culturally appropriate characters, events, and settings (e.g. a children's playground). Their ability to draw indirect inferences, beyond what was explicitly depicted (e.g. identifying why someone might not be eating ice cream), was assessed by answering specific questions (N=10) asked by the examiner about the characters (e.g. emotions, beliefs, intentions) and events. The task measured the ability to integrate contextual and linguistic information thus capturing higher-level comprehension skills. Performance was scored by the number of correct answers reflecting successful inference-making. Each correct response was awarded one point. In cases where questions could plausibly elicit more than one answer, any response that was logical and consistent with the context was also credited. The procedure and instructions were applied according to the manual.

Grammatical Judgement. Grammatical awareness was assessed using the Grammaticality Judgement subtask from the New Reynell Developmental Language Scales (NRDLS-HR), in which children listened to sentences that were either grammatically correct or contained an error (e.g. *She going to school yesterday*). The participant's task was to decide whether the sentence was correct and, if not, to explain the error and provide the corrected form. The task consisted of eight items in total, six of which contained a target error to be identified (in noun or verb morphology), while other two sentences served as fillers. Each item was scored on a scale from 0 to 2 points. A score of 0 was given for an incorrect response, 1 point if the participant detected the error but could not provide the correct form, and 2 points if the error was both identified and successfully corrected or explained. Thus, performance was scored based on the accuracy of error detection and correction, reflecting the child's metalinguistic knowledge of grammar, with maximum of 14 points.

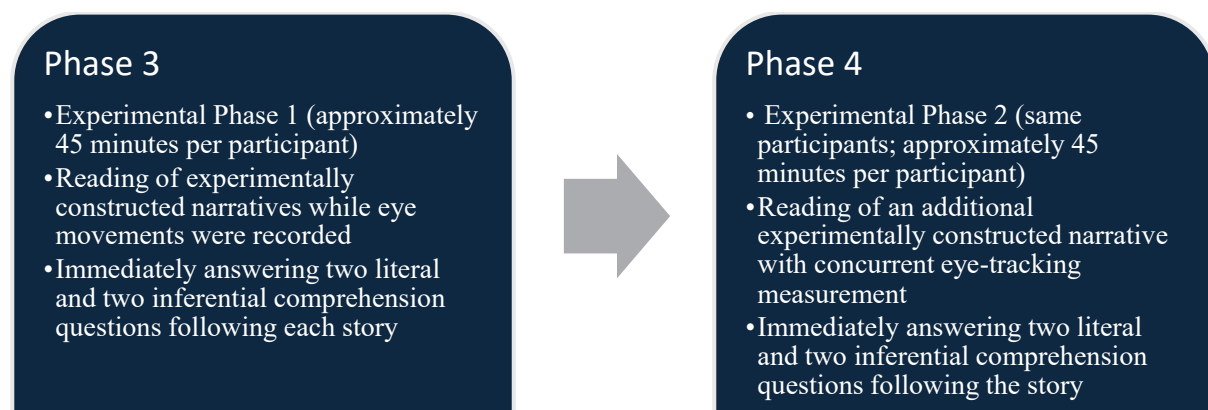
Vocabulary test. Vocabulary depth was assessed using the Vocabulary subtest of the WISC-IV-HR (Wechsler, 2009), comprising 32 items that require participants to define orally presented target words. This task evaluates how comprehensively and precisely a child understands word meanings. Participants are presented with words of increasing difficulty and asked to provide clear and accurate definitions (e.g., “What does the word *umbrella*

mean?”). Performance is scored out of a maximum of 68 points based on the accuracy and completeness of these definitions. This task provides an age-sensitive measure of semantic language development and reflects the child's depth of vocabulary, as well as their metalinguistic awareness, and ability to articulate abstract word meanings (Artuso & Palladiono, 2022).

Phonological working memory (phonological loop). The phonological loop was assessed using a pseudoword repetition task consisting of a list of 16 items that varied systematically in length (2-5 syllables) and phonological complexity (e.g. *tuma*, *mistakosnik*; Kelić, 2017). Children listened to each pseudoword read aloud by the examiner and were asked to repeat it exactly. Responses were scored for accuracy (total number of correctly repeated items), providing a measure of temporary storage and reproduction of phonological sequences independent of lexical knowledge.

3.4 Experimental procedure

The following section describes the experimental procedure (Phases 3 and 4) through which the core research data were collected.



Phase 3 and 4: The experiment

The experimental phase of the study involved both online (eye-tracking) and offline (question-answering) testing. The value of this combination lies in the complementary data provided by the two methods: offline measurements provide insight into the outcome of a specific cognitive-language process (i.e., success or failure in performing a task), while real-time methods - especially eye-tracking - reveal the underlying processes occurring during task performance. In addition, real-time methods enable the identification of immediate processing difficulties experienced by individual participants.

Both experiments took place in specially adapted rooms at the schools and institutions from which the participants were recruited. All tests were conducted individually.

3.4.1 Development of the experimental materials

In accordance with the Lexical Quality Hypothesis (Perfetti & Hart, 2002), four short narratives were constructed: two experimental narratives in which the complexity of (psycho)linguistic word properties (especially frequency and imageability) was manipulated across four paragraphs, and two structurally identical texts created to enhance the reliability and validity of the data. To control for the potential influence of topic familiarity on reading comprehension, the thematic content was based on data from the international project *Tell Me a Story: Personal Event Narratives from 10-Year-Old Children Around the World* (Westerveld et al., 2021), conducted by the Child Language Committee of the International Association of Communication Sciences and Disorders (IALP). From the collected database of personal narratives by 10-year-old Croatian speakers and the topic analyses, the theme of each story was defined. Each story focused on a specific theme, with no thematic overlap between the stories.

The paragraphs within each story were systematically varied according to the experimental conditions, defined as specific combinations of word frequency and imageability. Each paragraph represented one such condition. Discourse-level cohesion, syntactic complexity, sentence length, number of words, and word length were carefully controlled during the construction of the materials to ensure comparability across conditions. Each story consisted of four paragraphs of equal length and structure, with the same number and type of sentences (both simple and complex), and main clauses consistently preceding subordinate clauses.

The psycholinguistic properties of the target words, specifically frequency and imageability, were taken from the Croatian Lexical Database (<http://polin-hlb.erf.hr/>; Kuvač Kraljević & Olujić, 2018) and the Croatian Psycholinguistic Database (http://megahr.ffzg.unizg.hr/hr/?page_id=339; Peti Stantić et al., 2019). Both frequency and imageability were rated on a 5-point Likert scale, with greater values indicating higher frequency of occurrence and greater ease of forming a mental image, respectively. For this study, words were categorised as high or low on both dimensions: values of 2.5 and below were considered as low, and values of 3.5 and above were considered high. Items with intermediate values were excluded, to ensure a clear contrast between conditions. The placement of these words was carefully controlled during material construction, within predefined areas of interest (AOIs) in the text.

In this experimental design, the *extreme values* refer to words in the highest or lowest percentiles according to specific psycholinguistic properties. High-frequency words are those commonly used in everyday language (e.g., "table", "run") and have high corpus frequencies (e.g., more than 100 occurrences per million words), while low-frequency words are rarely used (e.g., "stinger", "cicada") and have fewer than 10 occurrences per million words. High-imageability words evoke clear mental images (e.g., "apple", "car") and are rated on average ≥ 6 on a scale from 1 to 7, while abstract, low-imageability words (e.g., "truth", "time") are rated ≤ 3 .

The controlled use of these normed, extreme-value words allows for:

1. **Clear differentiation between conditions:** Contrasting high and low values increases sensitivity to the effects of lexical manipulation.
2. **Reduction of data noise:** Words with "average" properties, which could complicate interpretation, are excluded.
3. **Control of confounding variables:** Extreme words often have additional predictable properties (e.g., word length), making it easier to isolate target effects more precisely.

Studies involving children with typical language development have shown that word familiarity and age of acquisition are among the most influential predictors of reading comprehension. However, due to the lack of normative data for these variables in Croatian, word frequency and imageability were selected as valid proxies to explore how children with different language profiles process psycholinguistic properties during reading, and how this in turn affects their comprehension.

It is also important to note that including a larger number of psycholinguistic variables would have increased task complexity and length, potentially placing additional demands on participants. For this reason, the study focused specifically on two psycholinguistic word properties: frequency and imageability, to maintain a theoretically focused and methodologically controlled design.

The four paragraphs within each story varied according to the experimental conditions, each representing combinations of high and low frequency and imageability. The matrix of these conditions is presented in the Table 6.

Table 6. Matrix of Word Frequency and Imageability Combinations and Corresponding Experimental Conditions

		<i>Imageability</i>	
<i>Frequency</i>		High	Low
	High	Condition 1	Condition 2
	Low	Condition 3	Condition 3

Each paragraph in the constructed narratives represented a separate experimental condition, with anaphoras and their antecedents manipulated for the complexity of their (psycho)linguistic properties serving as the target stimuli. This approach was chosen because, although previous studies have often investigated the frequency and imageability as isolated variables, their interaction has not yet been sufficiently explored, especially in clinical populations. By including combinations of these two factors, the present study aimed to deepen our understanding of how different psycholinguistic properties interact during language processing in ecologically valid reading contexts. This design reflects the complexity of natural language processing, where words are not considered in isolation, but as dynamic interactions of their various properties. Such an approach enables a more precise analysis of how readers function in everyday language environments.

In the constructed texts, anaphoras and their antecedents were defined as key areas of interest (AOIs), as they represent central coreferential elements whose processing directly impacts discourse coherence (Figure 6). The antecedent (i.e., the referential noun) and the anaphora (i.e., the linguistic element that refers to the antecedent) were chosen as focal points because establishing a cohesive referential link between them requires the integration of information at the discourse level, a crucial process for examining the effects of psycholinguistic properties on reading comprehension.

Figure 6. Illustration of areas of interest (AOI 1: antecedent; AOI 2: anaphora) using an example sentence from the experimentally constructed narrative. English translation: *Ivana was standing behind the antique wooden clock with an owl on its front. She rang it gently.*

AOI 1
AOI 2

Ivana je stajala iza starinske, drevne ure sa sovom na pročelju. Zvonila je nježno njišući kazaljke.

$n + 2$
 $n + 2$
 $n + 2$
 $n + 2$

To isolate the effects of the target variables - word frequency and imageability - two words preceding and two words following each antecedent or anaphora were also controlled. This

methodological step addressed the so-called *spill-over effect*, in which the cognitive load associated with processing word n may extend to subsequent fixations on words $n+1$ or $n+2$ (Schroyens et al., 1999), potentially confounding the interpretation of effects attributed to the intended target. For example, the low frequency of an antecedent could prolong fixations on subsequent words, thereby masking the true effects on the processing of the anaphora (Rayner & Duffy, 1986; Pollatsek et al., 2006).

The constructed stories (Appendix No. 3) expressed anaphoric relations by first introducing a character (the antecedent) with a full noun phrase, which was then maintained using null anaphora expressed through verb phrase - a structure prototypical in Croatian and consistent with anaphora resolution preferences in the language (Sorace et al., 2009; Hržica & Kuvač Kraljević, 2022). The number, gender, and order of character introduction, the grammatical function of the antecedent, and the position and proximity of the anaphoric element to its referent were carefully checked in all texts.

To assess reading comprehension (product), four questions were constructed for each of the four stories: two literal and two inferential questions referring to the text as a whole. This structure was consistent across all stories. The questions were directly linked to the anaphoric relationships embedded in the narrative.

The two literal questions were presented first, requiring participants to recognize and recall information explicitly stated in the text (e.g., "*Who sang?*"). These questions focused on details such as identifying characters or recognising specific events. For example, participants were asked who performed a particular action or what idea occurred to a character. To answer correctly, they needed to retrieve information directly from the text, without further interpretation.

This was followed by two inferential questions, which required deeper processing and integration of information not explicitly mentioned in the text - such as inferences about feelings or cause-and-effect relationships (e.g., "*Why was the boy angry?*"). Inferential comprehension was assessed through two types of questions. The first type required participants to generate a causal inference, such as explaining why a character behaved in a certain way. In this case, children had to integrate explicit textual information with its implied meaning. The second type targeted anaphoric inference, requiring participants to identify the referent of an expression by linking it to the appropriate element in the preceding context. In

both formats, the task required participants to go beyond surface-level recall and establish meaningful connections across the text.

All questions were open-ended and followed the recommendations of Cain & Oakhill (2006). This format encourages more active and generative cognitive processing, allowing for a deeper assessment of reading comprehension. Open-ended questions prompt participants to express their own interpretations and conclusions, enabling more accurate detection of their inferential reasoning skills, working memory capacity, and potential expressive language difficulties (Cain & Oakhill, 2006; Woolley, 2011). Unlike multiple-choice questions, which provide cues and constrain the range of responses, open-ended formats provide richer data on comprehension processes, especially under more complex linguistic and discourse conditions (Magliano et al., 2007; Graesser et al., 2010).

Moreover, studies have shown that open-ended questions and longer texts in languages with transparent orthographies, are more sensitive to fluctuations in language comprehension. Their use in this study therefore enabled a more comprehensive assessment of reading skills, including self-monitoring of comprehension, which cannot be fully captured by simpler tasks such as multiple-choice formats or short passages (Cain & Oakhill, 2006; Keenan et al., 2008).

As children with dyslexia were included in the sample, decoding skill was a potential confounding variable in the text processing. However, previous research has shown that similar cognitive and language processes support the comprehension of both spoken and written discourse (Bell & Perfetti, 1994; Nation, 2005). To minimise the influence of decoding demands on comprehension performance, comprehension questions were presented orally rather than requiring independent reading.

To ensure the linguistic validity and age-appropriateness of the materials, all constructed narratives and corresponding comprehension questions were reviewed in collaboration with a linguistics expert. This process involved verifying that the texts and questions were linguistically clear and developmentally appropriate for the target age group, and that the syntactic and lexical elements conformed to the norms and stylistic conventions of standard Croatian. This step was essential for ensuring the face validity of the materials in terms of linguistic relevance and pedagogical appropriateness, thus increasing the overall reliability of the research results.

3.5 Experimental procedure in online conditions (eye-tracking)

Procedure and Measures

Eye-tracking technology was used to measure the processes underlying discourse comprehension - specifically, the resolution of anaphoric relations - in real time. Eye-tracking is an objective, non-invasive, naturalistic, and highly controlled experimental method that provides valuable insight into complex, multi-level processes of language comprehension and yields reliable and valid data (Staub & Rayner, 2007). This method tracks how long and how often the eyes remain fixated on the area of interest (e.g. gaze behaviour, including fixation duration on individual words), how frequently they return to previous area of interest (e.g. the number of regressive saccades), and the speed at which the eyes move across the text, etc. It assumes that physiological eye movements are closely linked to cognitive processing.

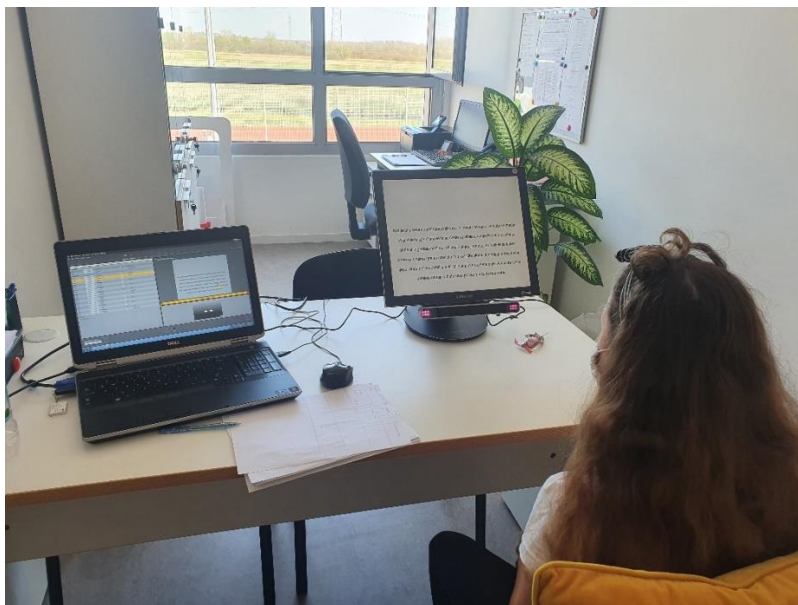


Figure 7. Experimental setup for real-time anaphora processing task using eye-tracking

In the real-time experimental condition, participants read a narrative displayed on a computer screen, paragraph by paragraph, in a self-paced manner without time constraints (Figure 7). Eye movements were recorded throughout natural reading to capture online processing. After reading the entire story, participants responded to comprehension questions, which were presented on the screen, and asked orally by the examiner. Areas of interest (AOIs) were predefined around potential antecedents (and anaphoras) to examine gaze patterns associated with anaphora resolution. Participants were expected to direct their gaze to the pre-defined area of interest (AOI) associated with the anaphora. This response pattern is typically characterised by shorter fixation durations and a higher number of regressive saccades.

Deviations from this expected pattern were interpreted as signs of difficulties in processing anaphoric relations.

Eye movements were recorded with a portable SMI Red-m eye-tracking system (120 Hz), used without a chin rest to preserve a more natural reading setting. A five-point calibration was carried out before the experiment and repeated whenever necessary to ensure data quality. Participants were seated approximately 60 cm from the screen. Stimuli were presented via SMI Experiment Center and subsequently analysed in BeGaze (SensoMotoric Instruments, 2014). Calibration accuracy was checked prior to recording and considered acceptable for further analysis. Gaze behaviour was examined through fixation durations and regressive saccades within predefined areas of interest (AOIs).

The specific eye-tracking measures that served as dependent variables in this experiment, along with their operational definitions, are presented in the Table 7.

Table 7. Overview of Eye-Tracking Measures and Their Function in the Study

<i>Eye-Tracking Measure</i>	<i>Description</i>	<i>Function in the Study</i>	<i>Language Processing Phase</i>
First Fixation Duration	Duration (in milliseconds) of the first fixation within an AOI during the first visit (initial pass), prior to any regressions.	Measure of initial lexical processing; sensitive to word frequency and imageability.	Indicators of early language processing phase: initial lexical activation, word recognition, and the early integration of word meaning into the current context. These measures are particularly relevant for examining the effects of word frequency and imageability, as such lexical properties influence the speed and ease of word recognition during reading.
First Pass Duration	Sum of fixation durations within an AOI from first entry until the gaze exits the AOI for the first time, excluding any subsequent re-entries (regressions).	Reflects immediate lexical processing and sensitivity to psycholinguistic properties.	
Dwell Time (Total Gaze Duration in BeGaze)	Total time spent within an AOI from entry to final exit, including all fixation durations and saccadic movements occurring within the AOI, as well as any regressions back to the AOI.	Indicator of cognitive load and difficulty in processing psycholinguistic (lexical) properties such as frequency and imageability.	Late phase measures - provide insight into subsequent cognitive processing, including reappraisal, reinterpretation, and the maintenance of semantic and referential coherence within a broader discourse. These measures are particularly useful when investigating how children solve more demanding

Number of Fixations	Total number of fixations within the AOI across all visits.	Reflects the processing intensity and potential processing interruptions in relation to key linguistic elements.	comprehension tasks, such as establishing coreferential connections between anaphoras and antecedents.
Number of Revisits	Number of times the gaze re-enters an AOI after having exited it (i.e., number of re-visits/regressions to the AOI).	Reflects processing difficulty and the need for reanalysis; may be associated with challenges in maintaining referential coherence, although revisits are not necessarily restricted to anaphora-driven regressions.	Measures of extended phase of language processing - indicating the need for reanalysis, additional processing and monitoring due to the of reader's uncertainty or a mismatch during reading.
Regression Out Probability	Probability (frequency) that a fixation within an AOI is followed by a regression to a previous region of the text.	Indicates the need to reprocess preceding text due to misunderstanding or lexical complexity.	

The scientific debate about which eye-tracking measures are most sensitive to different levels of language processing is ongoing (Sedivy, 2010). In this study, which aims to investigate children's reading comprehension at both the lexical and discourse levels - a set of measures was selected to capture both early and late phases of language processing (see Table 7 for detailed description; Figure 8 for a schematic representation).

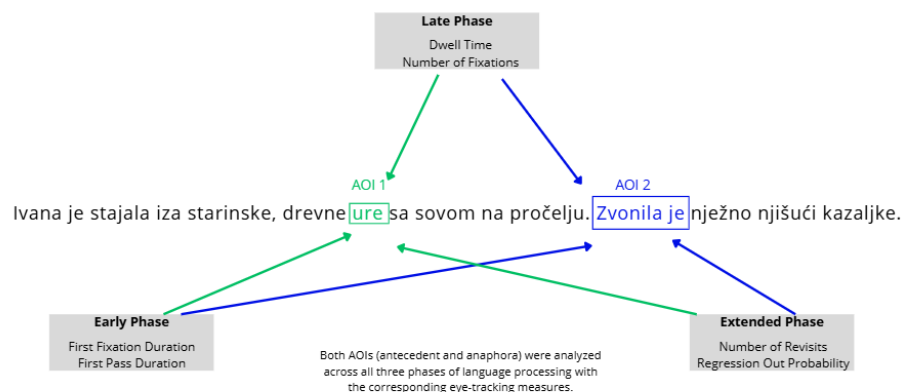


Figure 8. Schematic representation of Areas of Interest (AOI1: antecedent; AOI2: anaphora) and their relation to the three measurement phases in eye-tracking (early, late, extended).

By combining indicators from both early and late processing phases, a more comprehensive understanding of the real-time mechanisms underlying reading comprehension can be achieved. This multidimensional analytic approach is particularly valuable in research involving children with diverse language profiles, as it enables the identification of specific processing difficulties at multiple linguistic levels - from lexical recognition to sentence integration and discourse-level coherence.

The eye-tracking measures described above were analysed within pre-defined areas of interest (AOIs), specifically: the anaphora, the antecedent, and two words immediately before and after each anaphoric element. This design was implemented to account for the so-called *spill-over effect*, where the cognitive load associated with one word can influence the processing of adjacent words. However, for the word regions immediately preceding and following the anaphor or antecedent, only a targeted set of eye-tracking measures - First Fixation Duration, Dwell Time, Number of Fixations, and Number of Revisits - was analysed. These regions do not bear the primary load of referential integration, but they may reveal anticipatory attention before encountering the anaphora or residual difficulty immediately after its processing. The selected measures capture precisely these local fluctuations in language processing without assuming deeper integrative processes. Including the full set of metrics in these peripheral regions would introduce unnecessary complexity and obscure the central findings. Focusing on a smaller, theoretically motivated set of measures ensures a clearer interpretation and preserves analytic coherence across the main and adjacent regions.

In contrast to previous studies which often relied on short, single-line sentences to control for reading direction (Fernández, 2003), the present study used thematically relevant and experimentally balanced narratives (Figure 9). This approach allowed for a more naturalistic reading experience while strictly controlling the experimental conditions, thereby enabling reliable analysis of comprehension processes and referential resolution.

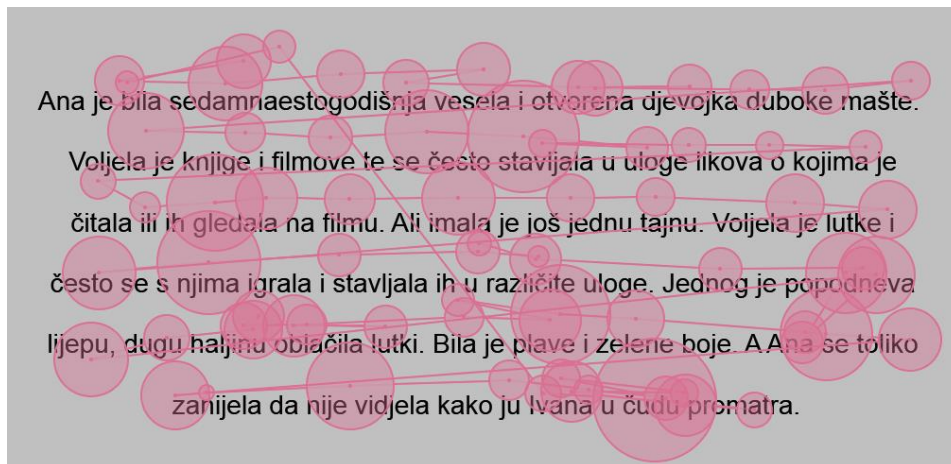


Figure 9. Example of a scanpath recorded during eye-tracking while reading a story paragraph (experimental condition). Circles represent individual fixations, with circle size corresponding to fixation duration, and connecting lines indicating the sequence of saccades across the text.

3.6 Experimental procedure in offline conditions (comprehension accuracy)

Procedure and Measures

After reading all four paragraphs, the participants were shown the entire story on the screen, followed by a comprehension question at the end. They were asked to respond orally to the questions (two literal and two inferential), which were read aloud by the examiner while the texts remained visible to allow reinspection of referential elements. All verbal responses were audio recorded. The complexity of the questions was controlled by both the experimental conditions (i.e., psycholinguistic and anaphoric manipulations) and the question type (literal vs. inferential). The two literal questions were presented first, followed by two inferential questions.

Each story included two literal and two inferential comprehension questions. This structure was consistent across all stories and experimental conditions, resulting in a total of eight literal and eight inferential questions across the four stories. Comprehension questions targeted anaphoric relations between referents and their antecedents across the narrative, rather than focusing on individual paragraphs. This approach was adopted to capture integrative aspects of discourse processing central to anaphora resolution, and to avoid fragmenting comprehension into isolated text segments. To reduce random error associated with individual items and to increase sensitivity to meaningful patterns of comprehension, final scores were computed at the story level rather than at the level of individual questions.

Although responses to individual questions were coded separately to preserve the distinction between literal and inferential comprehension, performance was analysed at the story level, separately for each question type. For each story, participants could obtain a maximum of one point per question type (literal or inferential). A story-level point was awarded if at least one response within a given question type was fully correct (i.e. received the maximum score of 3 on the ordinal scale), regardless of the accuracy of the second response.

This procedure yielded a maximum score of four points per participant for each question type, corresponding to performance across the four stories. To ensure scoring reliability, responses were independently coded by three raters using a four-point ordinal scale (0-3). Intraclass correlation coefficients (ICC) indicated excellent inter-rater reliability for both literal (ICC = 0.91) and inferential (ICC = 0.89) questions (Cicchetti, 1994).

To minimise fatigue effects, short breaks were provided after half of the experimental tasks. As two structurally identical but thematically unrelated versions of each condition were constructed, testing was conducted across sessions (24 hours apart) with the same participants. Each individual test session lasted a maximum of 45 minutes.

4. DATA ANALYSIS

This chapter outlines the analytical procedures used to analyse the data collected in the three experimental phases of the study. All data were analysed using IBM SPSS Statistics 27.

Assumption testing

Before conducting statistical analyses, the assumptions of normality and homogeneity of variances were tested for all relevant dependent variables. Normality was assessed using the Shapiro-Wilk and Kolmogorov-Smirnov tests and by visual inspection of histograms, Q-Q plots, and boxplots. While several variables showed approximately normal distributions, some measures - particularly those representing lexical-semantic ability and error-based results - showed noticeable skewness or kurtosis, with some scores clustering at the lower end of the scale (Appendix No. 1).

The homogeneity of variances between groups was analysed using the Levene's test, which revealed no significant violations that would preclude the use of parametric procedures. In cases where minor deviations were observed, robust analyses and visual inspections supported the use of standard tests, as their robustness to moderate non-normality has been documented in balanced designs (Schmider et al., 2010).

Although not all variables strictly met the normality assumption, these deviations did not affect the interpretability of the classification procedure described in the following section. This is because the classification thresholds were set based on the standardized distributions within the control group (see next section), allowing for consistent application of the thresholds regardless of the individual variable distributions in the entire sample. To address the first two research questions, a mixed design repeated measures ANOVA was performed, while the third research question was examined using regression analysis (see the next chapters).

4.1 Analysis 1: Real-Time Anaphora Processing during reading comprehension

The aim of this experiment was to investigate how children process anaphoric references in real-time under conditions of manipulated psycholinguistic properties (word frequency and imageability), with a particular focus on differences between language profiles.

Eye movement data were conceptually grouped but analysed individually, using multiple gaze-based measures reflecting different stages of language processing:

- **Early processing** of words, which includes initial visual recognition and early semantic integration: First Fixation Duration and First Pass Duration.
- **Late processing**, where the reader makes additional cognitive effort to comprehend, interpret, or connect the word to the preceding context: Dwell Time and Number of Fixations.
- **Extended processing and monitoring**, which involves regressive eye movements and reanalysis of the text prompted by uncertainty or a mismatch with the reader's expectations: Number of Revisits and Regression Out Probability.

Each measure was calculated across four experimental conditions that varied according to the frequency and imageability of the areas of interest (AOI): anaphora, antecedent words, and the two words before and after the AOIs. Raw eye-tracking data were collected across four narrative paragraphs, across four stories, and mean values for each measure and participant were subsequently computed for further analysis. Each eye-tracking measure was submitted to a mixed-design repeated measures ANOVA with two (dependent) within-subjects factors- **Frequency** (high vs. low) and **Imageability** (high vs. low) - and one (independent) between-subjects factor: **Group** (typical development vs. various language disorder profiles). Post-hoc tests (Scheffé and Bonferroni) were performed, and effect sizes (η^2) and power estimates

were reported. Profile plots were created to visualize interactions of condition between groups.

4.2 Analysis 2: Reading Comprehension Accuracy

The second analysis, conducted after the eye-tracking procedure, examined children's comprehension accuracy for literal and inferential questions relating to the same experimental narratives. The aim was to explore discourse-level comprehension and the influence of psycholinguistic properties on anaphoric interpretation in offline conditions. The study focused on the effects of question type (literal vs. inferential) and language-based disorder profiles on comprehension accuracy. Three clinical groups - children with decoding difficulties (poor decoders), children with language comprehension difficulties (poor oral language group), and children with mixed language difficulties (mixed group) - were compared with children with typical development. The aim was to investigate whether different question types have varying effects on reading comprehension success depending on the profile of language-based skills.

The primary analysis consisted of a mixed-design ANOVA 2 x (Question Type subjects: literal, inferential) x 2 (Group: typical language development and language-based disorders), with comprehension performance analyzed as a function of question type (within-subjects factor) and group (between-subjects factor). The second analysis included a mixed-design ANOVA 2 x (Question Type subjects: literal, inferential) x 4 (Group: poor decoders, poor oral language, mixed group, typical development). The dependent variable was comprehension accuracy (range: 0-4). The analysis included 70 participants (N = 37 in the language-based disorders group divided into three clinical subgroups based on previous assessments of phonological and non-phonological language skills: poor oral language group = 10; poor decoders group = 16; mixed group = 11; and N = 33 in the typical language development group).

4.3 Analysis 3: Predictors of Reading Comprehension

The third analysis investigated the predictive contributions of phonological (decoding) and non-phonological (language comprehension) (meta)linguistic composites as well as phonological processing (phonological working memory), primarily based on the Simple View of Reading theoretical model ($D \times C = RC$; Gough & Tunmer, 1986), to overall reading comprehension (sum of literal and inferential scores from Analysis 2), over and above the child's clinical language profile.

Although decoding relies on phonologically based processes, phonological processing was treated as a separate composite predictor. This distinction reflects the difference between foundational phonological mechanisms, which can be assessed independently of print, and decoding performance, which represents the functional outcome of reading-specific skills.

Phonological processing measures (pseudoword repetition and spoonerisms) index core phonological memory and manipulation abilities, whereas decoding measures (reading accuracy and speed) capture the efficiency with which these underlying processes are implemented during reading (Wagner & Torgesen, 1987; Scarborough et al., 2001; Ramus et al., 2013.). Separating these domains therefore allowed for a more precise examination of their distinct contributions to reading comprehension.

Given the theoretical and empirical relationships between the individual variables, composite predictors were constructed by standardising and aggregating the z-scores within the conceptual domains:

- Decoding (phonological (meta)linguistic) skills: reading accuracy and speed.
- Language comprehension (non-phonological (meta)linguistic) skills: receptive vocabulary, grammatical comprehension, vocabulary depth, inferential reasoning.
- Phonological processing: pseudoword repetition and spoonerisms.

These composites were used to reduce measurement error and increase internal consistency, allowing for a clearer interpretation of construct-level effects (Song et al., 2017). A hierarchical multiple regression analysis (enter method) was then conducted to assess the individual and combined contributions of these predictor domains to reading comprehension outcomes.

5. RESULTS

5.1 Analysis 1: real-time anaphora processing during reading comprehension

This section presents the data from real-time processing, in which anaphoric expressions in narrative texts were analysed using the eye-tracking method. The analysis focused on several eye movement measures, each reflecting different phases of language processing (early, late, and extended), depending on the psycholinguistic properties of the target words (frequency and imageability) and participants' language profiles. The results are first presented according to the area of interest at the group level (typical vs. atypical language

development), followed by a more detailed analysis of three subgroups of language-based disorders.

5.1.1 Comparison of results between children with typical language development and those with language-based disorder

Area of interest- Antecedent

Results on early language processing measures

The descriptive data (M and SD) for the region of interest *Antecedent* are presented in Table 8, for *First Fixation Duration* and *First Pass Duration*, which reflect the early phase of language processing (recognition and early lexical access).

Table 8

Descriptive Statistics for Early Language Processing Measures (in ms) TLD & LBD

Measures: First Fixation Duration and First Pass Duration

<i>Condition</i>	<i>Group</i>	<i>First Fixation Duration</i>		<i>First Pass</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	325.88	112.73	417.32	225.19
	LBD	359.99	111.16	476.59	185.32
<i>HILF</i>	TLD	358.70	163.74	445.29	163.98
	LBD	425.47	149.64	609.00	313.47
<i>LIHF</i>	TLD	307.55	142.90	397.02	245.75
	LBD	365.80	143.86	438.00	132.58
<i>LILF</i>	TLD	336.61	154.43	417.45	157.26
	LBD	363.24	161.33	511.91	240.34

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

For both measures, separate mixed-design ANOVAs were conducted with a 2 (imageability: low, high) $\times$ 2 (frequency: low, high) $\times$ 2 (group: TLD, LBD) design, with imageability and frequency as (dependent) within-subject factors and group as a (independent) between-subject factor. The results are shown in Tables 9 and 10.

Table 9

Mixed ANOVA Results for Early Language Processing Measure - First fixation duration in TLD & LBD

First Fixation duration	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	4.949	1/68	.029	.068	.592
Frequency	7.680	1/68	.007	.101	.780
Imageability x Frequency	4.059	1/68	.048	.056	.511
Imageability x group	0.135	1/68	.714	.002	.065
Frequency x group	0.001	1/68	.981	.000	.050
Imageability x Frequency x Group	3.254	1/68	.076	.046	.428

First Fixation Duration: The analysis of the *First Fixation Duration* measure in the antecedent region revealed significant main effects of both imageability and frequency. These results suggest that, regardless of language profile, children fixated antecedents for a shorter duration when the referent words were either more imageable or more frequent.

A significant interaction between imageability and frequency was also found, indicating that the effects of imageability and frequency interact on early processing phase. Post hoc comparison ($p < .005$) shows that the first fixation duration increases in conditions of low frequency, when the word is highly imageable, whereas when the word is highly frequent the level of imageability does not affect first fixation duration. The Figure 10 shows the interaction between these two psycholinguistic properties. The longest duration was observed in high-imageability and low-frequency (HILF) condition. For low-imageability antecedents, changes in frequency had minimal effect. There were no significant interactions with group, no significant main effect of group, and no significant two- or three-way interactions with imageability or frequency. This suggests that the observed effects were consistent for both groups: children with TLD and children with LBD.

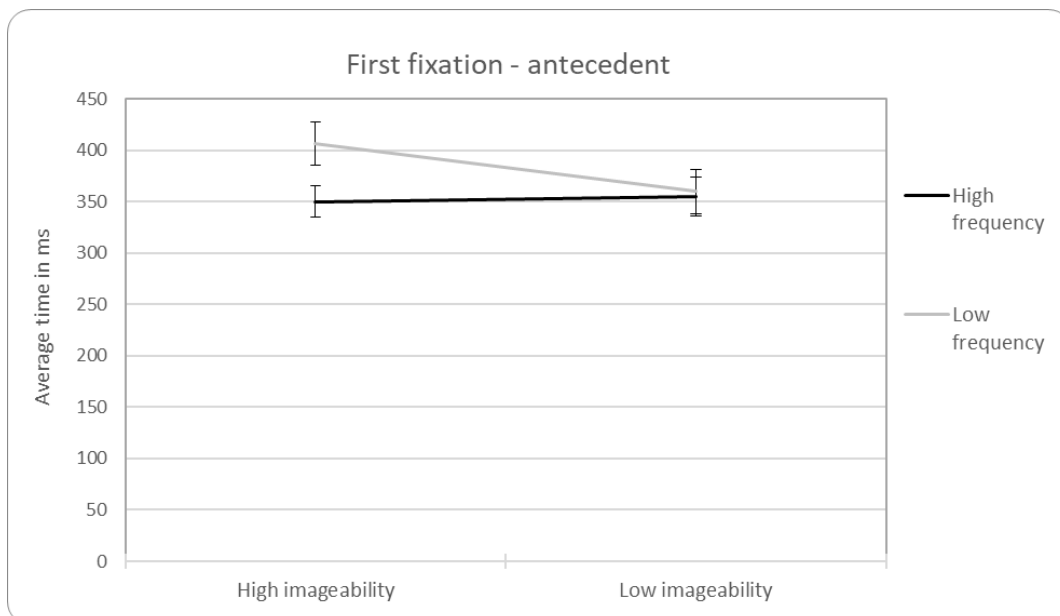


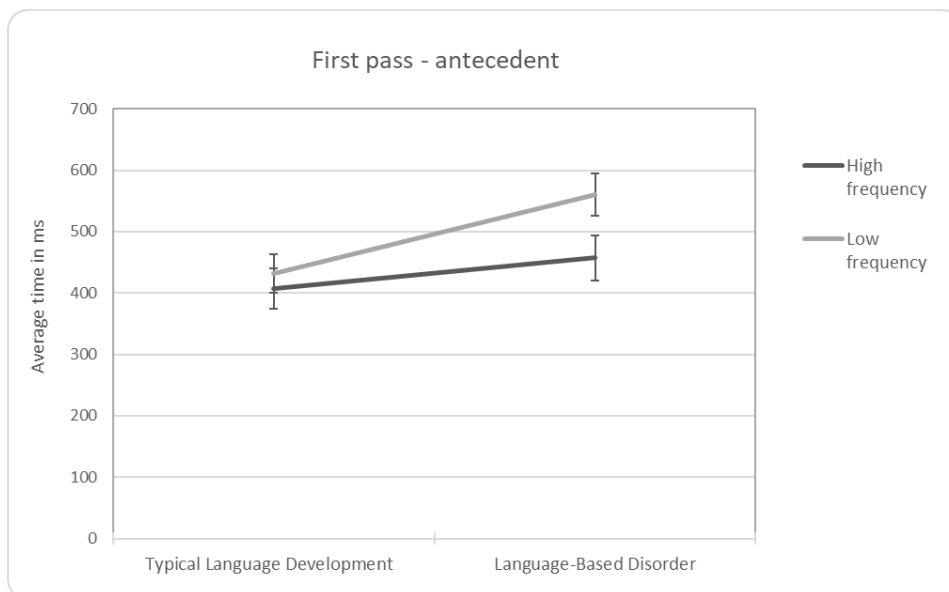
FIGURE 10. The interaction between imageability and frequency on the First Fixation Duration in the Antecedent Region

Note. Error bars represent standard errors of means

Table 10**Mixed ANOVA Results for Early Language Processing Measure - First Pass in TLD & LBD**

First Pass	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	10.832	1/68	.002	.137	.901
Frequency	12.121	1/68	<.001	.151	.929
Imageability x Frequency	1.429	1/68	.236	.021	.218
Imageability x group	2.456	1/68	.122	.035	.339
Frequency x group	4.659	1/68	.034	.064	.567
Imageability x Frequency x Group	0.851	1/68	.360	.012	.149
<i>Independent</i>					
Group	4.107	1/68	.047	.057	.515

First Pass Duration: A significant main effect of imageability and frequency was found, indicating that both properties independently facilitated faster processing of the antecedents. A significant main effect of group was also found - TLD children have shorter first pass duration in the AOI than LBD children. However, a significant interaction between frequency and group was also found. Post hoc comparison shows that low-frequency antecedents have a longer first pass time in the AOI in both groups, with a significant difference between low and high frequency antecedents in the LBD group ($p < .005$). Figure 11 shows the interaction between frequency and group. As shown, the interaction pattern between the groups remained consistent across conditions; however, First Pass values were higher in the LBD group. No other interactions reached statistical significance.

**FIGURE 11.** The interaction between frequency and group on the First Pass Duration in the Antecedent Region

Note. Error bars represent standard errors of means

Results on late language processing measures

The descriptive data (M and SD) for the region of interest *Antecedent* are shown in Table 11, for *Dwell Time* and *Number of Fixations*, which reflect the late phase of language processing. These measures indicate deeper processing and more complex linguistic operations.

Table 11

Descriptive Statistics for Late Language Processing Measures (in ms) for TLD and LBD

Measures: *Dwell Time* and *Number of Fixations*

Condition	Group	Dwell time		Number of fixations	
		M	SD	M	SD
HIHF	TLD	589.41	357.81	10.88	3.83
	LBD	829.84	407.01	15.81	8.18
HILF	TLD	713.41	309.95	14.33	4.77
	LBD	1166.23	782.55	16.78	8.65
LIHF	TLD	563.79	382.92	12.00	4.92
	LBD	856.24	499.32	16.38	7.80
LILF	TLD	749.27	361.34	13.45	5.44
	LBD	1078.08	632.75	17.32	7.85

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data indicates a longer Dwell Times and greater Number of Fixations in children with LBD across conditions, with higher values for low-frequency antecedents.

The results of separate mixed-design ANOVAs (2x2x2) are presented in Tables 12 and 13.

Table 12

Mixed ANOVA Results for Late Language Processing Measure - Dwell Time in TLD & LBD

Dwell Time	F	df	p	η^2	Observed power
<i>Dependent</i>					
Imageability	0.182	1/68	.671	.003	.071
Frequency	51.448	1/68	<.001	.431	1.000
Imageability x Frequency	0.186	1/68	.668	.003	.071
Imageability x group	0.356	1/68	.552	.005	.091
Frequency x group	4.229	1/68	.044	.059	.527
Imageability x Frequency x Group	2.045	1/68	.157	.029	.292
<i>Independent</i>					
Group	9.436	1/68	.003	.122	.857

Dwell Time: Analysis of the *Dwell Time* revealed a significant main effect of frequency, indicating that dwell time was shorter for high-frequency antecedents. There was also a

significant main effect of group, with children with LBD spending more time processing the antecedent region than their TLD peers. Additionally, a significant interaction between frequency and group revealed longer dwell times (total time of all fixations and saccades) for low-frequency antecedents (Figure 12) in both groups, with post hoc comparison showing a more pronounced difference between the processing of high- and low-frequency antecedents in the LBD group ($p < .005$). No significant effects or interactions were found for imageability.



FIGURE 12. The interaction between frequency and group on the Dwell Time in the Antecedent Region

Note. Error bars represent standard errors of means

Table 13

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations in TLD & LBD

Number of Fixations	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.529	1/68	.470	.008	.111
Frequency	12.832	1/68	<.001	.159	.942
Imageability x Frequency	1.007	1/68	.319	.015	.167
Imageability x group	0.217	1/68	.643	.003	.075
Frequency x group	2.461	1/68	.121	.035	.340
Imageability x Frequency x Group	0.954	1/68	.332	.014	.161
<i>Independent</i>					
Group	7.937	1/68	.006	.105	.793

Number of fixations: The results for the *Number of fixations* measure (Table 13) show a significant main effect of frequency, indicating that participants made more fixations on low-

frequency antecedents than on high-frequency antecedents, suggesting greater processing effort when lexical access is less automatic. There was no significant main effect of imageability, nor was there a significant interaction between imageability and frequency.

The main effect of group was statistically significant and showed that children with LBD produced a higher number of fixations than their TLD peers in all conditions. This group difference indicates increased cognitive load and lower processing efficiency in the LBD group during referential resolution.

No significant interactions between psycholinguistic word properties and group were observed.

Results on extended (early and late) language processing measures

The descriptive data (M and SD) for the region of interest *Antecedent* are presented in Table 14, for *Number of Revisits* and *Regression Out Probability*, which reflect both early and late phases of language processing. These measures provide information on regressive eye movements and the frequency of returns to the region of interest, which may indicate comprehension difficulties or attempts to verify referential links. The values are presented by group, imageability, and word frequency.

Table 14

Descriptive Statistics for Extended Language Processing Measures (in ms) TLD & LBD

Measures: Number of Revisits and Regression Out Probability

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>		<i>Regression Out Probability</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	3.27	2.55	0.25	0.08
	LBD	5.27	4.75	0.20	0.10
<i>HILF</i>	TLD	4.67	3.13	0.19	0.08
	LBD	5.27	4.36	0.20	0.13
<i>LIHF</i>	TLD	3.52	2.87	0.27	0.17
	LBD	5.54	4.59	0.20	0.09
<i>LILF</i>	TLD	4.15	2.99	0.23	0.16
	LBD	6.00	4.30	0.18	0.06

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data indicates higher values for the Number of Revisits in the LBD group, across conditions, and lower values of the Regression Out Probability.

The results of separate mixed-design ANOVAs are shown in Tables 15 and 16.

Table 15**Mixed ANOVA Results for Extended Language Processing Measure - Number of Revisits in TLD & LBD**

Number of Revisits	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.318	1/68	.575	.005	.086
Frequency	3.216	1/68	.077	.045	.424
Imageability x Frequency	0.065	1/68	.799	.001	.057
Imageability x group	0.973	1/68	.327	.014	.163
Frequency x group	1.280	1/68	.262	.018	.200
Imageability x Frequency x Group	1.085	1/68	.301	.016	.177
<i>Independent</i>					
Group	4.929	1/68	.030	.068	.590

Number of revisits: The analysis of the *Number of revisits* measure revealed a significant main effect of group, indicating that children with LBD made more revisits to the antecedent region than children with TLD. This suggests that the LBD group placed greater demands on reprocessing and required more time to integrate referential information.

No significant effects of imageability or frequency were observed and there were no significant interactions between psycholinguistic properties or between psycholinguistic properties and group. This indicates that the differences in revisit behavior were not influenced by the psycholinguistic properties of the referent.

Table 16**Mixed ANOVA Results for Extended Language Processing Measure - Regression Out Probability in TLD & LBD**

Regression out probability	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.722	1/68	.398	.011	.133
Frequency	9.805	1/68	.003	.126	.870
Imageability x Frequency	0.249	1/68	.619	.004	.078
Imageability x group	2.512	1/68	.118	.036	.346
Frequency x group	3.603	1/68	.062	.050	.465
Imageability x Frequency x Group	0.952	1/68	.333	.014	.161
<i>Independent</i>					
Group	4.092	1/68	.047	.057	.514

Regression out Probability: The analysis of Regression out probability measure revealed a statistically significant main effect of word frequency, indicating that frequency influenced the likelihood of regressions to the previous AOI. Interestingly, children showed more

regressions when the antecedents were of higher frequency. There was no significant main effect of imageability or any interaction with frequency or group (Table 16). However, a significant group effect was observed - TLD showed greater probability of returning to the previous region after fixation within the antecedent.

Area of interest - Anaphora

Results on early language processing measures

Table 17 shows the descriptive statistics (M and SD) for two measures reflecting early language processing in the domain of anaphoric resolution: First Fixation Duration and First Pass Duration.

Table 17

Descriptive Statistics for Early Language Processing Measures (in ms) TLD & LBD

Measures: First Fixation Duration and First Pass Duration

<i>Condition</i>	<i>Group</i>	<i>First Fixation</i>		<i>First Pass</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	309.08	94.79	477.40	171.38
	LBD	369.05	165.97	530.48	256.14
<i>HILF</i>	TLD	365.65	136.35	678.18	730.89
	LBD	371.54	184.33	704.82	631.13
<i>LIHF</i>	TLD	319.26	90.35	562.09	671.89
	LBD	365.88	152.48	736.15	715.13
<i>LILF</i>	TLD	339.95	129.88	500.02	181.12
	LBD	368.88	186.43	623.08	589.19

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

For both measures, separate mixed-design ANOVAs were conducted with a 2 (imageability: low, high) $\times$ 2 (frequency: low, high) $\times$ 2 (group: TLD, LBD) design, with imageability and frequency as (dependent) within-subject factors and group as a (independent) between-subject factor. The results are shown in Tables 18 and 19.

Table 18

Mixed ANOVA Results for Early Language Processing Measure - First fixation duration in TLD & LBD

First fixation duration	<i>F</i>	<i>Df</i>	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.179	1/68	.673	.003	.070
Frequency	2.529	1/68	.116	.036	.348
Imageability x Frequency	0.540	1/68	.465	.008	.112

Imageability x group	0.037	1/68	.848	.001	.054
Frequency x group	1.902	1/68	.172	.027	.275
Imageability x Frequency x Group	0.572	1/68	.452	.008	.116
<i>Independent</i>					
Group	1.584	1/68	.212	.023	.237

First Fixation Duration: The analysis of *First Fixation Duration* showed that the main effects of imageability and frequency were not statistically significant, nor was their interaction. Additionally, no significant interactions were found between these effects and group, including the three-way interaction among imageability, frequency, and group.

These results suggest that the psycholinguistic properties of the words and group membership did not significantly influence *First Fixation Duration*.

Table 19

Mixed ANOVA Results for Early Language Processing Measure - First Pass in TLD & LBD

First pass	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.015	1/68	.902	.000	.052
Frequency	0.705	1/68	.404	.010	.131
Imageability x Frequency	5.077	1/68	.027	.069	.603
Imageability x group	0.778	1/68	.381	.011	.140
Frequency x group	0.106	1/68	.746	.002	.062
Imageability x Frequency x Group	.010	1/68	.920	.000	.051
<i>Independent</i>					
Group	1.472	1/68	.229	.021	.223

First Pass Duration: The analysis of *First Pass Duration* measure revealed that the main effects of imageability and frequency were not statistically significant. The interaction between imageability and frequency was statistically significant, meaning that both psycholinguistic properties modulate early anaphora processing.

The interaction pattern suggests that participants processed these items most quickly in conditions of aligned psycholinguistic word properties: either when the lexical representation was well established (high frequency, high imageability) or when the word was not elaborated upon due to limited lexical representation (low frequency, low imageability). The longest first-pass durations seems to emerge when the two psycholinguistic properties were mismatched - high frequency combined with low imageability, or low frequency combined with high imageability - indicating that additional processing was required when only one

lexical cue supported rapid activation of the word. However, post hoc comparisons did not show any statistically significant differences between specific condition pairs (all $p > .05$, corrected), indicating that the nature of this interaction is not clear (Figure 13).

No significant interactions were found between the main effects and the participant group, and the three-way interaction was also not statistically significant.

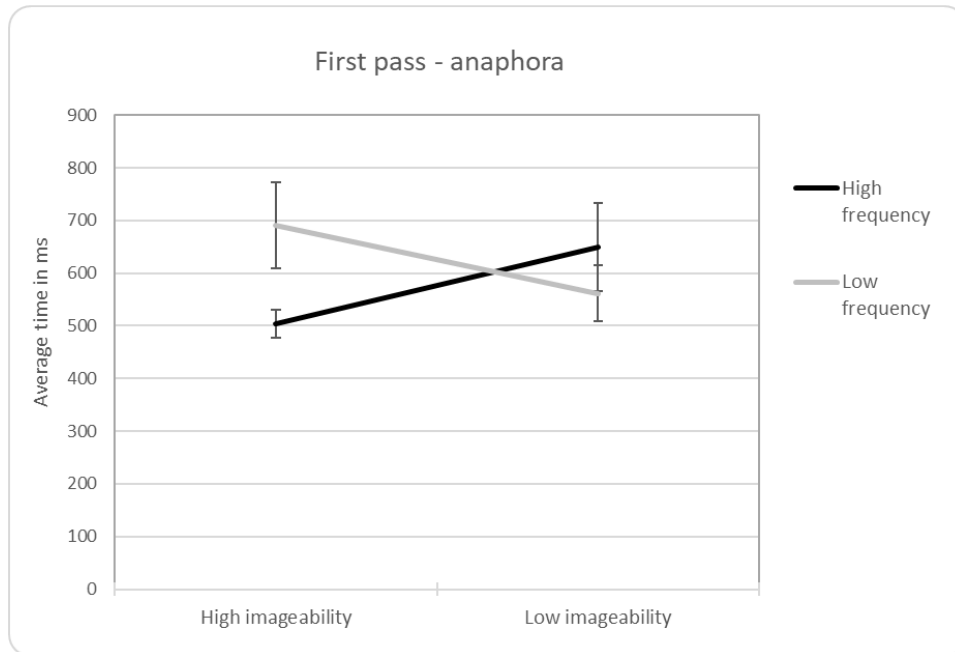


Figure 13. The interaction between imageability and frequency on the First Pass Duration in the Anaphora Region

Note. Error bars represent standard errors of means

Results on late language processing measures

Table 20 shows the descriptive statistics (M and SD) for two measures reflecting late language processing in the domain of anaphoric resolution: Dwell Time and Number of Fixations. The values are presented by group, imageability, and word frequency.

Table 20

Descriptive Statistics for Late Language Processing Measures

Measures: Dwell Time and Number of Fixations

<i>Condition</i>	<i>Group</i>	<i>Dwell time</i>		<i>Number of fixations</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	830.14	377.76	19.70	5.81
	LBD	1136.71	500.75	23.19	8.77
<i>HILF</i>	TLD	1121.13	684.81	21.82	9.23
	LBD	1566.66	958.26	27.73	16.23

<i>LIHF</i>	TLD	893.69	431.48	19.67	7.79
	LBD	1266.75	804.61	23.38	12.58
<i>LILF</i>	TLD	1073.50	649.13	21.82	9.13
	LBD	1414.41	910.18	26.46	19.64

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data indicates higher values of Dwell Time and Number of Fixations in low-frequency conditions. The results of separate mixed-design ANOVAs are shown in Tables 21 and 22.

Table 21

Mixed ANOVA Results for Late Language Processing Measure- Dwell Time in TLD & LBD

Dwell Time	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.002	1/68	.968	.000	.050
Frequency	28.758	1/68	.000	.297	1.000
Imageability x Frequency	6.001	1/68	.017	.081	.675
Imageability x group	0.058	1/68	.810	.001	.057
Frequency x group	0.298	1/68	.587	.004	.084
Imageability x Frequency x Group	1.135	1/68	.290	.016	.183
<i>Independent</i>					
Group	5.910	1/68	.018	.080	.669

Dwell Time: The results for the *Dwell Time* measure showed a significant main effect of word frequency, indicating that processing time was longer for low-frequency words, while the main effect of imageability was not statistically significant. A significant interaction was found between imageability and frequency. Post hoc comparison showed that differences between low and high frequency are significant both in high and low imageability ($p < .005$). However, in low frequency and in high frequency conditions, the differences between imageability were not significant. Figure 14 suggests that the interaction is potentially driven by a smaller difference between high and low frequency in the low imageability condition.

A significant group effect was also observed, indicating that children with LBD had longer dwell times than TLD children, reflecting slower processing of anaphora compared to TLD.

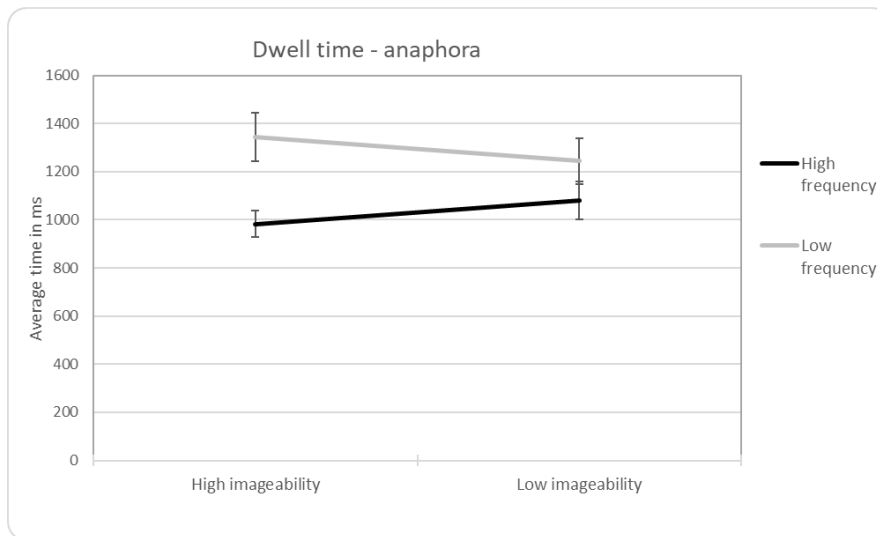


FIGURE 14. The interaction between imageability and frequency on Dwell Time in the Anaphora Region

Note. Error bars represent standard errors of means

Table 22

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations in TLD & LBD

Number of fixations	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.102	1/68	.750	.002	.061
Frequency	10.723	1/68	.002	.136	.898
Imageability x Frequency	0.208	1/68	.650	.003	.073
Imageability x group	0.092	1/68	.763	.001	.060
Frequency x group	0.850	1/68	.360	.012	.149
Imageability x Frequency x Group	0.226	1/68	.636	.993	.075
<i>Independent</i>					
Group	3.133	1/68	.081	.044	.415

Number of Fixations: The analysis of *Number of fixations* revealed a statistically significant main effect of frequency, with a higher number of fixations observed for low-frequency words. The main effect of imageability and the interaction between frequency and imageability, were not statistically significant.

Results on extended (early and late) language processing measures

Table 23 presents the descriptive statistics for two measures that, according to the literature, fall within the extended phase of language processing and include both early and late processing components: *Number of Revisits* and *Regression Out Probability*.

Table 23**Descriptive Statistics for Extended Language Processing Measures***Measures: Number of Revisits and Regression out probability*

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>		<i>Regression Out Probability</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	6.39	3.19	.15	.05
	LBD	7.89	4.64	.15	.06
<i>HILF</i>	TLD	6.97	4.86	.14	.06
	LBD	9.35	7.56	.12	.05
<i>LIHF</i>	TLD	5.88	3.78	.15	.05
	LBD	6.76	5.07	.14	.07
<i>LILF</i>	TLD	6.70	4.55	.14	.06
	LBD	8.86	8.89	.13	.06

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of separate mixed-design ANOVAs are presented in Tables 24 and 25.

Table 24
Mixed ANOVA Results for Early and Late Language Processing Measures - Number of revisits in TLD and LBD

Number of Revisits	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	2.217	1/68	.141	.032	.312
Frequency	7.320	1/68	.009	.097	.760
Imageability x Frequency	0.262	1/68	.611	.004	.080
Imageability x group	0.265	1/68	.608	.004	.080
Frequency x group	1.405	1/68	.240	.020	.215
Imageability x Frequency x Group	0.054	1/68	.816	.001	.056
<i>Independent</i>					
Group	2.317	1/68	.133	.033	.323

Number of revisits: The results for the *Number of Revisits* measure showed a significant main effect of word frequency, indicating that children made more regressions to the area of interest when processing low-frequency words. In contrast, the main effect of imageability was not statistically significant. The interaction between imageability and frequency was also not significant, suggesting that these two psycholinguistic properties do not interactively affect the number of regressions.

No significant interaction was found between either psycholinguistic properties and group, and the main effect of group was also not significant, suggesting that the overall number of revisits did not differ significantly between children with LBD and children with TLD. However, the descriptive statistics suggest a trend towards more revisits in the LBD group across all conditions, particularly for low-frequency items, but this was not statistically confirmed.

Table 25

Mixed ANOVA Results for Early and Late Language Processing Measures - Regression Out Probability in TLD and LBD

Regression Out Probability	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.13	1/68	.722	.002	.064
Frequency	4.85	1/68	.031	.067	.584
Imageability x Frequency	0.62	1/68	.436	.009	.121
Imageability x Group	0.01	1/68	.906	.000	.052
Frequency x Group	1.20	1/68	.277	.017	.191
Imageability x Frequency x Group	0.56	1/68	.456	.008	.114
<i>Independent</i>					
Group	0.79	1/68	.378	.011	.141

Regression out probability: The results for the *Regression Out Probability* measure revealed a significant main effect of word frequency, suggesting that children were more likely to make regressions out of the area of interest when processing high-frequency anaphora than low-frequency anaphora. The main effect of imageability was not statistically significant.

No significant interaction was found between imageability and frequency, and there were no significant interactions with group. The main effect of group was also not significant, suggesting that the probability of making regressions out of the target region did not differ significantly between children with LBD and children with TLD.

Area of interest- Words preceding AOIs

For the analysis of the word regions immediately preceding and following the anaphora or antecedent, four eye-tracking measures were selected: First Fixation Duration, Dwell Time, Number of Fixations, and Number of Revisits. These measures provide insight into basic patterns of attention and processing in the local context, without assuming that these words play a central role in the integration of referential information. The chosen measures - particularly Dwell Time and Number of Fixations - capture the extent to which attention is

sustained in the local region, which is sufficient to detect potential anticipatory or residual effects in anaphor processing. Including all available measures in regions that are not the primary focus of the study could result in an excessive data and reduce the clarity of the findings. Limiting the analysis to a smaller set of measures ensures a more focused examination and a clearer interpretation of results.

First Fixation Duration: Table 26 presents the descriptive statistics (M and SD) for First Fixation Duration measure.

Table 26

Descriptive Statistics for First Fixation Duration (in ms) TLD & LBD

<i>Condition</i>	<i>Group</i>	<i>First Fixation</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	312.90	71.43
	LBD	345.56	104.89
<i>HILF</i>	TLD	344.50	82.22
	LBD	344.51	111.29
<i>LIHF</i>	TLD	317.40	84.32
	LBD	336.63	109.00
<i>LILF</i>	TLD	329.43	103.30
	LBD	352.20	111.04

Note. TLD = Typical Language Development; LD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of mixed-design ANOVA are shown in Table 27.

Table 27

Mixed ANOVA Results for Early Language Processing Measure - First fixation duration before/ after AOI in TLD & LBD

First fixation duration	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>	0.400	1/68	.529	.006	.096
Imageability	10.177	1/68	.002	.130	.882
Frequency	0.024	1/68	.877	.000	.053
Imageability x Frequency	0.249	1/68	.619	.004	.078
Imageability x group	2.552	1/68	.115	.036	.350
Frequency x group	0.400	1/68	.529	.006	.096

Imageability x Frequency x Group	3.632	1/68	.061	.051	.468
<i>Independent</i>					
Group	0.703	1/68	.405	.010	.131

The results revealed a significant main effect of imageability, with low-imageability words generally eliciting a longer first fixation duration than high-imageability words. No significant main effect was found for frequency or for the group variable.

Although the descriptive means appeared to differ somewhat between the TLD and LBD groups, these differences did not reach statistical significance. The findings instead suggest that, at the level of first fixation duration, imageability exerted a general effect across participants, with low-imageability words requiring longer initial processing than high-imageability words.

No other two- or three-way interactions reached significance, including the interaction between imageability and frequency, or the three-way interaction with group.

Dwell time: Table 28 presents the descriptive statistics (M and SD) for measure the *Dwell Time*. The values are presented by group, imageability, and word frequency.

Table 28

Descriptive Statistics for Dwell Time

<i>Condition</i>	<i>Group</i>	<i>Dwell Time</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	1088.78	523.27
	LBD	1522.33	765.76
<i>HILF</i>	TLD	1516.56	741.81
	LBD	2120.99	1063.26
<i>LIHF</i>	TLD	1256.52	567.74
	LBD	1770.29	996.81
<i>LILF</i>	TLD	1393.26	812.23
	LBD	1898.86	1019.47

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data suggests higher values of Dwell Time in the LBD group, and in the conditions with low-frequency words.

The results of mixed-design ANOVA are shown in Table 29.

Table 29

Mixed ANOVA Results for Late Language Processing Measure - Dwell Time before/ after AOI in TLD & LBD

Dwell Time	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.350	1/68	.556	.005	.090
Frequency	42.255	1/68	<.001	.383	1.000
Imageability x Frequency	33.271	1/68	<.001	.329	1.000
Imageability x group	0.025	1/68	.876	.000	.053
Frequency x group	0.670	1/68	.416	.010	.127
Imageability x Frequency x Group	1.841	1/68	.179	.026	.267
<i>Independent</i>					
Group	7.309	1/68	.009	.097	.760

The analysis revealed a significant main effect of frequency, with low-frequency words eliciting longer dwell times than high-frequency words in both groups. A significant interaction between imageability and frequency was also found (Figure 15). Post hoc comparisons showed that the effect of frequency was present only for highly imageable words, with low-frequency words eliciting significantly longer dwell times than high-frequency words ($p < .001$), whereas no such difference was observed for low-imageability words. In addition, the effect of imageability reversed depending on frequency: under low-frequency conditions, highly imageable words produced longer dwell times than low-imageability words ($p = .002$), whereas under high-frequency conditions, low-imageability words elicited longer dwell times than highly imageable words ($p < .001$).

A significant main effect of group was found, with LBD children showing longer Dwell times, overall. The three-way interaction (imageability $\times$ frequency $\times$ group) was not statistically significant.

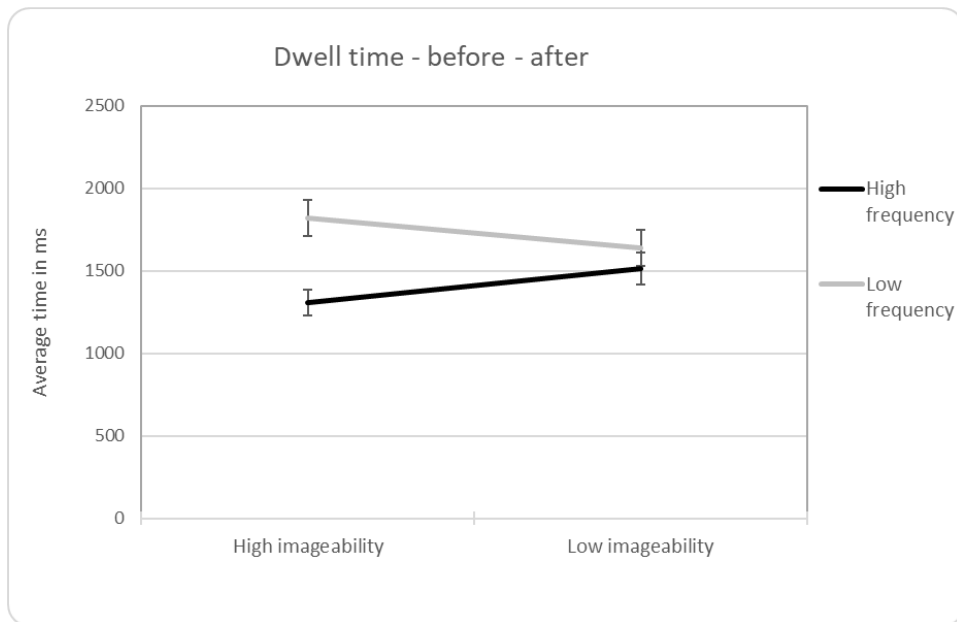


FIGURE 15. The interaction between imageability and frequency on Dwell Time in the Words preceding AOIs

Note. Error bars represent standard errors of means

Number of Fixations: Table 30 presents the descriptive statistics (M and SD) for the measure Number of fixations. The values are presented by group, imageability, and word frequency.

Table 30

Descriptive Statistics for Number of fixations

<i>Condition</i>	<i>Group</i>	<i>Number of fixations</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	71.70	16.50
	LBD	92.86	45.53
<i>HILF</i>	TLD	92.67	27.87
	LBD	114.81	60.05
<i>LIHF</i>	TLD	84.52	21.38
	LBD	104.76	50.86
<i>LILF</i>	TLD	87.15	27.65
	LBD	106.35	55.83

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data suggests that overall, children with LBD made more fixations in the local context surrounding the AOI, with some lower values in high-frequency conditions.

The results of the mixed-design ANOVA are shown in Table 31.

Table 31

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations before/ after AOI in TLD & LBD

Number of fixations	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	3.790	1/68	.056	.053	.484
Frequency	29.403	1/68	<.001	.302	1.000
Imageability x Frequency	39.122	1/68	<.001	.365	1.000
Imageability x group	0.493	1/68	.485	.007	.106
Frequency x group	0.000	1/68	.994	.000	.050
Imageability x Frequency x Group	0.106	1/68	.745	.002	.062
<i>Independent</i>					
Group	4.619	1/68	.035	.064	.563

The results showed a significant main effect of frequency, with low-frequency words triggering more fixations than high-frequency words across groups. Additionally, there was a significant interaction between imageability and frequency (Figure 16). Post hoc comparison shows that low-frequency words led to a higher number of fixations than high-frequency words, but only in the high-imageability condition ($p < .005$). In contrast, no difference between low- and high-frequency words was observed for low-imageability conditions.

There was no significant main effect of imageability alone, although the trend approached significance. Importantly, the group $\times$ frequency and group $\times$ imageability interactions were not significant, indicating that the effects of these psycholinguistic properties were relatively consistent across TLD and LBD groups. However, a significant main effect of group was found, showing that children with LBD made more fixations overall particularly with low-frequency words, reflecting greater cognitive effort.

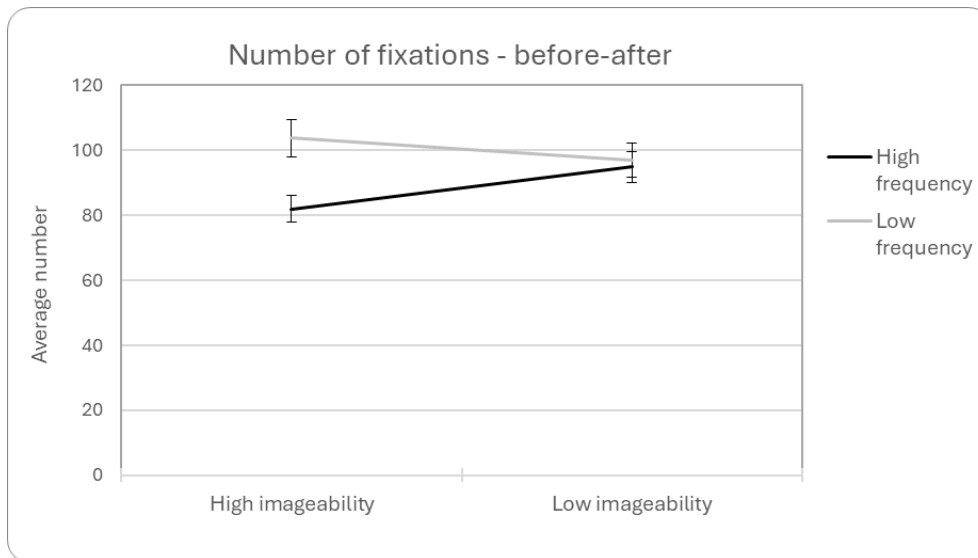


FIGURE 16. The interaction between imageability and frequency on Number of Fixations in the Words preceding AOIs

Note. Error bars represent standard errors of means

Number of Revisits: Table 32 presents the descriptive statistics for the *Number of Revisits* measure. The values are presented by group, imageability, and word frequency.

Table 32

Descriptive Statistics for Number of Revisits

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	19.24	7.90
	LBD	26.78	17.01
<i>HILF</i>	TLD	27.76	13.06
	LBD	34.08	23.31
<i>LIHF</i>	TLD	25.18	11.56
	LBD	31.43	21.77
<i>LILF</i>	TLD	25.48	12.48
	LBD	32.92	21.25

Note. TLD = Typical Language Development; LBD = Language-Based Disorder; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data suggests that the lowest number of revisits was observed in conditions with high imageability and high frequency words, while the highest values were observed in conditions with low frequency words.

The results of mixed-design ANOVA are shown in Table 33.

Table 33

Mixed ANOVA Results for Extended Language Processing Measure - Number of Revisits before/ after AOI in TLD & LBD

Number of revisits	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	4.378	1/68	.040	.060	.541
Frequency	26.779	1/68	<.001	.283	.999
Imageability x Frequency	13.282	1/68	.001	.163	.949
Imageability x group	0.003	1/68	.958	.000	.050
Frequency x group	0.000	1/68	.992	.000	.050
Imageability x Frequency x Group	0.389	1/68	.535	.006	.094
<i>Independent</i>					
Group	3.275	1/68	.075	.046	.430

The results showed a significant main effect of frequency, with low-frequency words triggering more regressions (revisits) than high-frequency words. There was also a significant main effect of imageability, suggesting that low-imageability words increase rereading behaviour. However, a significant interaction between imageability and frequency was observed (Figure 17). Post hoc comparison shows that the effect of frequency was present only for high imageability words ($p < .005$), with low-frequency conditions producing higher values than high-frequency conditions, whereas no meaningful difference was observed for low-imageability words.

No significant interactions were found between group and any of the within-subject factors, and the main effect of group was only marginally significant. This suggests a tendency towards more revisits in children with LBD, although the group differences were not consistent.

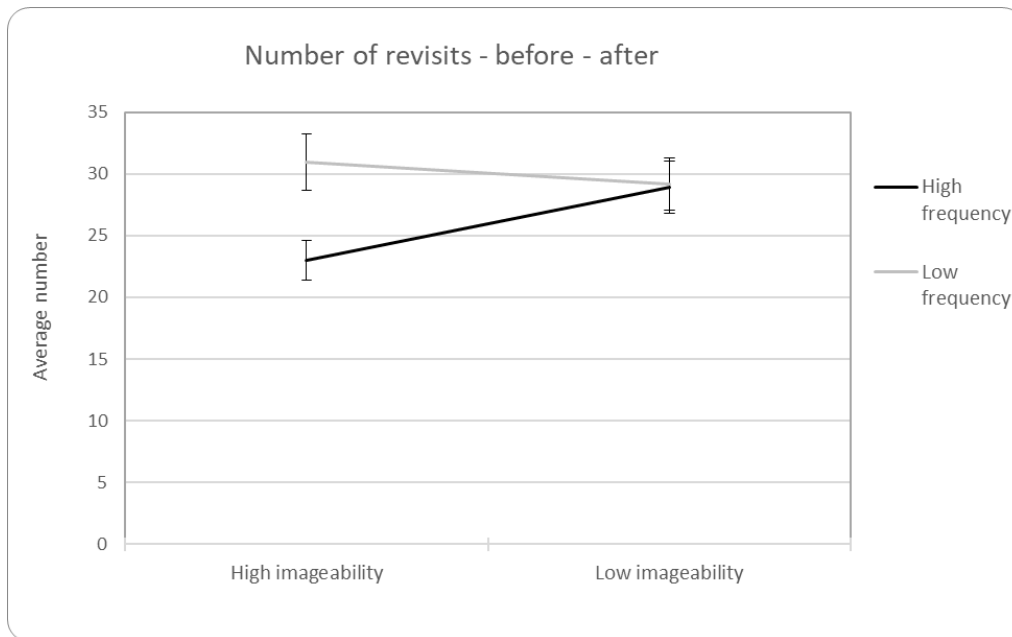


FIGURE 17. The interaction between imageability and frequency on Number of Revisits in the Words preceding AOIs

Note. Error bars represent standard errors of means

5.1.2 Summary of results for children with typical language development and those with language-based disorder across AOIs

The results indicate that word frequency is the primary factor in antecedent processing across all phases of processing. Imageability, on the other hand, shows time-limited and conditional contribution (due to frequency), present only in the early processing phases.

In the antecedent AOI, on the First fixation duration measure, high imageability prolongs initial lexical access, but only in conditions with low frequency antecedents, a pattern consistent for both groups (TLD and LBD). Group differences are found already in the early processing phase, on First Pass measure, where children with LBD process low frequency antecedents longer. These differences intensify in later phases through longer dwell time and higher number of fixations. In extended processing phase, patterns are qualitatively different: children with LBD make more revisits to the AOI, reflecting compensatory re-analysis, while children with TLD show greater probability of regressions out of the AOI, indicating more strategic and flexible processing of referential relationships.

For anaphors, the most evident finding is the absence of effects in early phase- on the First fixation duration there was no single significant effect found. In the First Pass, a significant interaction between imageability and frequency was found, which potentially indicates faster processing when psycholinguistic properties are aligned, however, statistical significance

remains absent. In later processing phases, frequency is confirmed as a reliable factor- low frequency anaphors require longer dwell time and higher number of fixations, while children with LBD show difference compared to TLD children only in dwell time. In extended phase both groups show similar pattern: more regressions for low frequency anaphors, and higher probability of regressions out of the AOI for high frequency anaphors.

In words preceding AOIs, imageability assumes a more prominent role when compared to other target AOIs: in early processing phase low imageability words elicit longer first fixations, regardless of frequency and group membership, thus reflecting difficulty activating mental imagery of the word representation. In later and extended phases, the interaction between frequency and imageability dominates and is consistent through all measures- frequency effect appears exclusively for high imageability words, while low frequency and high imageability words elicit the highest processing effort. Low imageability words remain demanding regardless of frequency. Group differences seems to be present on a global level- children with LBD generally process all conditions more slowly and intensively, without any specific interactions with psycholinguistic properties, which suggests a diffuse reduction in processing efficiency rather than a condition-specific deficit.

5.1.3 Comparison of results between children with TLD and specific profiles of LBD

Area of interest - Antecedent

Results on early language processing measures

Table 34 shows the descriptive statistics for *First Fixation Duration* and *First Pass Duration*. Values are shown for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF) and for each subgroup of children: those with TLD, and those with different profiles of LBD (Poor Oral Language Group - POL, Poor Decoders - PD, Mixed Language difficulties Group - MD).

Table 34

Descriptive Statistics for Early Language Processing Measures (First Fixation Duration and First Pass Duration) by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>First Fixation Duration</i>		<i>First Pass</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	325.88	112.73	417.32	225.19
	POL	339.43	66.80	443.12	117.10
	PD	369.64	134.85	499.54	246.85
	MD	364.63	112.24	473.65	133.86
<i>HILF</i>	TLD	358.70	163.74	445.29	163.98

	POL	396.11	85.76	507.03	145.95
	PD	438.92	186.55	617.46	293.71
	MD	432.60	143.48	689.41	433.06
<i>LIHF</i>	TLD	307.55	142.90	397.02	245.75
	POL	380.36	129.14	420.68	88.90
	PD	336.98	137.41	450.99	139.19
	MD	394.49	169.49	434.84	162.85
<i>LILF</i>	TLD	336.61	154.43	417.45	157.26
	POL	358.80	123.32	418.92	110.30
	PD	335.30	147.02	542.28	255.48
	MD	407.93	210.06	552.28	295.27

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; LIHF= High Imageability - High Frequency; LILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The descriptive data suggests that the values of First Fixation Duration and First Pass are the highest in high imageability and low frequency condition.

Separate mixed-design ANOVAs were conducted for both measures using a 2 (imageability: low, high) $\times$ 2 (frequency: low, high) $\times$ 4 (group: TLD, POL, PD, MD) design, with imageability and frequency as within-subject factors and group as a between-subjects factor. The results are shown in Tables 35 and 36.

Table 35

Mixed ANOVA Results for Early Language Processing Measure - First fixation duration in TLD & LBD subgroups

First Fixation duration	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	3.172	1/66	.080	.046	.419
Frequency	5.881	1/66	.018	.082	.666
Imageability x Frequency	6.662	1/66	.012	.092	.720
Imageability x group	1.961	3/66	.128	.082	.484
Frequency x group	0.109	3/66	.955	.005	.069
Imageability x Frequency x Group	1.102	3/66	.355	.048	.284
<i>Independent</i>					
Group	0.973	3/66	.411	.042	.254

First Fixation Duration: The analysis revealed a significant main effect of frequency, indicating that high-frequency words elicited shorter fixation durations, reflecting more efficient and automatic lexical processing. The main effect of imageability was not statistically significant.

Importantly, a significant interaction was observed between imageability and frequency (Figure 18), suggesting that the combined effect of these two psycholinguistic word

properties was stronger than their individual contributions. This was consistent with the pattern found in the two-group analysis. Specifically, post hoc comparison shows that the effect of frequency was found for high imageability antecedents, with low-frequency words eliciting longer processing times than high-frequency words ($p < .005$), whereas this difference was minimal for low-imageability words. As shown in Figure 18, the combination of high-imageability and low-frequency resulted in the longest fixation durations.

None of the interactions with the group reached statistical significance. The main effect of the group was also not significant.

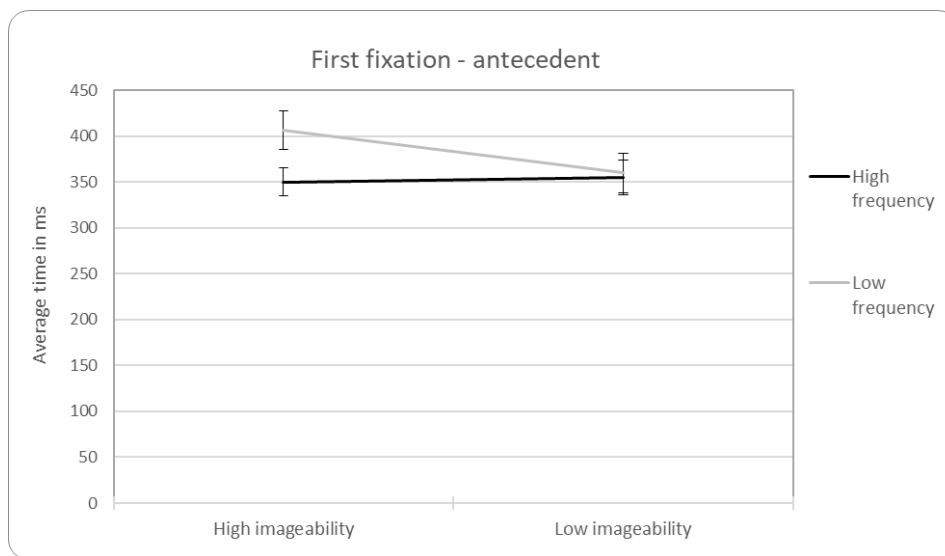


FIGURE 18. The interaction between frequency and imageability on First Fixation Duration in the Antecedent Region

Table 36

Mixed ANOVA Results for Early Language Processing Measure - First Pass Duration in TLD & LD subgroups

First Pass	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	13.302	1/66	.001	.168	.949
Frequency	16.629	1/66	<.001	.201	.980
Imageability x Frequency	2.548	1/66	.115	.037	.350
Imageability x group	0.958	3/66	.418	.042	.251
Frequency x group	3.025	3/66	.036	.121	.687
Imageability x Frequency x Group	0.488	3/66	.692	.022	.144
<i>Independent</i>					
Group	1.868	3/66	.144	.078	.463

First Pass: The analysis of the *First Pass* measure revealed significant main effects for both imageability and frequency. Words with high-imageability were associated with shorter first-pass durations than words with low-imageability, as well as words with high frequency.

The interaction between imageability and frequency was not statistically significant.

Importantly, a significant interaction between frequency and group was observed (Figure 19), the interaction probably arises from differences in sensitivity to word frequency within the groups. However, post hoc comparison indicates that low-frequency words elicited significantly longer first pass than high-frequency words, but only in the poor decoders and mixed group ($p < .005$).

In contrast, the imageability $\times$ group interaction was not significant, implying that all groups benefited similarly from more imageable words.

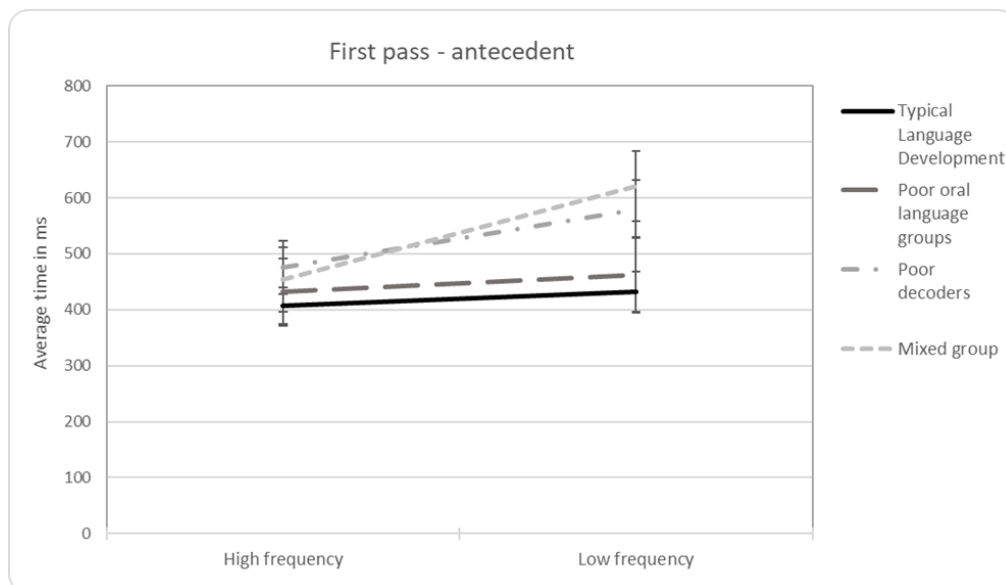


FIGURE 19. The Interaction of frequency and group on First Pass in the Antecedent Region

Results on late language processing measures

Table 37 presents the descriptive data for the measures *Dwell Time* and *Number of Fixations*, analysed in children with TLD and those with different profiles of LBD, in relation to the manipulated psycholinguistic properties of the target word - imageability and frequency.

Table 37

Descriptive Statistics for Dwell Time and Number of Fixations on the Antecedent by Condition and Subgroup

Condition	Group	Dwell Time		Number of fixations	
		M	SD	M	SD

<i>HIHF</i>	TLD	589.41	357.81	10.88	3.83
	POL	583.91	179.90	11.70	3.27
	PD	931.17	413.54	17.25	7.24
	MD	906.03	476.87	17.45	11.38
<i>HILF</i>	TLD	713.41	309.95	14.33	4.77
	POL	709.11	250.79	11.70	3.56
	PD	1260.19	716.07	20.00	10.04
	MD	1445.11	1034.51	16.73	8.03
<i>LIHF</i>	TLD	563.79	382.92	12.00	4.92
	POL	674.20	226.22	12.40	4.38
	PD	889.82	550.81	17.38	8.31
	MD	972.90	588.64	18.55	8.71
<i>LILF</i>	TLD	749.27	361.34	13.45	5.44
	POL	771.24	314.78	14.50	5.46
	PD	1224.15	532.42	19.69	9.53
	MD	1144.57	890.95	16.45	6.41

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of separate mixed-design ANOVAs are shown in Table 38 and 39.

Table 38

Mixed ANOVA Results for Late Language Processing Measure - Dwell Time in TLD & LBD subgroups

Dwell Time	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.311	1/66	.579	.005	.085
Frequency	53.491	1/66	<.001	.448	1.000
Imageability x Frequency	1.483	1/66	.228	.022	.224
Imageability x group	1.174	3/66	.326	.051	.302
Frequency x group	3.678	3/66	.016	.143	.779
Imageability x Frequency x Group	2.008	3/66	.121	.084	.494
<i>Independent</i>					
Group	5.544	3/66	.002	.201	.929

Dwell Time: The analysis of the Dwell Time measure revealed a strong and statistically significant main effect of frequency on dwell time, main effect of group, and frequency x group interaction (Figure 20). The frequency $\times$ group interaction was statistically significant, indicating that high frequency words were processed more efficiently with shorter dwell times, except in the POL group. Post hoc comparison shows that all groups benefited from word frequency- they required less time for all fixations and saccades for high-frequency compared to low-frequency antecedents, except the POL group (for other three groups $p < 0.005$).

The main effect of group was also significant, indicating that overall dwell time differed significantly between groups. Children with TLD processed referents faster than children from MD ($p = .018$) and PD ($p = .012$) groups indicating greater processing effort. Children with POL did not differ significantly from the other three groups ($p > .05$).

No significant effects were observed for imageability alone, imageability $\times$ group or for the imageability $\times$ frequency interactions, suggesting that the effect of imageability was consistently weak and did not interact meaningfully with other factors in this measure.

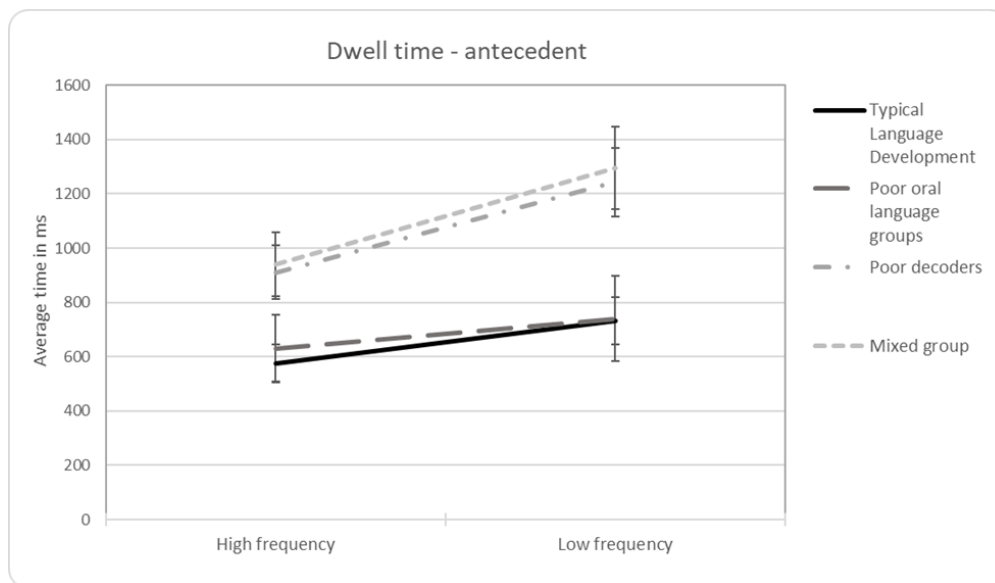


FIGURE 20. The interaction between frequency and group on Dwell Time in the Antecedent Region

Table 39

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations in TLD & LBD subgroups

Number of Fixations	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	1.111	1/66	.296	.017	.180
Frequency	5.102	1/66	.027	.072	.605
Imageability x Frequency	0.141	1/66	.708	.002	.066
Imageability x group	0.539	3/66	.657	.024	.155
Frequency x group	3.163	3/66	.030	.126	.708
Imageability x Frequency x Group	0.624	3/66	.602	.028	.174
<i>Independent</i>					
Group	5.320	3/66	.002	.195	.918

Number of fixations: The analysis for the *Number of fixation* measure revealed a significant main effect of frequency, main effect of group, and the interaction between frequency and

group. Low-frequency words triggered a higher number of fixations on the antecedent compared to high-frequency words. The main effect of group was also significant, indicating that children with TLD and POL made fewer fixations in the antecedent region overall compared to PD (p TLD vs PD = .005; p POL vs PD = .057). The number of fixations in MD is similar to the overall number of fixations made by TLD and POL.

A significant frequency $\times$ group interaction was observed (Figure 21). Post hoc comparisons indicated that only the TLD group showed a reliable effect of frequency (p = .014), with low-frequency words eliciting a higher number of fixations than high-frequency words. This pattern was not observed in the language-based disorder groups (all p > .05), indicating that only the TLD group showed the sensitivity to the change in levels of frequency. However, there are some tendencies in patterns noticed, such as for the MD group: they showed a tendency to the opposite pattern (but not statistically confirmed): more fixations for high-frequency words than for low-frequency ones, and also for the PD group: a tendency of a pronounced increase in the number of fixations on low-frequency antecedents (but not statistically significant). PD had the most fixations across both word frequency conditions.

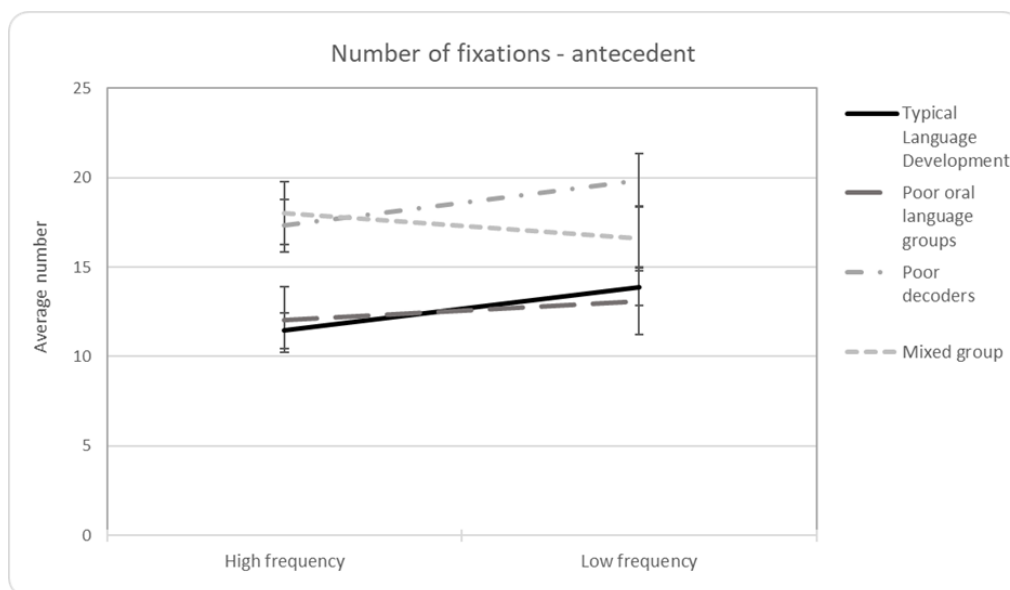


FIGURE 21. The interaction between frequency and group on Number of Fixations in the Antecedent Region

Results on extended (early and late) language processing measures

Table 40 presents the descriptive data for the measures *Number of Revisits* and *Regression out Probability* as indicators of both early and late stages of language processing. These measures were analysed in children with TLD and children with different profiles of LBD, in

relation to manipulated psycholinguistic properties of the target word - imageability and frequency.

Table 40

Descriptive Statistics for Number of Revisits and Regression out Probability on the Antecedent by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>		<i>Regression out probability</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	3.27	2.55	0.25	0.08
	POL	2.70	1.57	0.26	0.13
	PD	6.00	3.45	0.17	0.06
	MD	6.55	7.22	0.19	0.11
<i>HILF</i>	TLD	4.67	3.13	0.19	0.08
	POL	2.80	1.69	0.23	0.12
	PD	6.75	4.88	0.17	0.09
	MD	5.36	4.55	0.22	0.19
<i>LIHF</i>	TLD	3.52	2.87	0.27	0.17
	POL	4.50	3.03	0.23	0.09
	PD	5.19	3.95	0.20	0.11
	MD	7.00	6.36	0.18	0.07
<i>LILF</i>	TLD	4.15	2.99	0.23	0.16
	POL	5.40	2.84	0.18	0.06
	PD	6.88	5.25	0.16	0.07
	MD	5.27	3.95	0.19	0.06

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of separate mixed-design ANOVAs are shown in Tables 41 and 42.

Table 41

Mixed ANOVA Results for Late Language Processing Measure- Number of Revisits in TLD & LD subgroups

Number of revisits	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	1.860	1/66	.177	.027	.269
Frequency	0.727	1/66	.397	.011	.134
Imageability x Frequency	0.027	1/66	.869	.000	.053
Imageability x group	2.358	3/66	.080	.097	.567
Frequency x group	2.450	3/66	.071	.100	.585
Imageability x Frequency x Group	0.573	3/66	.635	.025	.162
<i>Independent</i>					
Group	3.125	3/66	.032	.124	.702

Number of revisits: The analysis for the *Number of Revisits* measure did not reveal any significant main effects of imageability or frequency, indicating that these psycholinguistic properties do not independently influence rereading behavior.

A significant main effect of group was found, suggesting that the number of revisits differed between language groups. However, post hoc comparisons (Bonferroni corrections) revealed no statistically significant pairwise differences ($p > .05$).

Table 42

Mixed ANOVA Results for Late Language Processing Measure- Regression out Probability in TLD & LD subgroups

Regression out probability	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.033	1/66	.856	.001	.054
Frequency	3.911	1/66	.052	.056	.496
Imageability x Frequency	0.571	1/66	.453	.009	.116
Imageability x group	1.344	3/66	.268	.058	.342
Frequency x group	2.055	3/66	.115	.085	.504
Imageability x Frequency x Group	0.456	3/66	.714	.020	.137
<i>Independent</i>					
Group	2.012	3/66	.121	.084	.495

Regression out Probability: The analysis revealed no significant main effect of imageability indicating that the probability of regression did not differ between high and low imageable words.

The main effect of frequency was marginally significant, suggesting a possible trend that words with low frequency may increase the likelihood of regressions. However, this effect did not reach the conventional threshold for significance and should therefore be interpreted with caution.

There were no significant interactions between imageability and frequency, or between psycholinguistic word properties and group.

The three-way interaction was not significant, nor was the main effect of group between-subjects. Although descriptive trends suggested some variation across groups, these differences within this measure were not statistically reliable.

Therefore, the regression out probability was not strongly influenced by psycholinguistic word properties or group membership. The pattern observed merely indicates a possible sensitivity to word frequency, particularly for certain groups.

Area of interest- Anaphora

Furthermore, the aim was to determine whether there are differences in the language processing patterns of children with different profiles of LBD compared to children with TLD, considering the psycholinguistic characteristics of the target word - its frequency and imageability - in the region of interest: the anaphora.

Results on early language processing measures

Table 43 shows the descriptive statistics for *First Fixation Duration* and *First Pass Duration*, which serve as indicators of the early phases of anaphora processing. The values are given for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF), and for each subgroup: children with TLD, and those with different profiles of LD (Poor Oral Language Group - POL, Poor Decoders - PD, Mixed Language difficulties Group - MD).

Table 43

Descriptive Statistics for Early Language Processing Measures (First Fixation Duration and First Pass Duration) by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>First Fixation Duration</i>		<i>First Pass</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	309.08	94.79	477.40	171.38
	POL	345.59	110.40	454.05	200.65
	PD	395.08	208.86	582.31	333.46
	MD	352.51	144.35	524.58	151.61
<i>HILF</i>	TLD	365.65	136.35	678.18	730.89
	POL	393.91	140.56	556.19	191.16
	PD	344.86	182.84	815.63	806.77
	MD	390.01	228.87	678.74	618.29
<i>LIHF</i>	TLD	319.26	90.35	562.09	671.89
	POL	335.38	84.43	874.04	1169.07
	PD	378.32	192.60	655.39	449.56
	MD	375.50	142.72	728.28	527.60
<i>LILF</i>	TLD	339.95	129.88	500.02	181.12
	POL	369.11	60.55	549.08	111.43
	PD	411.66	265.00	787.56	874.03
	MD	306.45	96.31	451.10	106.56

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of separate mixed-design ANOVAs are shown in Tables 44 and 45.

Table 44

Mixed ANOVA Results for Early Language Processing Measure - First fixation duration in TLD & LBD subgroups

First Fixation Duration	<i>F</i>	df	<i>P</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.297	1/66	.588	.004	.084
Frequency	0.343	1/66	.914	.014	.156
Imageability x Frequency	0.482	1/66	.500	.008	.107
Imageability x group	0.555	3/66	.604	.031	.191
Frequency x group	1.206	3/66	.315	.052	.309
Imageability x Frequency x Group	2.296	3/66	.086	.094	.554
<i>Independent</i>					
Group	0.642	3/66	.590	.028	.178

First Fixation Duration: The analysis of the *First Fixation Duration* measure revealed no significant main effects of imageability, frequency, or participant group. Moreover, none of the interactions between these variables reached statistical significance. These results suggest that the early stages of anaphora processing, indicated by initial fixation duration, were not systematically influenced by the psycholinguistic properties of the target words or by specific profile differences.

Table 45

Mixed ANOVA Results for Early Language Processing Measure - First Pass duration in TLD & LBD subgroups

First Pass	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.380	1/66	.540	.006	.093
Frequency	0.090	1/66	.765	.001	.060
Imageability x Frequency	4.984	1/66	.029	.070	.595
Imageability x group	0.620	3/66	.604	.027	.173
Frequency x group	0.942	3/66	.426	.041	.247
Imageability x Frequency x Group	0.312	3/66	.817	.014	.107
<i>Independent</i>					
Group	0.820	3/66	.487	.036	.218

First Pass Duration: Analysis for the *First Pass Duration* measure revealed no significant main effects of the psycholinguistic properties of the target words or group membership. However, a significant interaction between frequency and imageability was found. This

pattern mirrors that observed in the two-group analysis, as group effects were not significant in the present four-group model. However, as the post hoc comparisons did not reach significance, the precise nature of this interaction remains unclear ($p > .05$). The shortest early processing times were observed for anaphors with aligned psycholinguistic properties (high-frequency/high-imageability and low-frequency/low-imageability), whereas longer first-pass durations were evident for mixed conditions (low-imageability/high-frequency and high-imageability/low-frequency). This pattern tentatively indicates that congruence between frequency and imageability may facilitate processing, while mismatched properties may introduce additional processing demands. No significant interactions with the group were observed.

Results on late language processing measures

Table 46 presents descriptive data for the measures *Dwell Time* and *Number of Fixations* as indicators of later stages of language processing. These measures were analysed in children with TLD and those with different profiles of LBD, in relation to manipulated psycholinguistic properties of the target word - imageability and frequency.

Table 46

Descriptive Statistics for Dwell Time and Number of Fixations on the Anaphor by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>Dwell Time</i>		<i>Number of Fixations</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	830.14	377.76	19.70	5.81
	POL	891.27	185.87	20.40	5.32
	PD	1228.21	586.14	24.56	10.18
	MD	1226.75	523.17	23.73	9.23
<i>HILF</i>	TLD	1121.13	684.81	21.82	9.23
	POL	1013.94	387.51	19.30	7.47
	PD	1660.96	927.35	30.06	16.82
	MD	1931.95	1189.16	32.00	19.34
<i>LIHF</i>	TLD	893.69	431.48	19.67	7.79
	POL	895.76	270.98	18.50	6.72
	PD	1296.33	670.58	25.00	12.61
	MD	1560.97	1164.38	25.45	16.04
<i>LILF</i>	TLD	1073.50	649.14	21.82	9.13
	POL	962.99	319.43	17.70	5.85
	PD	1509.05	786.18	26.50	12.64
	MD	1687.12	1289.75	34.36	31.18

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of the separate mixed-design ANOVAs are shown in Tables 47 and 48.

Table 47**Mixed ANOVA Results for Late Language Processing Measure - Dwell Time in TLD & LBD subgroups**

Dwell Time	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.005	1/66	.944	.000	.051
Frequency	24.690	1/66	<.001	.272	.998
Imageability x Frequency	7.468	1/66	.008	.102	.768
Imageability x group	0.171	3/66	.915	.008	.080
Frequency x group	1.269	3/66	.292	.055	.324
Imageability x Frequency x Group	1.557	3/66	.208	.066	.392
<i>Independent</i>					
Group	4.308	3/66	.008	.164	.846

Dwell Time: Analysis of the *Dwell Time* measure revealed a significant main effect of word frequency, main effect of group, and significant imageability x frequency interaction (Figure 22). This observed interaction indicates that in the high-imageability condition, dwell time varied as a function of word frequency- the longest dwell time was observed in the high imageability and low frequency condition ($p < .05$), whereas no such significant effect of frequency was observed for low-imageability words ($p > .05$). No significant main effect of imageability was found, and no significant interactions were observed between the psycholinguistic variables and the participant group. The three-way interaction between imageability, frequency, and group was also not statistically significant.

A significant main effect of group was also observed. Post hoc comparisons revealed a statistically significant difference in dwell time between children in the mixed group- MD and those with TLD ($p = .028$). Specifically, MD group showed significantly longer dwell times for anaphoric expressions than their peers with TLD. No other pairwise comparisons between the groups reached statistical significance. These results suggest that children with co-occurring (mixed) oral language and decoding difficulties may process anaphoric references with greater effort or require more time, even when the main psycholinguistic variables (frequency and imageability) are held constant.

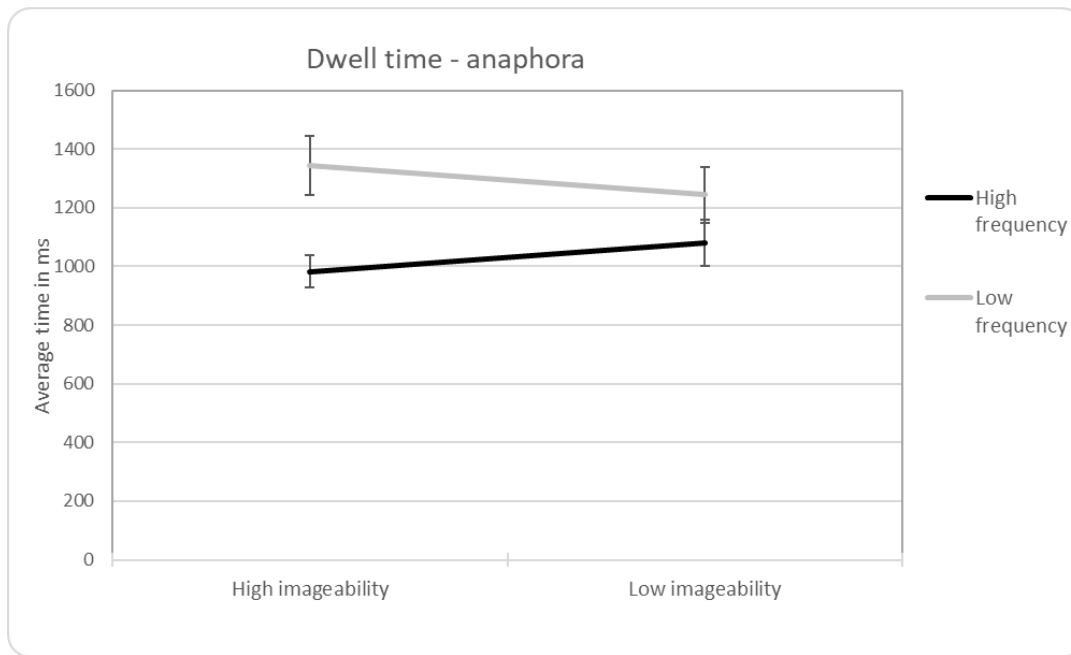


Figure 22. The interaction between imageability and frequency on Dwell Time in the Anaphora Region

Table 48

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations in TLD & LBD subgroups

Number of Fixations	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.110	1/66	.741	.002	.062
Frequency	11.946	1/66	.001	.153	.926
Imageability x Frequency	0.186	1/66	.667	.003	.071
Imageability x group	0.696	3/66	.558	.031	.190
Frequency x group	3.387	3/66	.023	.133	.741
Imageability x Frequency x Group	0.424	3/66	.736	.019	.130
<i>Independent</i>					
Group	2.900	3/66	.041	.116	.666

Number of fixations: The analysis revealed a significant main effect of word frequency, and main effect of group. No main effect of imageability was found, indicating that imageability alone did not influence the number of fixations. Although the interaction between frequency and imageability was not significant, a significant frequency $\times$ group interaction was observed (Figure 23). Post hoc comparisons showed that the effect of frequency was present only in the MD group, with low-frequency words eliciting a higher number of fixations than high-frequency words ($p < .05$). No such difference was observed in the remaining groups.

No other interactions, including the three-way interaction, reached significance.

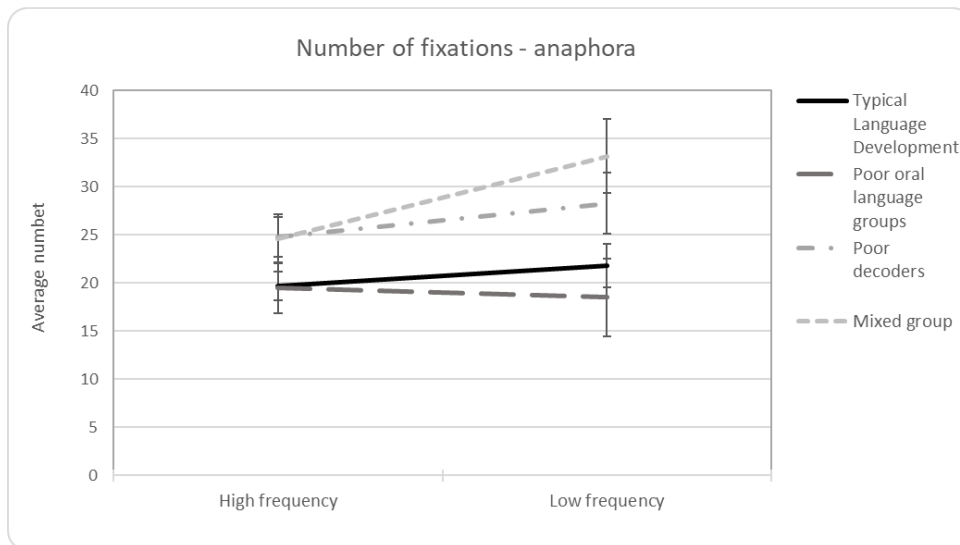


Figure 23. The interaction between frequency and group on Number of fixations in the Anaphora Region

Results on extended (early and late) language processing measures

Table 49 presents the descriptive data for the measures *Number of Revisits* and *Regression out Probability* as indicators of early and late stages of language processing. These measures were analysed in children with TLD and in children with different profiles of language-based difficulties, in relation to manipulated psycholinguistic properties of the target word - frequency and imageability.

Table 49

Descriptive Statistics for Number of Revisits and Regression out Probability on the Anaphor by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>		<i>Regression out Probability</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	6.39	3.19	.15	.05
	POL	6.30	3.02	.14	.04
	PD	8.00	4.62	.15	.06
	MD	9.18	5.74	.15	.06
<i>HILF</i>	TLD	6.97	4.86	.14	.06
	POL	6.50	4.20	.15	.05
	PD	9.94	8.43	.11	.05
	MD	11.09	8.44	.11	.05
<i>LIHF</i>	TLD	5.88	3.78	.15	.05
	POL	5.80	3.88	.15	.08
	PD	6.81	4.10	.13	.05
	MD	7.55	7.23	.15	.09
<i>LILF</i>	TLD	6.70	4.55	.14	.06
	POL	4.60	2.80	.16	.06
	PD	9.06	5.89	.12	.06

	MD	12.45	13.90	.11	.06
--	----	-------	-------	-----	-----

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of separate mixed-design ANOVAs are shown in Tables 50 and 51.

Table 50

Mixed ANOVA Results for Late Language Processing Measure - Number of Revisits in TLD & LBD subgroups

Number of Revisits	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	2.306	1/66	.134	.034	.322
Frequency	8.264	1/66	.005	.111	.809
Imageability x Frequency	0.309	1/66	.580	.005	.085
Imageability x group	0.295	3/66	.829	.013	.104
Frequency x group	2.486	3/66	.068	.102	.592
Imageability x Frequency x Group	0.677	3/66	.569	.030	.185
<i>Independent</i>					
Group	2.285	3/66	.087	.094	.552

Number of revisits: Analysis of the *Number of Revisits* measure revealed a significant main effect of word frequency, indicating that low-frequency anaphoric expressions elicited more revisits to the anaphoric region than high-frequency ones. This suggests that less frequent anaphoric expressions require additional visual inspection during anaphora resolution.

In contrast, there was no significant main effect of imageability, indicating that the vividness or concreteness of the anaphoric expression alone did not significantly affect returns to the referential region.

Overall, frequency was found to be the primary psycholinguistic variable influencing revisit patterns in anaphora processing.

Table 51

Mixed ANOVA Results for Late Language Processing Measure - Regression out Probability in TLD & LBD subgroups

Regression out probability	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.256	1/66	.615	.004	.079
Frequency	5.658	1/66	.020	.079	.649
Imageability x Frequency	0.813	1/66	.371	.012	.144
Imageability x group	0.198	3/66	.897	.009	.085
Frequency x group	2.428	3/66	.073	.099	.580
Imageability x Frequency x Group	0.362	3/66	.781	.016	.117

<i>Independent</i>					
Group	0.898	3/66	.447	.039	.236

Regression out probability: Analysis of *Regression out probability (RoP)* revealed a significant main effect of word frequency, with low-frequency words having lower RoP. This suggests that more regressive eye movements out of anaphoric region are triggered by high-frequency words, potentially supporting readers to revisit other areas for clarification or integration.

In contrast, imageability did not have a significant influence on the likelihood of regression, implying that this psycholinguistic property alone did not significantly affect the resolution process. The main effect of group was also not significant, suggesting that children with different language-based profiles did not differ significantly overall in their use of regressions as a rereading strategy.

Area of interest - Words preceding AOIs

First Fixation Duration: Table 52 shows the descriptive statistics for the *First Fixation Duration*. The values are given for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF), and for each subgroup of children: children with TLD, and children with different profiles of LD (Poor Oral Language Group - POL, Poor Decoders - PD, Mixed Language Difficulties Group - MD).

Table 52

Descriptive Statistics for First Fixation Duration by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>First Fixation Duration</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	312.90	71.43
	POL	346.37	64.33
	PD	342.19	93.64
	MD	349.74	150.63
<i>HILF</i>	TLD	344.50	82.22
	POL	351.86	53.62
	PD	338.97	91.11
	MD	345.89	171.59
<i>LIHF</i>	TLD	317.40	84.32
	POL	354.44	54.62
	PD	320.94	102.85
	MD	343.26	153.17
<i>LILF</i>	TLD	329.43	103.30

POL	355.31	71.59
PD	349.57	99.96
MD	353.19	157.36

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results for mixed-design ANOVA are shown in Table 53.

Table 53

Mixed ANOVA Results for Early Language Processing Measure - First Fixation Duration before/ after AOI in TLD & LBD subgroups

First fixation duration	<i>F</i>	Df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.045	1/66	.833	.001	.055
Frequency	3.953	1/66	.051	.057	.500
Imageability x Frequency	0.255	1/66	.615	.004	.079
Imageability x group	0.246	3/66	.864	.011	.094
Frequency x group	1.021	3/66	.389	.044	.265
Imageability x Frequency x Group	1.634	3/66	.190	.069	.410
<i>Independent</i>					
Group	0.280	3/66	.840	.013	.101

The analysis of the first fixation duration did not reveal any statistically significant main effects or interactions. Neither imageability nor the interaction between imageability and frequency influenced fixation duration. Although the main effect of word frequency was marginally significant, suggesting a possible tendency towards longer first fixations for low-frequency words, this effect did not reach conventional significance thresholds. Similarly, there were no significant interactions between the experimental variables and participant group, nor was there a main effect of group, although participants with LBD showed slightly longer fixations.

Dwell Time: Table 54 shows the descriptive statistics for the *Dwell Time*. The values are given for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF), and for each subgroup of children: children with TLD, and children with different profiles of LBD (Poor Oral Language Group - POL, Poor Decoders - PD, Mixed Language Difficulties Group - MD).

Table 54

Descriptive Statistics for Dwell Time by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>Dwell Time</i>
------------------	--------------	-------------------

		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	1088.78	523.27
	POL	1051.40	296.53
	PD	1627.34	737.07
	MD	1797.71	945.57
<i>HILF</i>	TLD	1516.56	741.81
	POL	1414.63	478.72
	PD	2341.26	895.45
	MD	2442.74	1403.72
<i>LIHF</i>	TLD	1256.52	567.74
	POL	1213.35	339.43
	PD	1879.22	923.39
	MD	2118.16	1316.09
<i>LILF</i>	TLD	1393.26	812.23
	POL	1319.97	417.12
	PD	2161.51	1005.38
	MD	2043.09	1265.08

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of mixed-design ANOVAs are shown in Table 55.

Table 55

Mixed ANOVA Results for Late Language Processing Measure - Dwell Time before/ after AOI in TLD & LBD subgroups

First fixation duration	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	0.154	1/66	.696	.002	.067
Frequency	35.159	1/66	<.001	.348	1.000
Imageability x Frequency	34.477	1/66	<.001	.343	1.000
Imageability x group	0.237	3/66	.870	.011	.093
Frequency x group	1.256	3/66	.297	.054	.321
Imageability x Frequency x Group	1.949	3/66	.130	.081	.481
<i>Independent</i>					
Group	5.321	3/66	.002	.195	.918

The analysis revealed a significant main effect of frequency, as well as a significant interaction between imageability and frequency (Figure 24). Post hoc comparison shows longer dwell times for low-frequency words in general, but even longer for low-frequency, high-imageability than for low-frequency, low-imageability words ($p < .05$). In high frequency conditions, the pattern was opposite: high-frequency, low-imageability words required longer dwell time than high-frequency, high-imageability words ($p < .05$). In low imageability conditions, there is no significant difference between low and high frequency (p

>.05). No other main effects or interactions involving imageability or group have reached significance.

Importantly, a significant between-subjects effect was observed for group, indicating differences in overall dwell time between participant subgroups. Post hoc comparison showed that TLD group spent significantly less time fixating on anaphora than the PD ($p = .025$) and MD ($p = .025$).

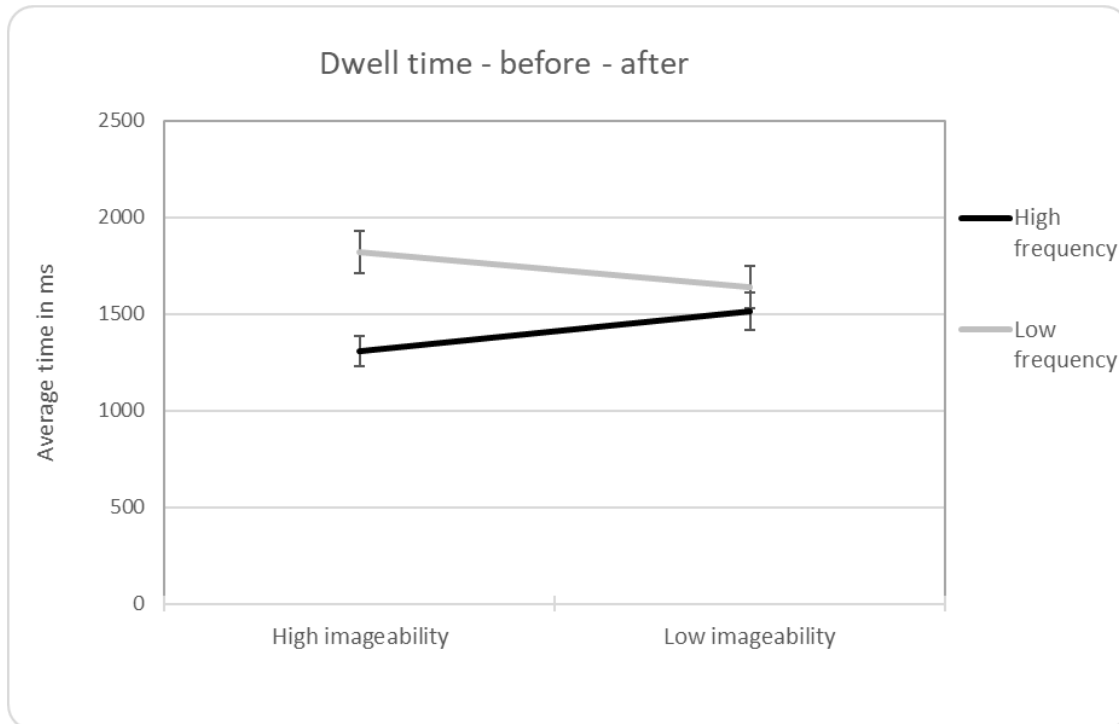


FIGURE 24. The interaction between imageability and frequency on Dwell Time in Words Preceding AOI

Number of fixations: Table 56 shows the descriptive statistics for the *Number of Fixations*. The values are given for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF), and for each subgroup of children: children with TLD, and children with different profiles of LBD (Poor Oral Language Group - POL, Poor Decoders - PD, Mixed Language Difficulties Group - MD).

Table 56

Descriptive Statistics for Number of Fixations by Condition and Subgroup

Condition	Group	Number of Fixations	
		M	SD
HIHF	TLD	71.70	16.50
	POL	67.60	15.37
	PD	97.25	37.18
	MD	109.45	64.84

<i>HILF</i>	TLD	92.67	27.87
	POL	79.30	19.28
	PD	127.25	55.86
	MD	129.00	78.99
<i>LIHF</i>	TLD	84.52	21.38
	POL	74.50	15.91
	PD	112.06	40.02
	MD	121.64	73.26
<i>LILF</i>	TLD	87.15	27.65
	POL	77.40	23.17
	PD	121.37	60.78
	MD	110.82	63.04

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of mixed-design ANOVA are shown in Table 57.

Table 57

Mixed ANOVA Results for Late Language Processing Measure - Number of Fixations before/ after AOI in TLD & LBD subgroups

Number of Fixations	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	1.563	1, 66	.216	.023	.234
Frequency	20.981	1, 66	.000	.241	.995
Imageability x Frequency	33.122	1, 66	.000	.334	1.000
Imageability x group	1.113	3, 66	.350	.048	.287
Frequency x group	1.919	3, 66	.135	.080	.474
Imageability x Frequency x Group	1.286	3, 66	.286	.055	.328
<i>Independent</i>					
Group	4.482	3, 66	.006	.169	.861

The results showed a significant main effect of word frequency, indicating that low-frequency words triggered more fixations than high-frequency words. A significant interaction between imageability and frequency was also observed (Figure 25). Post hoc comparison shows that low-frequency words elicited a higher number of fixations than high-frequency words, but only in the high-imageability condition. No difference between frequency levels was observed under low-imageability conditions.

There was a significant main effect of group. However, post hoc comparison did not find any significant difference between the groups (all $p > .05$), suggesting that the observed differences are not related to the specific language profile.

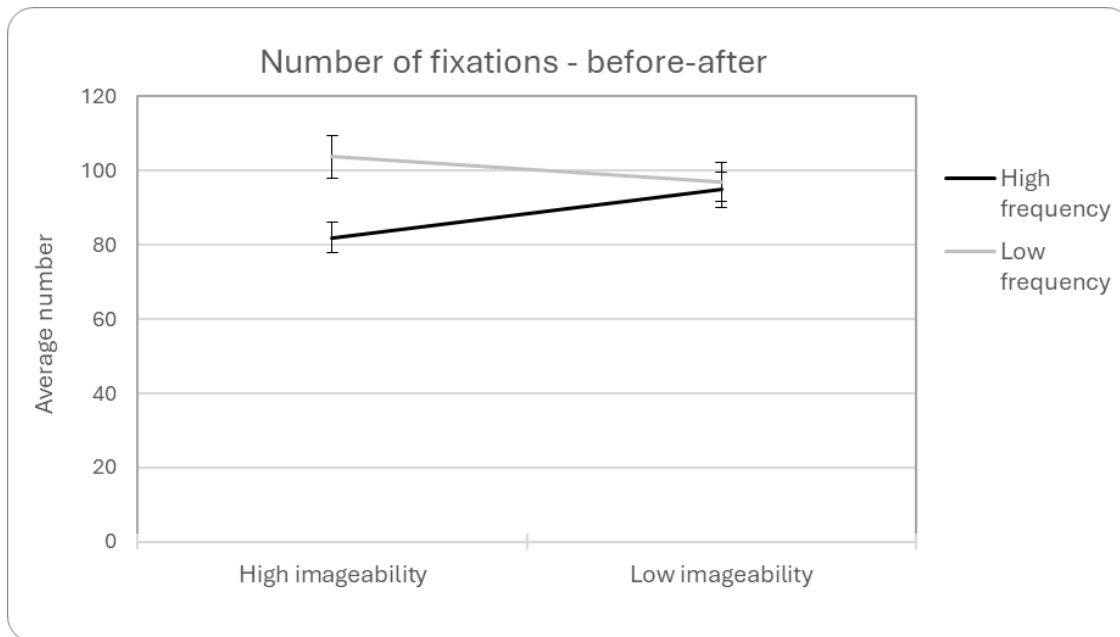


FIGURE 25. The interaction between imageability and frequency on Number of Fixations in Words Preceding AOI

Number of revisits: Table 58 shows the descriptive statistics for the measure *Number of Revisits*. The values are given for four experimental conditions, defined by high and low imageability and frequency of the target word (HIHF, HILF, LIHF, LILF), and for each subgroup of children: children with TLD, and children with different profiles of LBD (Poor Oral Language group - POL, Poor Decoders - PD, Mixed Language Difficulties Group - MD).

Table 58

Descriptive Statistics for Number of Revisits by Condition and Subgroup

<i>Condition</i>	<i>Group</i>	<i>Number of Revisits</i>	
		<i>M</i>	<i>SD</i>
<i>HIHF</i>	TLD	19.24	7.90
	POL	18.60	10.06
	PD	27.69	13.84
	MD	32.91	23.52
<i>HILF</i>	TLD	27.76	13.06
	POL	22.00	9.62
	PD	36.31	18.78
	MD	41.82	33.54
<i>LIHF</i>	TLD	25.18	11.56
	POL	21.20	8.95
	PD	30.81	14.56
	MD	41.64	33.10
<i>LILF</i>	TLD	25.48	12.48

	POL	23.80	12.63
	PD	36.00	21.92
	MD	36.73	25.45

Note. TLD = Typical Language Development; POL = Poor Oral Language; PD = Poor Decoders; MD = Mixed Language Difficulties; HIHF= High Imageability - High Frequency; HILF= High Imageability - Low Frequency; LIHF = Low Imageability - High Frequency; LILF = Low Imageability - Low Frequency.

The results of mixed-design ANOVA are shown in Table 59.

Table 59

Mixed ANOVA Results for Late Language Processing Measure - Number of Revisits before/ after AOI in TLD & LBD subgroups

Number of Revisits	<i>F</i>	df	<i>p</i>	η^2	Observed power
<i>Dependent</i>					
Imageability	3.540	1, 66	.064	.051	.458
Frequency	19.069	1, 66	.000	.224	.990
Imageability x Frequency	9.719	1, 66	.003	.128	.867
Imageability x group	0.026	3, 66	.994	.001	.054
Frequency x group	1.243	3, 66	.301	.053	.318
Imageability x Frequency x Group	1.528	3, 66	.215	.065	.385
<i>Independent</i>					
Group	3.341	3, 66	.024	.132	.734

The analysis revealed a main effect of frequency, with low-frequency words resulting in more revisits across all groups.

A significant interaction between imageability and frequency (Figure 26) indicated that their combined effect influenced revisitation patterns. Post comparison shows that frequency effect was observed only in the high-imageability condition, with fewer revisits for high-frequency than for low-frequency words ($p < .05$). No difference between frequency levels was found under low-imageability conditions, indicating possibility that frequency influenced revisiting behaviour only when words were highly imageable (Figure 26).

No interactions with group were statistically significant. However, the main effect of group was significant, suggesting general differences between participant groups. Post hoc comparisons did not find any statistically significant pairwise differences between the groups (all $p > .05$), suggesting that the observed differences reflect overall disorder-level effects rather than specific group-to-group differences.

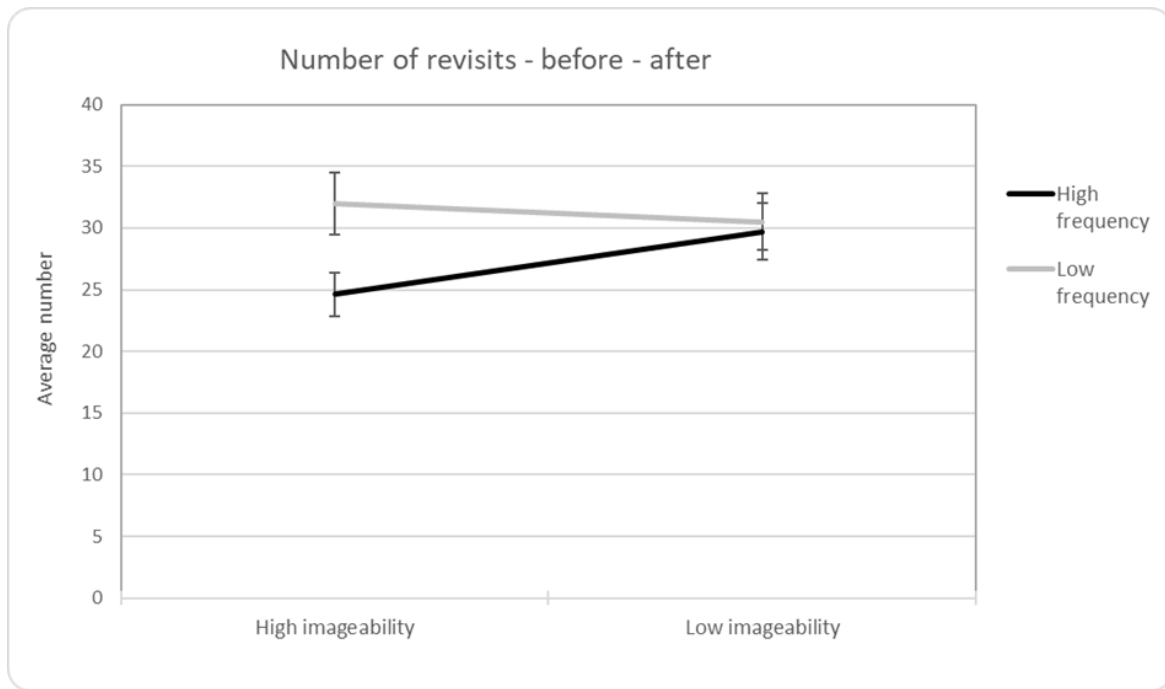


FIGURE 26. The interaction between imageability and frequency on Number of Revisits in Words Preceding AOI

5.1.4 Summary of results for children with typical language development and specific profiles of language-based disorder across AOIs

The analyses on four subgroups (TLD vs. LBD: PD, MD, POL) reveal that children with language-based disorders do not constitute a homogeneous category. In the early processing phase of the antecedent region, these group differences are absent - pattern of the initial lexical access is consistent in all groups. On the first pass measure, only PD and MD groups show sensitivity to frequency effect, with longer duration for low frequency antecedents. In later processing phases emerges crucial difference: all groups process high frequency antecedents faster, except the POL group, which does not benefit from high frequency despite a global processing speed similar to the others, indicating a weakened quality of lexical representations. The PD group makes the most fixations, reflecting the least degree of automaticity, while MD tends towards a potentially reversed pattern of attentional deployment. TLD group only exhibits a flexible, frequency-modulated fixation strategy, as a hallmark of an efficient reader.

In the processing of anaphora into four subgroups, group differentiation is significantly narrowed compared to the antecedent. In the early processing phase no effect is significant in any group, and the interaction of imageability and frequency in the First Pass remains unclear as in the two-group analysis. In later processing phase, the MD group stands out as the most distinctive: it is the only one that shows significantly longer dwell time than the TLD and the

only one that responds to the frequency of the anaphora in the number of fixations, meaning that low-frequency anaphors elicit more fixations exclusively in this group. The interaction of imageability and frequency consistently confirms that high imageability, low frequency anaphors are the most demanding. In the extended processing phase, the group differences completely disappear, because all subgroups show the same pattern of multiple regressions for low frequency anaphora.

In the processing of words preceding the AOIs, the psycholinguistic pattern was consistent with the previous analyses. Frequency emerged as the dominant predictor of cognitive effort, with its effects across all late and extended processing measures confined exclusively to the high imageability condition. Low frequency, high imageability words were consistently the most demanding, whereas under low imageability conditions no frequency-related differences were observed. Group differentiation was diffuse and without completely clearly defined subgroup distinctions: the TLD group showed shorter overall fixation times than the PD and MD groups, whilst the POL group again did not differ significantly from the TLD group. During the extended stages of processing, no subgroup-specific differences were identified, suggesting that the observed effects operate primarily at the broader disorder level rather than within individual language-based profiles.

5.2 Analysis 2: Offline comprehension

5.2.1 Reading comprehension performance of children with TLD and children with LBD

This section presents the results of Analysis 2, in which children's offline reading comprehension was assessed by analysing their accuracy in answering literal and inferential questions about narrative texts. The aim was to examine the interaction between question type and language profile, and to determine whether different comprehension profiles lead to different performance patterns. The descriptive data are presented in Table 60 by question type and participant group, following Figure 27, which visually highlights group differences and possible effects of question type on reading comprehension performance. The accuracy results for literal and inferential questions are then presented for each participant group. The graph enables a comparison of comprehension patterns between children with TLD and children with LBD.

Table 60

Descriptive statistics for literal and inferential questions by group (TLD vs. LBD)

<i>Question Type</i>	<i>Group</i>	<i>M</i>	<i>SD</i>	<i>N</i>
<i>Literal questions</i>	TLD	3.70	0.64	33
	LBD	2.38	1.16	37
<i>Inferential questions</i>	TLD	3.36	0.86	33
	LBD	1.84	1.28	37

Note. TLD= typical development; LBD= language-based disorder

The descriptive data show that participants were more accurate on literal questions than on inference questions, with children with TLD performing better on both types of questions than children with LBD.

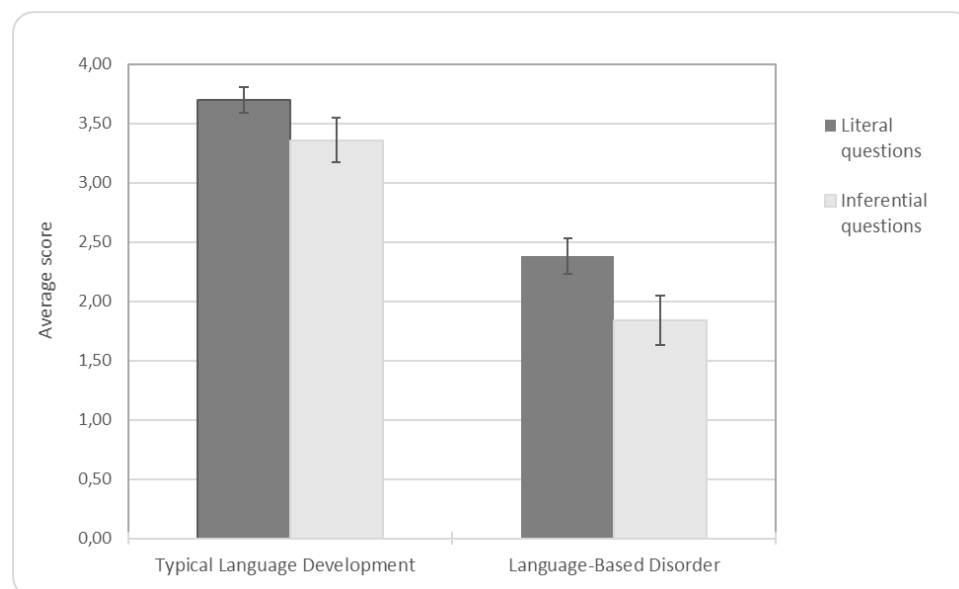


FIGURE 27. Accuracy on Literal and Inferential Questions by Groups

Note: Error bars represent standard errors of the mean. TLD = Typically Developing; LBD = Language-Based Disorders

A mixed-design ANOVA revealed a significant main effect of question type, $F(1, 68) = 10.98, p = .001, \eta^2 = .14$, and a significant main effect of group, $F(1, 68) = 46.55, p < .001, \eta^2 = .41$. The interaction between question type and group was not statistically significant, $F(1, 68) = 0.62, p = .435, \eta^2 = .009$.

5.2.2 Reading comprehension performance in children with TLD and children with different LBD profiles

The same analysis was conducted for the results of all four participant groups. The results are presented in Table 61 and shown in Figure 28.

Table 61

Descriptive statistics for literal and inferential questions by group (TLD vs. LBD profiles)

Question Type	Group	M	SD	N
<i>Literat questions</i>	TLD	3.70	0.64	33
	Poor oral language group	2.30	1.50	10
	Poor decoders	2.75	0.86	16
	Mixed group	1.91	1.14	11
<i>Inferential questions</i>	TLD	3.36	0.86	33
	Poor oral language group	1.80	1.40	10
	Poor decoders	2.25	1.24	16
	Mixed group	1.27	1.10	11

Note. TLD= typical development; M = mean; SD = standard deviation. Maximum score per question type was 4.

The descriptive data show that children with TLD were, on average, more successful in reading comprehension than all groups of children with LBD, while children with combined language difficulties (mixed group) achieved the lowest scores. Across all groups, lower accuracy was generally observed for inferential questions than for literal questions, with the differences most pronounced in children with mixed language difficulties (mixed group).

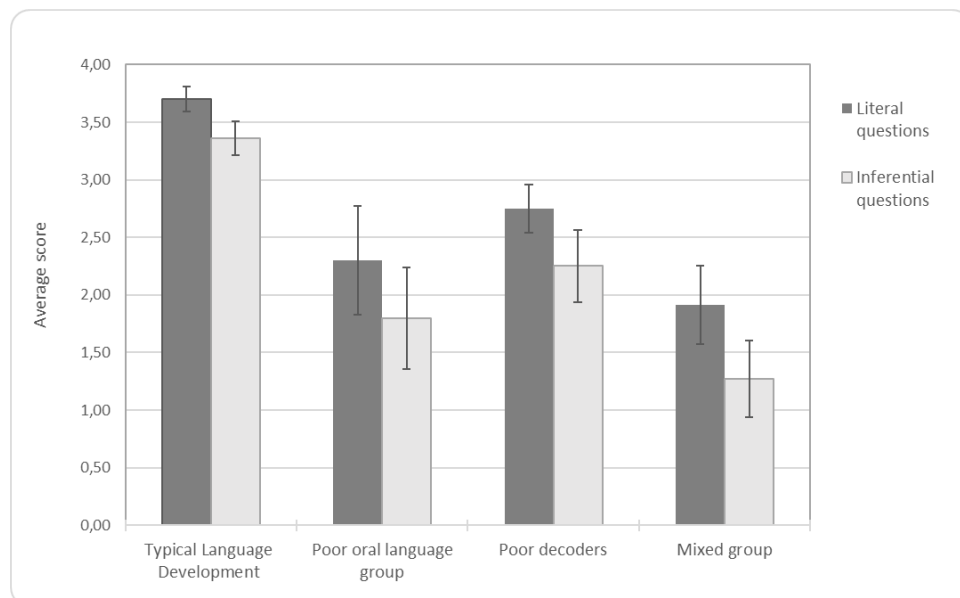


FIGURE 28. Accuracy on Literal and Inferential Questions by Group

The analysis revealed a statistically significant main effect of question type, $F(1, 66) = 10.96, p = .002, \eta^2 = 0.14$, and a significant main effect of group, $F(3, 66) = 19.43, p < .001, \eta^2 = 0.47$. However, the interaction between these two factors was not significant, $F(3, 66) = 0.24, p = .869, \eta^2 = .011$.

Post-hoc comparisons by group using the Scheffé test showed that children with typical language development performed significantly better than all clinical subgroups ($p < .01$), while the differences among the clinical subgroups were not statistically significant, except between poor decoders and mixed group (Bonferroni's: $p = .043$).

5.3 Analysis 3: Predictive components

This section presents the results of the predictive analysis conducted to determine which (meta)linguistic and cognitive factors, based on the Simple View of Reading (Gough & Tunmer, 1986), contribute to individual differences in reading comprehension, over and above the child's clinical language profile. Hierarchical regression analyses were used to examine the individual and combined contributions of the SVR component skills: a) decoding (phonological language skills); b) language comprehension (non-phonological language skills); and the additional individual component c) phonological working memory - to both literal and inferential comprehension (reading comprehension). This allowed for an assessment of the contributions of different cognitive-linguistic mechanisms to reading comprehension, in the context of the SVR model.

Reading comprehension performance (the criterion variable) was defined as the total number of correctly answered literal and inferential questions based on a read discourse (four narrative stories, as in the previous assessments).

Considering the number of individual predictor variables included in the study, and the fact that most are primarily theoretically (construct-wise) and empirically interrelated (SVR; Gough & Tunmer, 1986), theoretically and methodologically sound composite scores were formed. In other words, the individual variables were combined into composite measures, based primarily on theoretical justifications (the SVR model) and subsequently on methodological considerations. This approach aimed to integrate the results of multiple related items, create more coherent and stable measures of the targeted constructs (in this case, (meta)language phonological and non-phonological skills), reduce the influence of random measurement error, improve internal consistency, and enable a clearer interpretation of construct relationships within theoretical models (Song et al., 2017).

Three groups of variables were included in the study as potential composite predictors of the reading comprehension (RC), based on the Simple View of Reading model (Gough & Tunmer, 1986) and its specific components:

- **Decoding (phonological (meta)linguistic skills):** word and pseudoword reading accuracy and speed, and phonological awareness (spoonerisms).
- **Language comprehension (non-phonological (meta)linguistic skills):** grammatical comprehension (TROG:2-HR), vocabulary breadth (receptive vocabulary - PPVT-III HR), inferencing (NRDLS-HR), grammatical judgment (NRDLS-HR), verbal semantic fluency, and vocabulary depth (WISC-IV).
- **Phonological working memory (phonological loop):** pseudoword repetition.

The first step in creating the composite scores was to examine the intercorrelations between the individual variables within each theoretically predefined composite. Almost all measures showed at least one significant correlation ($p < .05$ or $< .01$) with another measure (see Table 38). However, the measure of *Verbal Semantic Fluency* showed exceptionally weak and non-significant associations with *Inferencing* ($r = .08$) and *Grammatical Judgement* ($r = .11$). As it did not correspond well with any potential composite, this variable was excluded from further analysis.

In contrast, the Spoonerisms measure showed strong but negative correlations with certain phonological measures - for example, with *Word Reading Errors* ($r = -.58$) and *Word Reading Time* ($r = -.52$). Nevertheless, it was not excluded. Instead, its relationship to other variables, particularly *Pseudoword Repetition*, was analysed. Given their theoretical complementarity, these two measures were combined into a newly constructed composite representing phonological processing. Thus, the *Spoonerisms* variable was combined with *Pseudoword Repetition*, a measure of the phonological loop, to form a theoretically grounded composite reflecting broader phonological processing skills.

Table 62. Pearson Correlation Matrix Among Phonological, Non-phonological language skills, and Phonological Working Memory

Variables	TROG	PPVT	VD	Infer	VSF	GJ	WRE	WRT	PWRE	PWRT	PW R	SP
TROG	1	.352**	.504**	.362**	.338**	.399**	-.205	.011	-.090	.009	.528**	.470**
PPVT		1	.506**	.444**	.235	.132	-.018	-.136	-.008	-.109	.259*	.290*
Vocabulary Depth (VD)			1	.439**	.391**	.295*	-.311**	-.174	-.241*	-.068	.489**	.544**
Inferencing (Infer)				1	.084	.157	-.196	-.131	-.096	-.089	.416**	.297*

Variables	TROG	PPVT	VD	Infer	VSF	GJ	WRE	WRT	PWRE	PWRT	PW R	SP
Verbal Semantic Fluency (VSF)					1	.110	-.211	-.217	-.200	-.173	.241*	.325**
Grammatical Judgement (GJ)						1	-.227	-.037	-.083	-.003	.544**	.308**
Word Reading Errors (WRE)							1	.547**	.775**	.326**	-.469**	-.576**
Word Reading Time (WRT)								1	.480**	.870**	-.329**	-.524**
Pseudoword Reading Errors (PWRE)									1	.336**	-.409**	-.541**
Pseudoword Reading Time (PWRT)										1	-.162	-.403**
Pseudoword Repetition (PW R)											1	.519**
Spoonerisms (SP)												1

Note: Significant correlations are marked as * $p < .05$ and ** $p < .01$.

In the second step of the composite construction process, only those phonological and non-phonological (meta)linguistic variables, as well as the phonological loop measure that showed statistically significant group differences between TLD and LBD children based on t -tests were included (Appendix No. 4). The only variable that did not show a statistically significant difference was Grammatical Judgement ($t(1, 68) = 1.67, p = .09$) which was therefore excluded from further analysis.

To assess internal consistency, Cronbach's alpha values were calculated for each composite variable. The results showed adequate to strong internal reliability, supporting the inclusion of the selected variables in the proposed composites:

- the decoding (phonological (meta)linguistic) composite ($\alpha = .83$), representing phonological decoding accuracy and speed;
- the language comprehension (non-phonological (meta)linguistic) composite ($\alpha = .76$), reflecting language comprehension skills, reasoning, and semantic knowledge; and
- the phonological processing composite ($\alpha = .68$).

Table 63 shows the final set of composites used in the subsequent regression analyses.

Table 63. Final Composite Variables and Their Constituent Measures

Composite	Variables
Decoding composite	Word Reading Errors
	Pseudoword Reading Errors
	Word Reading Time
	Pseudoword Reading Time
Language comprehension composite	Receptive vocabulary (PPVT-III-HR)
	Vocabulary Depth (WISC-IV)
	Inferencing (NRDLS-HR)
	Grammatical comprehension (TROG-2:HR)
Phonological processing composite	Spoonerisms (Phonological Awareness)
	Pseudoword Repetition

All variable scores were standardised (expressed as *z*-scores). Each composite was calculated by averaging the *z*-scores of the variables comprising the composite for each participant. These composite scores were then used in the regression analysis to examine the extent to which each constructed composite explained the variance in the criterion variable (i.e., the total score on the comprehension questions).

Within the three composite constructs, the direction of the individual variable scores was consistent in each composite. For the language comprehension (non-phonological (meta)linguistic) and phonological processing composites, higher composite scores reflected better performance. For the decoding (phonological (meta)linguistic) composite, however, the direction was reversed - higher composite scores indicated poorer performance, as the variables contained in this composite (e.g., error counts) were inversely related to task success.

Table 64 shows the results of the regression analysis conducted to test the hypotheses for the third research question. The hierarchical multiple regression analysis showed that group membership and the language comprehension (composite significantly predicted reading comprehension performance, $F(2, 67) = 37.31, p < .001, R^2 = .527$. The addition of the decoding composite and the phonological processing composite did not significantly increase

the explained variance ($p > .05$). The strongest individual predictor was the language comprehension composite ($\beta = .365, p = .007$), while the decoding composite ($\beta = .007$) and phonological processing ($\beta = .192$) were not statistically significant.

Table 64. Hierarchical multiple regression predicting Reading comprehension from group membership and decoding (phonological), language comprehension (non-phonological) and phonological processing composite predictors

Step	Predictor	β	p	F	p	R	R ²	Adjusted R ²
1	Group (0 = TLD, 1 = LBD)	−0.64	<.001	46.550	<.001	.637	.406	.398
2	Group (0 = TLD, 1 = LBD)	−0.38	.001	37.314	<.001	.726	.527	.513
	Language comprehension composite	0.44	<.001					
3	Group (0 = TLD, 1 = LBD)	−0.31	.026	24.847	<.001	.728	.530	.509
	Language comprehension composite	0.46	<.001					
	Decoding composite	−0.08	.488					
4	Group (0 = TLD, 1 = LBD)	−0.29	.038	19.315	<.001	.737	.543	.515
	Language comprehension composite	0.36	.007					
	Decoding composite	0.01	.954					
	Phonological processing composite	0.19	.184					

Note. Group is coded as 0 = typical language development (TLD), 1 = language disorder (LBD).

Predictors include language comprehension composite, decoding composite, and phonological processing composite.

Dependent variable: *questions_total*.

β = *standardized regression coefficient*.

6. DISCUSSION

Understanding how children comprehend text remains a central concern in research on written language and literacy development. Reading comprehension is not a unitary skill, but a dynamic process shaped by multiple language, cognitive, and experiential factors. These influences may combine in different ways across children, giving rise to diverse language-based profiles and developmental pathways. The purpose of this dissertation was to examine the mechanisms that support written discourse comprehension (through anaphora resolution) in school-aged children with or without language-based disorders. Particular attention was given to the role of psycholinguistic word properties (imageability and frequency), broader (meta)language skills, and inferential processes. By adopting contemporary experimental and analytical approaches, this research aimed to situate its findings within existing theoretical

frameworks (Simple View of Reading, Gough & Tunmer, 1986; Lexical Quality Hypothesis, Perfetti & Hart, 2002) and to clarify how individual differences in language functioning contribute to variability in reading comprehension outcomes.

In this study, two levels of frequency and imageability (high and low) were manipulated using four passages across four stories, focusing on the following areas of interest: antecedent, anaphora, and the two words before and after the area of interest. Children silently read the story on a computer display while their eye movements were recorded, followed by an (open-ended) question-answering task (literal and inferential). Using both online (real-time; eye-tracking) and offline experimental procedures (question answering), three interrelated analyses were conducted. Three core aspects of comprehension were targeted: (a) real-time anaphora processing (Research Question 1); (b) accuracy in literal and inferential comprehension (Research Question 2); and (c) the contribution of various (meta)linguistic factors to reading comprehension (Research Question 3). Croatian-speaking children (5th and 6th Grade) were initially classified as having typical language development or language-based disorders (DLD, dyslexia, comorbidity of DLD and Dyslexia), primarily in line with Bishop & Snowling's (2004) Two-dimensional framework, further refined with Catts et al.'s (2006) Fully Distinct Hypothesis. Based on the SVR model (Reading Component Model; Gough & Tunmer, 1986), participants were further assigned to poor oral language, poor decoders, or mixed language difficulties groups, according to their reading-based skills.

6.1 Lexical Quality and Language-based profiles as determinants of anaphora resolution in reading comprehension

The first research objective was to examine the role of psycholinguistic word properties - specifically frequency and imageability - in the real-time processing of anaphoric relations in children, across different language profiles (Gough & Tunmer, 1986; Bishop & Snowling, 2004; Catts et al., 2005): a) TLD vs. LBD; b) TLD, PD, POL, MD, using eye-tracking methodology. The following eye-tracking measures were used as dependent variables: first fixation duration and first pass duration as early-phase measures (initial lexical access), dwell time and number of fixations as late-phase measures (integration, interpretation), and number of revisits and regression out probability (re-analysis) as extended-phase measures. Data was analysed applying repeated measures mixed ANOVA. The effects of frequency and imageability, as well as their interaction, were first examined across groups of children classified as having typical language development (TLD; N = 37) and language- based

disorders (LBD; N = 33). The subsequent analysis was conducted across specific language-based disorder profiles, based on the classification frameworks proposed by Catts et al. (2005) and the Simple View of Reading (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002), in accordance with the hypotheses posed under the first research question (poor oral language group, POL = 10; poor decoders group, PD = 16; mixed disorder group, MD = 11; TLD = 37). The discussion is structured in two parts: it first addresses the results for the TLD and LBD groups, in line with the stated hypotheses, according to areas of interest: antecedent, anaphora, words preceding the AOIs, followed by the analysis of the specific language-based disorder profiles (TLD vs. PD vs. POL vs. MD).

6.1.1 Word frequency and imageability effects in TLD and LBD

The results are discussed in relation to the first research question and its accompanying first hypotheses (subhypotheses H1, H1.1 and H1.2), which guided the examination of the effects of psycholinguistic properties on the processing of anaphoric relations in the typical language development (TLD) group and the language-based disorder (LBD) group. It was generally assumed that word frequency and imageability would influence processing, with low frequency and low imageability words eliciting longer processing times, longer fixation durations, and a higher number of regressions compared to high frequency and high imageability words. These hypotheses are grounded in the Lexical Quality Hypothesis (LQH; Perfetti, 2007). The LQH posits that high-quality lexical representations, encompassing accurate phonological, semantic, and orthographic encoding, facilitate rapid and automatic word recognition and meaning retrieval, thereby supporting higher-level processes such as anaphora resolution. Psycholinguistic properties such as word frequency and imageability contribute to the accessibility and integration of lexical information within a broader discourse context (van Berkum, 2009).

The discussion regarding the first research question is organized by addressing the general effects and interactions of the psycholinguistic word properties (imageability and frequency) with respect to the areas of interest (AOIs), followed by the interpretation of the analysis of the group-level differences (TLD vs. LBD).

Empirical findings from other studies suggest that low-frequency words require greater processing effort due to slower and less automatised retrieval from the mental lexicon (Rayner et al., 2004). In anaphoric contexts, where integration with a prior referent is required, low lexical frequency may increase processing demands, particularly in developing

readers or those with weaker reading/ language skills, as reflected in longer fixation durations and increased regressions (Nation, 2014; Tiffin-Richards & Schroeder, 2015). High-imageability words are generally more easily recognised and integrated (Sadoski & Paivio, 2004), and previous studies have shown that imageability can facilitate both early and late phases of language processing (Juhasz & Rayner, 2003).

The following findings regarding psycholinguistic effects in the target AOIs are reported for the sample as a whole, given that no significant group interactions were observed for imageability, and group-related modulations of frequency effects are addressed separately below. In the present study, within the antecedent region, where the word had already been anchored in the discourse, the results indicate that higher imageability and higher frequency each showed overall facilitatory effects, on the sample-level, though their specific contributions were not uniform across processing phases and were in certain conditions modulated by their mutual interaction, or by interaction between frequency and group, as elaborated in later sections. Imageability effects were observed primarily in early measures, such as first fixation duration and first pass duration, suggesting faster and easier processing in conditions involving highly imageable antecedents during initial lexical access. This finding is consistent with previous studies reporting early imageability effects (Schilling et al., 1998; Juhasz & Rayner, 2003, 2006; Kliegl et al., 2004; Jones et al., 2008; Dambacher et al., 2013), and may reflect the rapid activation of conceptual features associated with highly imageable words. However, this interpretation requires caution, as the effect of imageability was not uniform across conditions. Specifically, a significant interaction between imageability and frequency in first fixation duration revealed that imageability performed its influence selectively: high imageability antecedents elicited longer, not shorter, first fixations when paired with low frequency, producing the longest fixation durations of all conditions. Conversely, when antecedents were high in frequency, imageability had no significant effect, and for low imageability antecedents, variations in frequency had minimal impact. This pattern could be interpreted as a paradoxical effect of imageability under conditions of low frequency. Rather than compensating for the processing demands associated with low frequency antecedents, high imageability may introduce an additional processing challenge by activating a rich conceptual representation that cannot be efficiently anchored to a weak or less automatised lexical form (due to low frequency). This interpretation would be consistent with the Dual Coding Theory (Paivio, 1971, 1990) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002), which distinguish between conceptual richness and the automaticity

of lexical access: a vivid mental image may be readily activated, yet full lexical integration may still be slow when the orthographic-phonological-semantic connections are insufficiently established. The idea that early imageability effects could depend on frequency is consistent with the evidence from other studies reporting that imageability effects can be modulated by co-occurring psycholinguistic properties (Schilling et al., 1998; Juhasz & Rayner, 2003, 2006; Kliegl et al., 2004; Jones et al., 2008; Dambacher et al., 2013). According to the Dual-Route Cascaded Model (Coltheart et al., 2001), conceptual features can become available early in processing, even before full lexical identification and integration; however, the present findings suggest that the utility of this early conceptual activation is contingent upon the strength and the quality of the underlying lexical representation. Importantly, although the observed pattern suggests that the contribution of imageability may depend on frequency, the available data do not support strong conclusions regarding specific combinations of these variables. Rather than indicating that one factor systematically outweighs the other, the interaction is better understood as reflecting differences in how these properties influence processing at different phases.

In the first pass duration measure, imageability and frequency operated as independent facilitators: both high imageability and high frequency were associated with shorter first pass durations, with no significant interaction between the two properties. This suggests that, at this slightly later phase of early processing, which captures not only initial fixation but the entire first sweep through the AOI, the ‘paradoxical’ interaction observed in first fixation duration has resolved, and both properties contribute independently to the ease of the initial reading, in the antecedent AOI.

Frequency effects were consistently observed across multiple eye-movement measures. Low frequency antecedents were associated with longer first fixation duration, longer first pass duration, longer dwell time, and a higher number of fixations within the AOI. These findings align with a substantial body of literature demonstrating that word frequency influences both early and later phases of processing (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Livversedge et al., 1998; Rayner, 1998; Kliegl et al., 2004; Ashby et al., 2005; Jones et al., 2008; Joseph et al., 2007, 2013; Rayner et al., 2011; Blythe & Joseph, 2011; Dambacher et al., 2013; Staub, 2015; Meziere et al., 2023). For example, Ashby et al. (2005) and Rayner et al. (2011) report shorter first fixation durations for high frequency words, explaining that this measure reflects early processing, whereas low frequency words prolong the duration of the first fixation on the AOI. Livversedge et al. (1998) identify first fixation duration as the

earliest marker of psycholinguistic effects, and Staub (2015) adds that it is the measure most sensitive to manipulations of psycholinguistic word properties. Tiffin-Richards and Schroeder (2015) note that low frequency words lead to slower decoding, greater cognitive load, and increased integration difficulty, which are the patterns observed predominantly in children rather than adults. Similarly, Kliegl et al. (2004) and Jones et al. (2008) found longer total fixation time within the AOI when processing low frequency words, while Kuperman & Van Dyke and Joseph et al. (2013) highlight that this effect on first pass duration is particularly pronounced in children.

Thus, while both variables contribute to processing, their effects differ in timing and functional role. Imageability operates primarily in the earliest phase of processing (first fixation duration) but only in interaction with frequency, and its independent contribution is evident in first pass duration. Critically, imageability did not significantly influence dwell time or any later measures in the antecedent region, indicating that its role is temporally limited to early lexical access. Frequency, by contrast, showed a robust and consistent association across all phases, from first fixation duration through dwell time and number of fixations, reflecting its role in both initial word recognition and sustained processing effort. This distinction is consistent with the view that frequency influences all levels of a word's representation: lexeme, lemma, and concept, whereas imageability exerts its primary effect at the conceptual level (Levelt, 1992; Kittredge et al., 2008), and with the LQH's account of frequency as a determinant of lexical representation quality and automaticity (Perfetti & Hart, 2002). Furthermore, effects observed at the level of the antecedent reflect lexical processing rather than anaphoric integration, as the anaphoric expression has not yet been encountered. This distinction is essential for interpreting early eye-movement measures.

In the anaphora region, effects of psycholinguistic properties emerged primarily in later phases of processing. No significant effects were observed in the earliest measure, the first fixation duration, for any psycholinguistic property or group, a finding that is theoretically expected given the interpretive dependency of anaphors on their antecedents: the anaphor does not carry its referential meaning independently, and initial fixation therefore does not trigger intensive lexical analysis in the same way as a content word encountered for the first time. In the first pass duration measure, a significant interaction between frequency and imageability was observed; however, post hoc comparisons did not reveal reliable differences between specific conditions, and the nature of this interaction therefore remains insufficiently specified. Nevertheless, descriptive data point to a potential trend: processing appeared

fastest when frequency and imageability were aligned (high-frequency, high-imageability or low-frequency, low-imageability), whereas mismatched combinations appeared to require greater processing effort. Although this pattern did not reach statistical significance, it is theoretically informative. One possible explanation is that aligned conditions either enable rapid activation of a well-established lexical representation (high-frequency, high-imageability) or do not trigger additional processing effort when both cues indicate a weak representation (low-frequency, low-imageability), resulting in relatively brief fixation times. Mismatched conditions may instead induce processing conflict, where one property facilitates access while the other constrains it, prompting additional verification and longer gaze durations. Such an interpretation is broadly consistent with models of lexical quality (Perfetti & Hart, 2002; Perfetti & Stafura, 2014); however, given the absence of statistically reliable post hoc differences, this pattern should be regarded as tentative.

In later measures, frequency emerged as the dominant predictor: low frequency anaphors elicited longer dwell times, more fixations, and more revisits than high frequency anaphors. A significant imageability $\times$ frequency interaction was also observed in dwell time, suggesting that the difference between high and low frequency anaphors was more pronounced in conditions of high imageability, while in low imageability conditions the frequency effect was attenuated. This pattern mirrors, at a later processing phase, the interaction identified in the antecedent region, and further supports the interpretation that the combination of high imageability and low frequency consistently places the greatest demands on lexical processing across regions and phases. Any interpretation of this interaction should remain cautious, however, as it emerged in only one late-phase measure and was not supported by statistically significant post hoc comparisons in all conditions.

According to models of word recognition (Coltheart et al., 2001; Grainger & Holcomb, 2009) and lexical representation quality (Perfetti & Hart, 2002), high frequency words have lower activation thresholds within the lexical network, resulting in faster recognition and shorter fixation durations (Jurafsky, 2003; Balota et al., 2004; Ferrand et al., 2011; Joseph et al., 2013; Tiffin-Richards & Schroeder, 2015). Operating at multiple levels of linguistic processing, frequency may therefore exert strong influence both in early word recognition and in the integration of referential information during anaphora resolution. The finding that frequency effects were evident in the anaphor region can also be interpreted in light of findings by Masterson et al. (2008), who demonstrated that frequency effects vary depending on the syntactic function of a word and the age group. It is possible that, when the anaphora is

expressed through a null form or verb morphology, processing relies more heavily on automatic lexical mechanisms, where frequency acts as either a facilitating or inhibiting factor independently or in interaction with imageability. Low frequency in such cases likely delays the automatic activation of the referent, especially when the linguistic structure is demanding and semantically reduced.

In the region preceding the target AOIs, the early processing measure revealed a different pattern from that observed at the antecedent: imageability, rather than frequency, was the primary driver of first fixation duration, with low imageability words eliciting longer initial fixations regardless of frequency. Frequency reached only marginal significance in this earliest measure. This dissociation suggests that for contextual words not serving a direct referential function, the ease of forming a mental image is the primary determinant of initial lexical access. In later measures, however, frequency again became the dominant predictor, and its effect was modulated by imageability: low frequency words elicited longer dwell times and more fixations, but only in conditions of high imageability, while for low imageability words, variations in frequency had minimal impact. A similar pattern emerged in the number of revisits measure, where the frequency effect was likewise confined to high imageability conditions. These consistent interactions across late and extended measures suggest that the combination of high imageability and low frequency represents a particularly demanding processing, potentially reflecting the same mismatch between conceptual richness and lexical automaticity identified in the antecedent region. The absence of frequency effects for low imageability words across this region may indicate that processing of abstract, low imageability words relies more heavily on contextual cues, rendering distributional frequency less influential as a facilitating factor. These patterns may be partly attributed to spill-over effects from adjacent regions (Juhasz & Rayner, 2003).

Having addressed the general effects of psycholinguistic word properties, the following section turns to group-level differences in processing between children with TLD and those with LBD.

Previous studies show that difficulties of children with language-based disorders are also evident across a continuum of phonological, lexico-semantic, and syntactic levels of processing (Catts et al., 2005, 2006; Nation et al., 2001; Adlof & Catts, 2015). With respect to group differences in our study, the LBD group showed greater processing difficulty than the TLD group across multiple measures and processing phases in the antecedent region.

Group differences were evident from the first pass duration onward: children with LBD showed longer first pass durations than children with TLD, and this difference was further modulated by a significant frequency $\times$ group interaction, indicating that low frequency antecedents produced disproportionately longer first pass durations in the LBD group relative to the TLD group. In the later processing phase, children with LBD showed longer dwell times and made more fixations overall, with dwell time again showing a significant frequency $\times$ group interaction: the processing advantage for high-frequency over low-frequency antecedents was more pronounced in the LBD group. In the extended processing phase, children with LBD made significantly more revisits to the antecedent region than children with TLD, suggesting greater difficulty in integrating referential information and a need for compensatory reanalysis. Similarly, Van der Schoot (2008) reports that weaker readers produce more non-strategic revisits and require longer processing times (Hyönä & Olson, 1995; Blythe & Joseph, 2011). Taken together, these findings indicate that group differences in antecedent processing are not limited to a single phase or measure, but accumulate across the processing phases, consistent with the interpretation that children with LBD show reduced automaticity of lexical access and less efficient referential integration, in contrast to the TLD group (Rayner, 1986; Ashby et al., 2005; Joseph et al., 2013; Staub, 2015).

An interesting and unexpected pattern emerged in the regression out probability measure. Based on previous findings, one would expect low frequency or low imageability conditions to elicit a higher likelihood of regressions outside the AOI (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Kuperman & van Dyke, 2013). However, in both the antecedent and anaphor regions, the opposite pattern was observed: participants were more likely to make a regression out of the AOI following fixation on high frequency, rather than low frequency, words. Furthermore, in the antecedent region, children with TLD showed a significantly greater regression out probability than children with LBD, a finding that is likewise counterintuitive if regressions are interpreted solely as markers of comprehension difficulty. It is important to note that regressions vary in duration and function, and only long regressions necessarily reflect comprehension breakdown (Rayner, 1998). Murray and Kennedy (1988) distinguish strategic from non-strategic regressions, while Wotschack and Kliegl (2011) report that readers strategically increase regressions when they anticipate higher task demands. Cook & Wei (2019) and Rayner et al. (2009) further state that textual inconsistencies increase regression out probability, and Laubrock et al. (2006) emphasise that regressions serve to check cohesion. It is therefore possible that in high frequency conditions

(but only for antecedents), children with TLD, having rapidly identified the word, accessed its meaning, and attempted to integrate it into the discourse context, made regressions outside the AOI to retrieve additional contextual information and strengthen referential cohesion. In this sense, the higher regression out probability in TLD, particularly for high frequency words, may reflect active comprehension monitoring rather than processing failure. By contrast, children with LBD, whose processing of the antecedent itself appears more effortful and prolonged, may lack the residual processing capacity to engage in such strategic backward scanning. Low frequency conditions may not elicit the same pattern because these words are not processed rapidly enough to prompt a cohesion-checking regression before the eyes move on. These interpretations align with developmental patterns in eye-movement control (Schroeder et al., 2015) but should be treated with caution, as regression out probability alone cannot distinguish between strategic and comprehension-driven regressions, and measures such as regression path duration or second-pass reading time, which were not included in the present study, would be needed for a more targeted interpretation.

In the anaphora region, by contrast, group differences were more circumscribed: a significant group effect emerged only in the dwell time measure, with children with LBD showing longer total processing durations. No significant group differences were observed in the number of fixations, number of revisits, or regression out probability in the anaphora region, suggesting that group-level processing differences during anaphoric resolution are largely confined to total processing time rather than reflecting qualitatively distinct reading strategies.

Within the framework of the Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti & Stafura, 2014), these findings suggest that differences in processing efficiency between TLD and LBD are likely related to the quality and stability of lexical representations. The frequency $\times$ group interactions, observed in both first pass duration and dwell time in the antecedent region, indicate that children with LBD show differential sensitivity to word frequency, particularly for low frequency words, reflecting reduced automaticity in lexical access during referential processing. Imageability, by contrast, did not interact with group membership in any measure across any region, suggesting that both groups activate conceptual features of words in a comparable manner once lexical access has been initiated. This finding implies that what differentiates children with TLD from those with LBD is not the richness or availability of conceptual representations, but the speed and reliability with which orthographic-phonological-semantic connections are activated - a process strongly

shaped by the quality and automatisedness of lexical representations (Perfetti & Hart, 2002; Perfetti & Stafura, 2014; Levelt, 1992). Although high imageability generally facilitates early conceptual activation, it was not sufficient to compensate for the reduced lexical access efficiency observed in the LBD group, nor did it differentiate the two groups' processing trajectories. In this sense, interindividual differences in anaphoric processing appear to be driven primarily by variability in the automaticity of lexical access rather than by differences in conceptual representation per se.

6.1.2 Word frequency and imageability effects across different language-based disorder profiles

The results are discussed in relation to the first research question and its accompanying second hypotheses (subhypotheses H2, H2.1, and H2.2), which guided the examination of the effects of psycholinguistic properties on the processing of anaphoric relations across the typical language development (TLD) group and groups with language-based disorders. The discussion is organized by addressing the general effects and interactions of the psycholinguistic word properties (imageability and frequency) at the group-level (TLD vs. groups with language-based disorders).

Within the framework of the Lexical Quality Hypothesis (Perfetti & Hart, 2002), the same analysis was conducted across four groups defined in line with the classification proposed by Catts et al. (2005), and interpreted within the context of the Simple View of Reading (SVR; Gough & Tunmer, 1986): typical language development, poor decoders, poor oral language, and the mixed group. It was hypothesised that children with language-based disorders would experience greater processing demands in tasks involving lexically and semantically demanding processes such as anaphora resolution, particularly when processing words with low levels of imageability and frequency, compared to children with typical language development (H2).

It was further assumed that children with weaker semantic knowledge and less developed vocabulary would have difficulty constructing coherent discourse representations and integrating word meaning into context (Nation & Snowling, 1998; Catts et al., 2006). Accordingly, the effect of word imageability on anaphoric processing was expected to be more pronounced in children with poor oral language (POL) than in those with typical development (H2.1), reflecting their reliance on semantic richness to support referential integration. When faced with semantically sparse, low imageability words, their ability to resolve referential links is further compromised, resulting in greater processing effort. In

contrast, for children with decoding difficulties (PD), it was assumed that reduced automatization of word recognition would limit the availability of cognitive resources for higher-order processing (Shankweiler et al., 1996), leading to a stronger influence of word frequency on processing, due to less efficient access to orthographic and phonological representations during lexical processing (H2.2).

The findings provide only partial support for these hypotheses. Across AOIs (regions) and processing phases, word frequency consistently emerged as the most robust factor, whereas the effects of imageability were weaker and less stable. Crucially, the expected differential effects of these psycholinguistic properties across specific language profiles were not consistently supported by the statistical analyses, although meaningful group-specific patterns did emerge in several measures and are discussed below.

In the antecedent region, which is critical for establishing referential representations, early processing phases showed sensitivity to psycholinguistic properties, with group-related variation emerging selectively. In the First Pass duration measure, a significant interaction between word frequency and group was observed. Post hoc comparisons revealed that low frequency antecedents elicited significantly longer first pass durations than high frequency antecedents specifically in the PD and MD groups, while this contrast was not statistically confirmed in the TLD and POL groups. This suggests that, rather than reflecting global group difference in processing speed, the reduced automaticity of decoding places particular demands on lexical access precisely for low frequency words. Within the LQH framework (Perfetti & Hart, 2002), children with decoding difficulties already operate with weaker phonological-orthographic representations; when these weaker representations are further influenced by words with low distributional frequency, which themselves have less robust lexical entries, the cumulative effect on processing becomes pronounced. In other words, the combination of a weaker decoding system and a weaker lexical entry creates a compounding burden that is evident in early integrative processing. This interpretation is consistent with previous findings indicating that decoding difficulties disproportionately affect the processing of low frequency words in poor readers (Hyönä & Olson, 1995; Perfetti & Hart, 2002; Catts et al., 2005; Blythe & Joseph, 2011; Joseph et al., 2013). The absence of this interaction in the TLD and POL groups suggests that their phonological-orthographic representations are sufficiently robust to support relatively efficient first-pass processing even for low frequency antecedents. Thus, the early processing stage reflects a contribution of psycholinguistic word

properties, while group differences become even more clearly differentiated in the later stages of processing.

In the later phase of processing, high frequency antecedents were associated with shorter total fixation and saccade durations (dwell time) compared to low frequency antecedents across most groups; however, the pattern was not uniform. Specifically, the TLD, PD, and MD groups all showed a significant processing advantage for high frequency over low frequency antecedents in dwell time, whereas TLDs had overall shorter dwell time compared to PD and MD groups. This contrast was not statistically reliable in the POL group. The POL group showed a qualitatively similar pattern to the TLD group, but with less consistent sensitivity to frequency. This finding identifies the POL group as qualitatively distinct from the others in the sense that this group specifically fails to benefit from the facilitatory effect of high word frequency, rather than showing globally slower processing. This profile is consistent with the characterisation of poor comprehenders, who typically show intact decoding but weaknesses in higher-level language skills, including semantic processing and discourse integration (Cain et al., 2001; Oakhill & Cain, 2012; Catts et al., 2005). Within the LQH framework, it suggests that children with poor oral language may have lexical representations that are insufficiently specified at the semantic-conceptual level, such that even high frequency words do not afford the processing advantage they typically provide. Their phonological-orthographic representations may support accurate word recognition, but the semantic and syntactic dimensions of lexical quality may be less fully developed, reducing the benefit of frequency (which activate processes both on the lexeme and lemma level) as a facilitating factor.

The number of fixations measure provides additional nuance. At the group level, TLD and POL made fewer fixations than the PD group, with MD occupying an intermediate position. Critically, a significant interaction between word frequency and group revealed that only the TLD group showed a statistically reliable difference in number of fixations between low and high frequency conditions, making more fixations on low frequency antecedents. The LBD subgroups did not demonstrate this pattern in a statistically confirmed manner; moreover, a descriptive tendency in the MD group suggested a potential reversal, with slightly more fixations for high frequency antecedents, though this did not reach significance. Therefore, it seems that TLD children appear to modulate their fixation behaviour in accordance with the lexical demands of individual words, allocating more processing resources to less frequent items. This adaptive, frequency-sensitive strategy may reflect well-developed metacognitive

reading skills and efficient allocation of attentional resources. The absence of this modulation in LBD subgroups, and its potential reversal in MD, may indicate that atypical attentional and lexical processing patterns in these children reduce the flexibility of eye-movement behaviour during referential processing. These findings indicate that late phase antecedent processing is sensitive to differences in the efficiency of underlying linguistic systems, particularly those related to decoding and lexical access. According to the classification proposed by Catts et al. (2005), children with TLD and POL are characterised by relatively preserved decoding skills, whereas the PD and, to some extent, the MD groups are marked by phonological or combined language weaknesses. This distinction is consistent with the Reading Component Model and the Simple View of Reading (Gough & Tunmer, 1986; Catts & Kamhi, 1999; Joshi & Aaron, 2000; Tunmer & Chapman, 2007; Tunmer & Greaney, 2009; Torppa et al., 2010), which emphasises the interplay between decoding and language comprehension processes.

As already mentioned, within this framework, the POL group most closely corresponds to the profile of poor comprehenders (intact decoding, weaknesses in higher-level language skills). As noted by Catts et al. (2005), only a subset of these children meets the formal diagnostic criteria for developmental language disorder, which may explain why the POL group does not show marked deviations in earlier processing stages. Their phonological-orthographic representations appear sufficiently robust to support efficient word recognition (Perfetti & Hart, 2002), while their difficulties become apparent in later processing phases that require semantic integration and coherence building - specifically, in the failure to benefit from lexical frequency. This interpretation is consistent with previous findings showing that poor comprehenders experience difficulties in inference generation and discourse-level processing despite intact decoding (Cain et al., 2001; Oakhill & Cain, 2012).

In contrast, the PD group most closely reflects the profile associated with dyslexia (Lyon et al., 2003; Catts & Kamhi, 2005), while the MD group reflects the comorbidity, the co-occurrence of decoding and broader language difficulties (Snowling et al., 2020). Children with dyslexia often experience difficulties in establishing stable phonological representations, which constrain efficient lexical access and slow the transition to higher-level processing (Nation et al., 2001, 2010; Perfetti & Hart, 2001). These difficulties are particularly evident for low frequency words, which carry inherently weaker lexical entries and therefore place additional demands on an already less efficient decoding system. Although compensatory strategies such as contextual prediction may emerge, they are often insufficient when early decoding processes are not automatised (Cain et al., 2001; Oakhill & Cain, 2012). This helps

explain why PD and MD children show frequency-related processing difficulties even in early integrative measures (e.g. first pass), and why PD children produce the highest overall number of fixations across all conditions.

In the anaphora region, early processing measures did not show stable main effects of either frequency or imageability, nor clear group differences, which is consistent with the interpretive dependency of anaphors on their antecedents. A significant interaction between imageability and frequency was observed in first pass duration, but post hoc comparisons did not reveal reliable condition-specific differences, and the pattern should therefore be viewed as tentative. No systematic modulation by group membership was found in early measures.

In later stages of anaphora processing, frequency again emerged as a reliable factor, with low-frequency words associated with increased processing effort reflected in longer dwell times and more fixations. At the group level, the dwell time measure revealed a significant main effect of group, with the MD group showing significantly longer dwell times on anaphors than the TLD group, while other group contrasts did not reach significance. A significant imageability $\times$ frequency interaction in dwell time further indicated that the frequency effect was most pronounced for high-imageability anaphors, whereas in low-imageability conditions the frequency effect was attenuated, a pattern broadly consistent with findings across other AOIs in this study.

A more differentiated group-specific pattern emerged in the number of fixations measure. A significant interaction between word frequency and group revealed that the frequency effect was statistically reliable only within the MD group, where low frequency anaphors elicited a higher number of fixations than high frequency anaphors. This suggests that children with mixed language difficulties are particularly sensitive to frequency-related processing demands at the level of anaphoric resolution. This pattern also extends their frequency sensitivity observed in antecedent processing into the anaphoric region, and that may reflect the compound burden of co-occurring decoding and language comprehension difficulties on sustained lexical processing during referential integration (Blythe & Joseph, 2011; Snowling et al., 2020).

Low frequency anaphors were also associated with increased rereading behaviour as reflected in the number of revisits measure, indicating the need for additional processing when lexical access is less efficient. A main effect of group was observed, though post hoc comparisons did not identify significant differences between specific groups, suggesting that the broader

tendency for increased rereading in language-based disorder is distributed across profiles without a single subgroup driving the effect. When phonological-orthographic representations are not accessed rapidly enough, difficulties in establishing referential cohesion arise, prompting reanalysis (Rayner et al., 2004). Frequency also emerged as a reliable predictor in the regression out probability measure, with low frequency anaphors associated with a higher probability of regressions outside the AOI across all groups. This direction of effect (greater regressions for low-frequency items) is consistent with typical expectations and suggests that when lexical access is effortful, readers are more likely to seek additional contextual support from preceding text. No significant group differences were observed in this measure, indicating that the fundamental mechanism of frequency-driven regressions is preserved across all language profiles.

Regarding the eye-movement patterns observed in the region preceding the target AOIs, frequency effects emerged robustly across later and extended processing phases, modulated by imageability in a pattern consistent with those observed at the antecedent. Group differences in this region were diffuse: TLD showed shorter dwell times than PD and MD, while POL again did not differ significantly from other groups. Furthermore, post hoc analyses did not identify specific subgroup contrasts in number of fixations or revisits, suggesting that for contextually supporting words not carrying referential function, processing difficulties may be more globally distributed across LBD profiles.

Overall, the findings support the central role of lexical quality in reading comprehension (Perfetti & Hart, 2002; Perfetti & Stafura, 2014), showing that word frequency, as an index of lexical representation strength, consistently facilitates processing across AOIs and phases. However, the results do not point to a uniform effect of language-based disorders on psycholinguistic processing. Rather, they suggest meaningful specificity across profiles: decoding-related disorders (PD and, particularly, MD) show heightened sensitivity to word frequency in integrative processing phases, consistent with the compounding effect of weak decoding on already weak lexical entries for low frequency words. The POL profile shows a distinct pattern of preserved processing speed alongside a specific failure to benefit from lexical frequency, pointing to a deficit in the semantic-conceptual dimension of lexical quality rather than in phonological-orthographic automaticity. These dissociations across profiles suggest that the type of language-based disorders modulates not only how efficiently anaphoric processing proceeds, but also which aspects of lexical quality - phonological,

orthographic, or semantic - are most critically implicated in the breakdown of referential integration during reading comprehension.

6.2 The effect of question type and specific language difficulties profiles on reading comprehension

The results are discussed in relation to the second research question and its accompanying third and fourth hypotheses (subhypotheses H3, H4, H4.1 and H4.2), which guided the examination of the effects of the question type and specific language-based disorder profile on reading comprehension.

The second research question concerned the analysis of oral responses to two literal and two inferential questions, which were read aloud by the examiner after participants had read the same four stories ($N = 4$) on a computer screen. The stories and corresponding questions (total $N = 8$) remained visible on the screen throughout the entire assessment. The questions were open-ended (Cain & Oakhill, 2006; Graesser et al., 2010) and directly related to the anaphoric relations in the text. Literal questions required participants to recognise and retrieve explicitly stated information from the text (e.g., Who sang?), whereas inferential questions required deeper processing and the integration of implicitly conveyed information (e.g., Why was the boy angry?). The analysis was conducted first at the level of two groups (TLD and LBD), and then across the four groups defined by specific language profiles (TLD, PD, POL, MD).

According to classical models of discourse comprehension (Kintsch & van Dijk, 1978; Johnson-Laird, 1980), readers must construct a situation model that links information in the text and enables the generation of both explicit and implicit inferences. Comprehension involves the semantic and syntactic integration of information, as well as narrative comprehension (Daneman & Merikle, 1996; Hagtvet, 2003; Tanenhaus et al., 1995). Key processes include connecting ideas, evaluating coherence, and monitoring one's own understanding (Graesser et al., 1994; Oakhill et al., 2003; Nation, 2005). Linguistic markers of cohesion, particularly anaphoric links, further guide meaning integration, consistent with the Lexical Quality Hypothesis (Perfetti, 2007) and the Reading System Framework (Perfetti & Stafura, 2014), both of which emphasise the importance of high-quality lexical representations for coherent discourse processing. Once reading becomes automatised, these skills increasingly determine reading comprehension (Geva & Farnia, 2012; Cain, 2015; LARRC, 2015; Lervåg et al., 2017). Inference generation is a central mechanism of coherence (Dahl et al., 2021) and one of the strongest predictors of reading comprehension

from early childhood onwards (Kendeou et al., 2009; Oakhill & Cain, 2012). Thus, inference-making is fundamental to both language and reading comprehension (SVR; Gough & Tunmer, 1986) and relies on efficient access to phonological, syntactic, and semantic information as well as their integration into a coherent mental representation (Perfetti & Hart, 2002; Perfetti & Stafura, 2014). From this perspective, inferential comprehension is expected to be more demanding for all language profiles. Literal comprehension involves reproducing surface-level and text-based information, whereas inferential comprehension requires constructing a situation model, linking information across different parts of the text, and activating real-world knowledge (Kintsch & van Dijk, 1978; Perfetti & Stafura, 2014).

Previous studies consistently show that inferential questions are more difficult for children across different language profiles, but the same pattern was also evident for the children with typical language development (Cain & Oakhill, 2006). However, it has also been assumed that the effects of question type would depend on language profile, in line with multi-phase models of reading (Perfetti & Stafura, 2014) and empirical findings on distinct patterns of difficulty (Catts et al., 2006; Kendeou et al., 2009; Oakhill & Cain, 2012). Children with weak decoding skills (PD) often redirect cognitive resources to phonological processing, which hinders the construction of a coherent mental representation (Shankweiler et al., 1996; Perfetti & Hart, 2002). They are therefore expected to experience difficulties with both literal and inferential questions. Children with weak language comprehension (POL) have intact decoding but impaired semantic and discourse skills (Nation, 2005; Bishop & Snowling, 2004; Catts et al., 2006). Accordingly, literal questions are expected to be relatively accessible, whereas inferential questions should pose greater difficulty due to the demands of semantic integration and flexible vocabulary use. Children with mixed language difficulties (MD) show global comprehension impairments due to combined phonological, working-memory, and semantic weaknesses (Catts et al., 2006). As a result, difficulties are expected across all question types, with inferential questions being particularly challenging because they require multi-phase integration of knowledge.

In the present study, the results addressing the second research question revealed two key findings:

- a) all participants, regardless of language profile, achieved lower scores on inferential questions than on literal ones, a pattern that was consistent across groups and did not differ in magnitude between TLD and LBD children;
- b) children with language-based disorders performed more poorly overall.

The results revealed a statistically significant main effect of question type, with all participants, regardless of language profile, achieving higher scores on literal than on inferential questions. This finding is consistent with theoretical models of reading comprehension, such as the Reading Systems Framework (Perfetti & Stafura, 2014), and the situation model framework (Kintsch & van Dijk, 1978), which emphasize that inferential questions require a higher level of information integration and cognitive processing. Literal questions that require linking adjacent sentences and referents are processed more easily and quickly during reading (McMaster et al., 2014), whereas questions that require integrating broader text segments or predicting events are more cognitively demanding (Snow, 2002; Yang et al., 2007). Importantly, ceiling effects were not observed even in the TLD group, suggesting that higher inferential integration skills are still developing. This indicates that the ability to construct and maintain a coherent situation model across multiple sentences or paragraphs continues to mature beyond basic decoding and literal comprehension.

In addition, a statistically significant main effect of group was observed, with children with typical language development outperforming those with language-based disorders on all types of questions. Within the clinical sample, the poor decoders performed better than the mixed group but did not outperform the POL group. This suggests that the combined difficulties characteristic of the mixed profile have the strongest negative impact on comprehension, whereas children with TLD consistently show the best performance across all question types. According to Graesser and colleagues (2004), skilled readers compensate for missing information by drawing on knowledge of story structure to maintain coherence, while weaker readers struggle to establish such links. Successful readers dynamically integrate textual information with background knowledge, efficiently connect ideas, and suppress irrelevant details (Graesser et al., 1994; Côté et al., 1998), whereas weaker comprehenders experience difficulties in monitoring meaning and linking concepts. These differences are particularly evident in the ability to generate inferences. Skilled readers spontaneously produce causal and goal-related inferences (Graesser et al., 1994; Côté et al., 1998), while weaker comprehenders often rely on superficial or irrelevant associations. Oakhill's research (1982, 1984) showed that although TLD and LBD groups may perform similarly on literal recall tasks, children with stronger comprehension outperform weaker comprehenders on both literal and inferential questions, reflecting more efficient integration of textual information with background knowledge. Subsequent work has confirmed that less skilled readers struggle to form both text-connecting and gap-filling inferences, even when

given opportunities to reread (Cain & Oakhill, 1999), indicating reduced spontaneity in inference-making and poorer comprehension monitoring rather than memory limitations. However, the most relevant result showed that the interaction between question type and language profile was not statistically significant, in either the two-group comparison or the four-group analysis. Thus, both literal and inferential questions were proportionally demanding for all clinical groups, while children with TLD consistently outperformed them on both question types. This finding contradicts the initial expectation that question type would interact with language profile. Despite previous research (Catts et al., 2005; Oakhill & Cain, 2012; García & Cain, 2014) showing that children with weaker vocabulary, syntax, and higher-level comprehension skills (POL, poor comprehenders), as well as those with weaker general reading skills (PD), would —though for different reasons— experience greater difficulty with inferential or literal questions, this pattern did not emerge in the present study. The results of this study appear to partly support a developmental perspective on the progression of inferencing skills. Previous research has shown that young and less skilled readers tend to generate fewer inferences than older and more skilled readers, and that inferencing ability improves with age (Oakhill, 1984; Casteel & Simpson, 1991; Pike et al., 2010). This may explain the uniform pattern of reduced accuracy on inferential questions across all language groups, reflecting a broader developmental weakness in inferencing. Oakhill (1982) also found that both skilled and less skilled readers performed equally well when asked to recognise exact phrases from short texts they had read.

Considering the current findings, the work of Cain and Oakhill (1999; 2001) is particularly relevant, as it emphasizes that inferential processing is not merely a matter of language competence or knowledge, but also of cognitive strategy use during reading— such as activating background knowledge, monitoring comprehension, and spontaneously applying inferential processes. They pointed out that even children with sufficient linguistic skills may underperform on inference questions if they do not spontaneously engage relevant strategies. This may explain why even typically developing children showed reduced accuracy on inferential questions —inference generation requires additional metacognitive steps that many children, particularly in the elementary school years, do not yet apply automatically. Conversely, children with language-based disorders tend to underperform overall, likely due to both their language limitations (e.g., in vocabulary, syntax, or discourse processing) and reduced flexibility in applying comprehension strategies. Nonetheless, the observed pattern of decreased performance on inferential items was similar across groups, supporting the notion

that inferencing is broadly demanding and not solely determined by language competence, but also by cognitive control and strategic knowledge use, along with developmental trajectory. Additionally, it is possible that the type of reasoning and information integration required for generating *anaphoric* or *bridging inferences*, which serve to link an anaphora to its corresponding antecedent or to bridge gaps between different parts of the text, is systematically easier or more difficult for all groups (Snow, 2002; Yang et al., 2007). It may also be the case that, under different conditions and when examining performance on other, more challenging types of inferencing, such as drawing predictive inferences about upcoming events, differences between specific language profiles in relation to question type would be more pronounced. Thus, these findings further support the general vulnerability of children with language-based disorders in reading comprehension, regardless of whether the task involves surface-level (literal) or deeper-level (inferential) understanding, in line with the SVR model, (Gough & Tunmer, 1986a; 2020b) and the Fully Distinct Hypothesis (Catts et al., 2005). In addition, variability in lexical features may have introduced extra processing load. Since the antecedents and anaphors systematically varied in frequency and imageability, all groups were required to process words that differed in lexical difficulty. According to the Lexical Quality Hypothesis (Perfetti & Hart, 2002), the quality of lexical representations determines the speed and success of integrating meaning into the broader context. If certain words in the texts were low-frequency or low in imageability, this may have imposed a proportional burden on all participants—regardless of language profile—thereby “flattening” group differences that would otherwise have been more apparent on inferential questions.

6.3 The predictive role of (meta)linguistic skills in reading comprehension

The results are discussed in relation to the third research question and its accompanying fifth hypotheses (subhypotheses H5, H5.1, H5.2, H5.3 and H5.4), which guided the examination of the predictive role of cognitive (phonological working memory) and various (meta)linguistic measures in discourse comprehension.

The third research objective was to identify which (meta)linguistic and cognitive factors contribute to individual differences in reading comprehension among the same participants from the previous analysis. Reading comprehension was defined as the cumulative accuracy score on literal and inferential questions. Hierarchical regression analyses were conducted to examine the unique and combined contributions of phonological skills, semantic knowledge, inferential reasoning, and phonological working memory to literal and inferential

comprehension. The analysis involved constructing composite scores for three groups of variables, based on the Simple View of Reading model (Gough & Tunmer, 1986): decoding (phonological (meta)linguistic skills), language comprehension (non-phonological (meta)linguistic skills), and a phonological processing composite (phonological working memory, spoonerisms). These composites were then used as predictors in the regression analysis. The results confirmed only a partial contribution of these components. Specifically, group membership (typical development vs. language-based disorders) and the language comprehension composite significantly predicted reading comprehension performance. Together, group membership and the language comprehension composite accounted for 52.7% of the variance in reading comprehension outcomes. The addition of the decoding composite and the phonological processing composite did not significantly increase the explained variance. The strongest individual predictor was the language comprehension composite (explaining 36.5% of the variance), while the decoding composite and phonological processing were not statistically significant.

The fact that the language comprehension composite, which includes vocabulary breadth and depth, grammatical understanding, and inferential reasoning, was the strongest individual predictor of reading comprehension performance once group was considered, is supported by both the SVR model (Gough & Tunmer, 1986) and the Lexical Quality Hypothesis (Perfetti & Hart, 2002). Both models emphasize that semantic knowledge stored in the lexicon, inferencing skills, and general language comprehension abilities enable the fluent integration of textual information and facilitate the construction of meaning-based coherence (Cain & Oakhill, 2014; Silva & Cain, 2015). In other words, these abilities are essential for building a coherent mental representation of the text to achieve comprehension. Thus, in addition to the quality of lexical representations, vocabulary breadth and depth, inferential reasoning, and general language comprehension function not merely as isolated components, but as integrated elements of the language system that contribute to dynamic text understanding. This finding is further supported by recent studies highlighting that inferencing skills and vocabulary depth are among the best predictors of comprehension (Cain & Oakhill, 2014; Silva & Cain, 2015). Moreover, language comprehension and decoding jointly predict reading comprehension outcomes in the early phases of literacy development. However, once reading becomes automatised, typically around the third or fourth year of schooling, the relative contribution of each predictor shifts: language comprehension becomes a stronger

and more dominant predictor, while the predictive role of decoding diminishes (Geva & Farnia, 2012; Cain, 2015; LARRC, 2015; Lervåg et al., 2017).

Research also shows that the relationship between language comprehension (LC) and reading comprehension (RC) is particularly strong in languages with transparent orthographies. Müller and Brady (2001) found a stronger association between LC and RC in beginning readers of transparent languages than in those learning to read in non-transparent orthographies. Florit and Cain's (2011) meta-analysis demonstrated that LC is only a moderate predictor of RC for beginning readers of English, whereas in transparent orthographies LC is a stronger predictor than decoding (accuracy) as early as the first and second years of primary school. Similar findings come from research on Italian-speaking children (Tobbia & Bonifacci, 2015), which confirmed the primary role of LC in the early primary years. Croatian data likewise show that LC is the main - and indeed the only - predictor of reading comprehension in the second year of primary school among typically developing readers (Kuvač Kraljević et al., 2021).

Contrary to hypotheses H3.1 and H3.4, the decoding composite, which included accuracy and speed in reading words and pseudowords, did not show significant predictive value. This finding can also be interpreted in light of theoretical models that emphasize a developmental shift in the role of decoding: although phonological skills are fundamental for the initial acquisition of reading (Gough & Tunmer, 1986; Verhoeven & van Leeuwe, 2012; Cain, 2015; Lervåg et al., 2017), their predictive power for reading comprehension decreases once a minimal threshold of decoding proficiency is reached (Benson, 2008; Kim, 2015). Given that the sample included children in the middle years of primary school, it can be assumed that most participants had already mastered basic decoding skills - at least to the extent that phonological components no longer serve as a limiting factor in reading comprehension.

Similarly, the phonological processing composite, which included measures of phonological awareness (spoonerisms) and phonological working memory (nonword repetition), was not a significant predictor of reading comprehension performance. Therefore, hypothesis H3.1, which posited a positive relationship between phonological working memory and reading comprehension, was not supported. Kershaw and Schatschneider (2012) also found no predictive role for phonological processing in reading comprehension among children in Grades 1-3, nor did Keresteš et al. (2019). However, it is possible that, at this age, other aspects of working memory play a greater role in comprehension than the phonological loop

alone (Cain, 2006; Carretti et al., 2009). Although some studies highlight the importance of the phonological loop for temporarily holding phonological information during reading (Snowling et al., 1991; Gathercole & Pickering, 2001), its contribution appears to vary depending on task type, age of the children, and the degree of automatisisation in both language and reading processes (Cain et al., 2004; Florit et al., 2009). In the present study, it is possible that the comprehension tasks did not place sufficient demands on phonological memory, or that other resources (e.g., semantic) were more efficient for carrying out inferential processes.

Regarding the participants' language profile (TLD vs. LBD), it remained a strong predictor of reading comprehension performance at all steps of the regression analysis. These findings confirm the presence of qualitative differences in (meta)linguistic and cognitive resources between children with language-based disorders and their typically developing peers (Adlof & Hogan, 2018; Nation et al., 2010) and support componential models of language-based disorders (Catts et al., 2005; Ramus et al., 2013). Children with language-based disorders (LBD) performed significantly worse in discourse comprehension than children with typical language development (TLD), even when other linguistic and cognitive factors were controlled for. These results indicate qualitative differences in language-cognitive resources between groups, consistent with previous research (Catts et al., 2005; Nation et al., 2010; Adlof & Hogan, 2018). Furthermore, the finding that group membership remained a significant predictor of reading comprehension after accounting for the specific (meta)language and processing composites, aligns with theoretical perspectives suggesting that language-based disorders involve broader language and cognitive deficits extending beyond the specific skills assessed. In other words, group membership appears to account for additional variance in comprehension performance not fully explained by the measured variables (phonological and non-phonological (meta)language skills, phonological memory). This suggests that there is something inherent to the profile of children with language-based disorders that affects comprehension performance but is not captured by the individual measures included in the model.

The findings also suggest that measuring only a limited set of targeted abilities (e.g., vocabulary, grammar comprehension, nonword repetition) may not be sufficient to account for the broader range and interaction of deficits that affect children's reading comprehension in authentic contexts. Language-based disorders may also involve co-occurring language, cognitive, and behavioural challenges — such as pragmatic difficulties, attention deficits,

memory and executive function impairments, or weaknesses in comprehension monitoring — which were not directly assessed in this study but could help explain the persistent group-level differences. As these dimensions were not included in the regression model, it is possible that their influence is "absorbed" within the group variable. It is also likely that children with language disorders experience a non-linear combination of multiple mildly impaired components, which — when co-occurring — exert a disproportionately negative effect on comprehension. These “fragmented” profiles — e.g., relatively intact vocabulary combined with weak syntax and phonological processing — may be averaged out within a composite score yet still result in functional comprehension difficulties. Such patterns may be effectively captured by the group variable, which could explain its retained statistical significance. Similarly, children with cumulatively lower scores across multiple domains — even if not falling below clinical thresholds in any one area — often demonstrate difficulties that emerge only during more complex comprehension tasks.

According to the Component model of language-based disorders, in line with the Fully Distinct Hypothesis (Catts et al., 2005), children with LBD do not form a homogeneous group. Instead, they include children with predominant phonological, semantic, syntactic, or inferential deficits, as well as those with combined difficulties. In the context of the obtained results, the group variable can be interpreted as a latent indicator of broader developmental deviations and cognitive-linguistic configurations not fully captured by the existing measures. This interpretation is supported by theoretical models emphasizing the multidimensional nature of language-based disorders (Catts et al., 2005; Ramus et al., 2013; Adlof & Hogan, 2018), as well as empirical findings pointing to the existence of “poor comprehenders” (Cain & Oakhill, 2006; Nation et al., 2010), whose deficits often become apparent only at the discourse level, rather than in isolated language tasks.

7. HYPOTHESIS VERIFICATION

H1: Psycholinguistic word properties — specifically frequency and imageability — will influence the processing of anaphoric relations in both children with language-based disorders and children with typical language development.

This hypothesis is partially accepted. Psycholinguistic word properties influenced anaphoric processing in both groups; however, this effect was driven primarily by word frequency, while imageability effects were weaker, less consistent and limited in scope.

H1.1: Both groups of children will exhibit longer fixation durations on low-frequency words compared to high-frequency words.

This hypothesis is fully accepted. Low-frequency words consistently elicited longer fixation durations than high-frequency words across multiple eye-tracking measures, areas of interest and groups, indicating increased processing demands during anaphora resolution.

H1.2: Both groups of children will exhibit longer fixation durations on low-imageability words compared to high-imageability words.

This hypothesis is partially accepted. Effect of imageability were limited and phase-dependent, emerging primarily in early processing and in interaction with frequency, rather than as a stable main effect across measures and regions.

H2: Psycholinguistic word properties — frequency and imageability — will exert a greater influence on anaphora processing during text comprehension in children with language-based disorders compared to children with typical language development.

This hypothesis is partially accepted. Children with language-based disorders showed greater overall processing costs; however, there was no consistent evidence that the effects of frequency or imageability were systematically stronger in this group, as no robust interaction between group and psycholinguistic properties was observed. The observed differences were primarily driven by general group differences and the main effect of frequency.

H2.1: The effect of word imageability will be more pronounced in children with poor reading comprehension, such that the difference in processing low- versus high-imageability words (in terms of longer gaze durations in the region of interest, increased number of fixations, longer total fixation durations, and a greater number of regressions) will increase as reading comprehension scores decrease.

This hypothesis is rejected. Word imageability did not differentially affect processing across language-based profiles, as no significant interaction between imageability and group (including poor reading comprehension profiles) was observed.

H2.2: The effect of word frequency will be more pronounced in children with poor decoding skills, such that the difference in processing low- versus high-frequency words (in terms of longer gaze durations in the region of interest, increased number of fixations, longer total fixation durations, and a greater number of regressions) will increase as reading comprehension scores decrease.

This hypothesis is fully accepted. A significant Frequency $\times$ Group interaction was observed across several processing measures. Low-frequency words were associated with longer first-pass duration and dwell time in the PD and MD groups, and with a greater number of fixations on antecedents in these groups. Thus, frequency effects were more pronounced in children with poor decoding skills, as observed in the PD and MD groups.

H3: Both children with typical language development and children with language- based disorders will demonstrate lower accuracy on inferential questions compared to literal questions.

This hypothesis is fully accepted. Both children with typical language development (TLD) and children with language-based disorders (LBD) showed lower accuracy on inferential questions compared to literal questions. Although children with TLD outperformed children with LBD overall, the absence of a significant interaction indicates that the effect of question type was comparable across groups.

H4: The type of question (literal vs. inferential) will differentially affect reading comprehension depending on the language-based disorder profiles.

This hypothesis is partially accepted. Reading comprehension accuracy differed between children with language-based disorders and those with typical language development, but this effect was not modulated by question type across profiles.

H4.1: Poor decoders will show difficulties in accurately answering both literal and inferential questions.

This hypothesis is rejected. Although poor decoders demonstrated lower overall accuracy, this effect did not differ as a function of question type, as indicated by the absence of a significant interaction between question type and group.

H4.2: Children with poor reading comprehension will show greater difficulty accurately answering inferential questions compared to literal ones.

This hypothesis is rejected. Children with poor reading comprehension did not show a disproportionate difficulty with inferential questions, as there was no significant interaction between question type and group.

H5: Phonological working memory (phonological loop), (meta)phonological skills, semantic skills, and inferential reasoning ability will significantly predict performance in discourse comprehension.

This hypothesis is partially accepted. Discourse comprehension was predicted by group membership and non-phonological metalinguistic skills, with inferential reasoning and semantic abilities (vocabulary) making the strongest contributions.

H5.1: There is a significant positive correlation between phonological working memory and discourse comprehension.

This hypothesis is rejected. Phonological working memory was not significantly associated with discourse comprehension once other (meta)linguistic factors were considered.

H5.2: There is a significant positive correlation between (meta)linguistic measures (phonological and semantic) and discourse comprehension.

This hypothesis is partially accepted. Metalinguistic skills were related to discourse comprehension, but this effect was driven by non-phonological rather than phonological language measures.

H5.3: There is a significant positive correlation between inferential reasoning skills and discourse comprehension.

This hypothesis is fully accepted. Inferential reasoning skill showed a robust positive association with discourse comprehension performance.

H5.4: Each independent variable (phonological working memory and (meta)linguistic skills) will serve as a significant individual predictor of discourse comprehension success.

This hypothesis is partially accepted. Only non-phonological metalinguistic skills emerged as independent predictors of discourse comprehension, while phonological working memory and phonological metalinguistic skills did not.

8. LIMITATIONS OF THE STUDY

While the findings of this study provide valuable insights into how children with different language profiles process text and respond to comprehension questions, several methodological limitations should be noted, as they may affect interpretation.

First, due to the lack of an appropriate data source, the frequency and imageability values of words were based on corpora of adult speakers. As children's lexical experience can differ significantly from that of adults, certain word properties may not have reflected children's actual familiarity or usage frequency, which may have influenced the results to some extent.

Second, although the study included a detailed breakdown of subgroups within the language-based disorder population, the number of participants in some of these subgroups — particularly those with poor oral language (POL) and mixed language difficulties (MD) — was relatively small. This lack of statistical power limits the reliability of detecting finer distinctions between groups, especially when examining interactions between psycholinguistic properties and language profiles. Additionally, although careful classification within the LBD group (PD, POL, MD) was applied, subgroup boundaries are not always clear-cut. Language profiles often overlap, and small sample sizes further complicate interpretation.

Third, the study primarily focused on anaphoric processing, while other discourse-level processes, such as establishing global coherence or making predictive inferences, were not included. This provided an important, but still limited, view of the complex mechanisms involved in text comprehension. Furthermore, although both online and offline methods were employed, the comprehension tasks did not directly assess a broader range of cognitive and metacognitive processes that might play a role, such as attention, comprehension monitoring,

or reading strategies. These factors can contribute to individual differences but were not specifically isolated in this study.

Fourth, although the composite model included inferencing ability, vocabulary, and grammatical understanding, these components were assessed through isolated tasks, which may not fully reflect the language demands of real-world discourse processing. For example, the inferencing test used in this study measured inference generation in a controlled verbal context, whereas narrative comprehension requires integrating multiple pieces of information, maintaining global coherence, and tracking events over time - skills not necessarily captured by a single task.

Together, these limitations do not diminish the value of the results but rather highlight the need for caution in their interpretation and application, while opening important avenues for future research aimed at expanding and deepening current understanding.

9. SCIENTIFIC CONTRIBUTIONS

The present study makes one primary theoretical contribution, and one closely related methodological-conceptual contribution. All other contributions refine existing theoretical and methodological frameworks within this broader perspective.

The primary theoretical contribution of this dissertation is the extension of existing models of reading comprehension to discourse-level processing. While frameworks such as the Simple View of Reading and the Lexical Quality Hypothesis have established the importance of lexical quality for word recognition and sentence-level comprehension, its role in discourse integration has received comparatively less direct attention. This dissertation shows that lexical quality, operationalised through psycholinguistic word properties, continues to shape comprehension during the real-time establishment of reference and coherence in written discourse. Lexical quality therefore functions not only as a prerequisite for successful comprehension, but as an active mechanism that determines how efficiently readers link referents across sentences and maintain a coherent discourse representation as reading unfolds. In this way, the dissertation adds a process-oriented perspective to accounts that have largely relied on outcome-based interpretations of psycholinguistic effects. Crucially, previous research has typically examined the effects of individual psycholinguistic word properties in isolation. This dissertation extends existing work by considering the contribution of both word frequency and imageability within the same analytical framework. This approach shows that frequency exerts a broader and more robust influence on discourse

processing, affecting both antecedents and anaphoric expressions, whereas the effects of imageability are more restricted and largely confined to the antecedent. In doing so, the study clarifies how different aspects of lexical quality contribute in distinct ways to discourse integration.

The second main contribution is methodological and conceptual. In majority of the existing literature, difficulties with discourse comprehension have been inferred indirectly from performance on comprehension tasks or response accuracy. Although reading is widely recognised as a dynamic process, discourse-level processing itself has rarely been examined directly, particularly in clinical populations, including readers with language-based disorders, and in pro-drop languages (i.e. Croatian). Importantly, the process-level perspective used in this dissertation makes it possible to identify systematic differences between language-based profiles, revealing that readers whose difficulties are rooted in decoding (i.e. poor decoders and those with mixed language difficulties) show the greatest and most persistent processing limitations during discourse integration. In this way, the study adds process-level evidence to existing accounts of discourse-level reading difficulties, which have predominantly relied on behavioural measures.

The dissertation further provides secondary theoretical support for multidimensional models of language-based disorders by refining the view that reading comprehension difficulties arise from partially independent mechanisms rather than from a single continuum of severity. In addition, the findings refine developmental accounts by suggesting that inferential processing remains closely tied to general discourse integration in children aged 10-12 years and may not yet function as a stable source of individual differentiation.

Finally, this dissertation strengthens key theoretical frameworks of reading comprehension by clarifying how different sources of difficulty contribute at this stage of development. Within the Simple View of Reading, the findings support the central role of language comprehension in reading comprehension around the age of 11, once decoding is largely established. Within the Lexical Quality Hypothesis, the results show that word frequency plays a stronger and more far-reaching role than imageability in written anaphora comprehension when both properties are considered together. With respect to the Fully Distinct Hypothesis, the observed performance patterns across different language-based profiles in this study do not point to a fully clear-cut distinction between decoding- and language-based difficulties. Instead, the results suggest a more complex picture, with both shared and profile-specific

characteristics evident across measures. This highlights the importance of considering not only differences between profiles, but also their areas of overlap, in both theoretical accounts and clinical practice. Additionally, by examining anaphora resolution in Croatian, the study supports the cross-linguistic applicability of process-based models of discourse comprehension in school-aged children with and without language-based disorders.

10. IMPLICATIONS FOR FUTURE RESEARCH

Several directions for future research could further enhance our knowledge of individual differences in reading comprehension among children with varying language profiles. As children's lexical experiences differ both quantitatively and qualitatively from those of adults, the use of age-appropriate language corpora could yield more accurate and ecologically valid insights into the relationship between lexical properties and processing. Additionally, replicating the study with a larger and more balanced sample would allow for more reliable statistical comparisons across groups and more precise identification of patterns associated with specific language weaknesses. Beyond anaphoric processing as one aspect of discourse comprehension, future studies should broaden their scope to include other key components of text understanding- such as causal inference, global coherence maintenance, content prediction, and the interpretation of pragmatic implications. This broader perspective would provide a more comprehensive view of how different language and cognitive skills contribute to multiple levels of discourse processing. Another important and still underexplored area is the role of executive functions and metacognition — particularly in relation to comprehension monitoring and the strategies children spontaneously or explicitly use while reading. Future research could integrate qualitative methods, such as introspective interviews, think-aloud protocols, or combined online (e.g., eye-tracking) and offline methods, to gain deeper insights into how children approach reading tasks and what hinders or supports their construction of meaning.

Finally, longitudinal studies on the development of inferencing skills and their relationship to other language and executive functions would be a crucial step towards identifying developmental profiles and predictors of comprehension difficulties, as well as designing timely and targeted forms of educational and clinical support.

11. CONCLUSION

The aim of this study was to examine how psycholinguistic word properties (imageability and frequency), question types (literal and inferential), and varying language skills influence reading comprehension in children with specific language-based profiles (typical language development, language-based disorder: poor oral language group, poor decoders, mixed language difficulties).

The findings of this study demonstrate that reading comprehension in school-aged children is shaped by the interaction between foundational reading components, as described by the Simple View of Reading (Gough & Tunmer, 1986), and the quality of lexical representations, as proposed by the Lexical Quality Hypothesis (Perfetti & Hart, 2002; Perfetti, 2007). Across all analyses, word frequency emerged as the most robust and consistent psycholinguistic predictor of processing efficiency, exerting effects not only during early lexical access but also during the integration of referential information in anaphora resolution. While the facilitative role of frequent words in recognition and retrieval is well established (Inhoff & Rayner, 1986; Brysbaert et al., 2018), the present findings extend this effect to discourse-level processing, demonstrating its influence on maintaining textual cohesion.

Imageability contributed to processing efficiency both as an independent factor and more frequent in interaction with frequency, particularly during antecedent processing, where conceptual activation plays a central role. In earliest processing phase, imageability effects were observable alongside frequency, suggesting that semantic support is available from the initial phases of lexical access, although its influence remains more limited and context-dependent compared to frequency. A somewhat paradoxical pattern emerged in the earliest processing phase: high imageability increased processing demands when paired with low frequency words, suggesting that a vivid conceptual representation places additional burden on processing when the underlying lexical form is weakly automatised. For high frequency words, imageability had no significant effect, and it did not emerge as a reliable predictor in later processing phases or in the anaphora region.

In contrast, anaphor processing showed a different pattern. While frequency emerged as a robust predictor in later processing stages, early stages were not solely driven by frequency. Some evidence of imageability effects was observed in the earliest phase; however, this was limited to a single measure, indicating that its role at this stage is less stable and should be interpreted with caution. Thus, although frequency plays a dominant role overall, early

anaphor processing potentially reflects a more complex interplay of lexical and semantic factors rather than a purely frequency-driven mechanism. These findings are broadly consistent with models emphasising frequency-based differences in lexical activation (Coltheart et al., 2001; Balota et al., 2004; Grainger & Holcomb, 2009), while also highlighting the contribution of semantic richness in the initial phases of processing. Taken together, the results suggest that frequency operates across multiple levels of the mental lexicon (Levelt, 1992; Kittredge et al., 2008), reinforcing lexical stability through repeated exposure and strengthening phonological, orthographic, and semantic connections (Perfetti, 2007; Perfetti & Stafura, 2014), but that its effects are modulated by other psycholinguistic properties, particularly in early processing.

Group differences were evident across processing phases but were neither uniform nor reducible to a simple efficiency gradient. The findings reveal partially distinct profiles across language-based disorder subgroups, providing partial support for the hypothesised differential effects of psycholinguistic properties. Children in the PD and MD groups showed specific sensitivity to word frequency in early integrative antecedent processing, with low frequency words producing disproportionately longer first pass durations. This pattern reflects a compounding interaction between weakened phonological-orthographic representations and the inherently weaker lexical entries of low-frequency words, a cumulative burden that becomes measurable already in early processing. The MD group further demonstrated unique frequency sensitivity in anaphoric processing, with the frequency effect on number of fixations confirmed exclusively within this group, suggesting that co-occurring decoding and language comprehension difficulties extend the frequency-related processing burden into referential resolution.

The POL group exhibited a qualitatively distinct profile: rather than showing globally slower processing, this group specifically failed to benefit from high word frequency in antecedent dwell time, a pattern not observed in any other group. This suggests that their difficulties reflect insufficiently specified semantic-conceptual dimensions of lexical quality, consistent with the profile of poor comprehenders whose difficulties arise at the level of semantic integration rather than word recognition (Cain et al., 2001; Oakhill & Cain, 2012). Although hypothesis regarding a more pronounced imageability effect in POL was not supported, the POL findings nonetheless reveal a meaningful frequency-related deficit consistent with weaknesses in the semantic dimension of lexical quality.

Taken together, the type of language-based disorder modulates which dimension of lexical quality is most critically implicated in referential processing during reading, a distinction with both theoretical and clinical relevance. At the level of comprehension outcomes, all groups showed lower accuracy on inferential than on literal questions, confirming that inferencing remains a developmentally sensitive process throughout early adolescence (Cain & Oakhill, 2001; Pike et al., 2010). While lexical processing provides the foundation for establishing referential links, successful inference-making further requires higher-order integration processes that continue to mature beyond the age range studied. Importantly, although children with language-based disorders performed less accurately overall than their typically developing peers, the absence of a significant interaction indicates that the effect of question type was comparable across groups.

Regression analyses further demonstrated that language comprehension skills: vocabulary, grammatical knowledge, and inferential reasoning - were stronger predictors of reading comprehension than phonological decoding at ages 11-12, particularly in a transparent orthography such as Croatian (Silva & Cain, 2015; Tobbia & Bonifacci, 2015; Kuvač Kraljević et al., 2021). Group membership remained a significant predictor even after controlling for specific linguistic skills, indicating that language-based disorders encompass broader developmental differences not fully captured by isolated measures (Keenan et al., 2008).

Methodologically, the integration of online eye-tracking measures with offline comprehension outcomes provided a more precise account of how comprehension is achieved, not merely whether it is successful. Differentiating subgroups within the language-based disorder population further revealed distinct processing profiles that would remain obscured in homogeneous group analyses. Together, these findings emphasize the necessity of multidimensional assessment and targeted intervention approaches that simultaneously address lexical quality, language comprehension, and higher-level discourse processes to support reading comprehension in children with language-based disorders.

12. REFERENCES

- Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, 6(3), 251–278. [https://doi.org/10.1076/1380-3611\(200009\)6:3;1-A;FT251](https://doi.org/10.1076/1380-3611(200009)6:3;1-A;FT251)
- Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing*, 14(1), 61–89. <https://doi.org/10.1023/A:1008128417862>
- Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, 102(2), 281–298. <https://psycnet.apa.org/doi/10.1037/a0019318>
- Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, 8(3), 323–338. <https://doi.org/10.1177/107319110100800308>
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.
- Adlof, S. M., & Catts, H. W. (2015). Morphosyntax in poor comprehenders. *Reading and Writing*, 28(7), 1051–1070. <https://doi.org/10.1007/s11145-015-9562-3>
- Adlof, S. M., & Hogan, T. P. (2018). Understanding dyslexia in the context of developmental language disorders. *Language, Speech, and Hearing Services in Schools*, 49(4), 762–773. https://doi.org/10.1044/2018_lshss-dyslc-18-0049
- 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., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, 19(9), 933–958. <https://doi.org/10.1007/s11145-006-9024-z>
- Ahn, S., Kelton, C., Balasubramanian, A., & Zelinsky, G. (2020). Towards predicting reading comprehension from gaze behavior. *Proceedings of the Eye Tracking Research and Applications Symposium (ETRA)*. <https://doi.org/10.1145/3379156.3391335>

- Aksu-Koç, A., & Nicolopoulou, A. (2015). Character reference in young children's narratives: A crosslinguistic comparison of English, Greek, and Turkish. *Lingua*, 155, 62–84. <https://doi.org/10.1016/j.lingua.2014.04.006>
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1061–1070. <https://doi.org/10.1037/0278-7393.19.5.1061>
- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, 25(2), 92–98. <https://doi.org/10.1027/1015-5759.25.2.92>
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Alonzo, C. N., McIlraith, A. L., Catts, H. W., & Hogan, T. P. (2020). Predicting dyslexia in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(1), 151–162. https://doi.org/10.1044/2019_JSLHR-L-18-0265
- Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *WIREs Cognitive Science*, 4(2), 149–168. <https://doi.org/10.1002/wcs.1215>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.
- Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal*, 14(4), 367–381. <https://doi.org/10.2307/1162336>
- Andersson, U. (2010). The contribution of working memory capacity to foreign language comprehension in children. *Memory*, 18(4), 458–472. <https://doi.org/10.1080/09658211003762084>

- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439–461.
<https://doi.org/10.3758/BF03214334>
- Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, 107, 868–883.
<https://doi.org/10.1037/edu0000006>
- Archibald, L. M. D., & Gathercole, S. E. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language*, 57(2), 177–194. <https://doi.org/10.1016/j.jml.2006.11.004>
- Artuso, C., & Palladino, P. (2022). Working memory, vocabulary breadth and depth in reading comprehension: A study with third graders. *Discourse Processes*, 59(9), 685–701.
<https://doi.org/10.1080/0163853X.2022.2116263>
- Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology Section A*, 58(6), 1065–1086.
<https://doi.org/10.1080/02724980443000476>
- 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](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](https://doi.org/10.1016/s0021-9924(03)00019-4)
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481. <https://doi.org/10.1037/a0023851>
- Barbosa, P. (2011). Pro-drop and theories of pro in the Minimalist Program part 2: Pronoun deletion analyses of null subjects and partial, discourse and semi pro-drop. *Language and Linguistics Compass*, 5(8), 571–587. <https://doi.org/10.1111/j.1749-818X.2011.00292.x>

- Barry, C., & Gerhand, S. (2003). Both concreteness and age of acquisition affect reading accuracy but only concreteness affects comprehension in a deep dyslexic patient. *Brain and Language*, 84(1), 84–104. [https://doi.org/10.1016/S0093-934X\(02\)00522-9](https://doi.org/10.1016/S0093-934X(02)00522-9)
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Batinić Angster, M. (2020). *Rečenična anafora: Morfologija i sintaksa povratnih imenskih izraza u hrvatskom i talijanskom jeziku*. Institut za hrvatski jezik i jezikoslovlje.
- Bawayan, R., & Brown, J. A. (2022). Diagnostic decisions of language complexity using informal language assessment measures. *Language, Speech, and Hearing Services in Schools*, 53(2), 466–478. https://doi.org/10.1044/2021_LSHSS-21-00065
- Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology*, 86(2), 244–255. <https://doi.org/10.1037/0022-0663.86.2.244>
- Benkí, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America*, 113(3), 1689–1705. <https://doi.org/10.1121/1.1534102>
- Benson, N. (2008). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, 26(1), 27–41. <https://doi.org/10.1177/0734282907301424>
- Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools*, 44(2), 133–146. [https://doi.org/10.1044/0161-1461\(2012/12-0093\)](https://doi.org/10.1044/0161-1461(2012/12-0093))
- 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., & 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., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child Development*, 80(2), 593–605. <https://doi.org/10.1111/j.1467-8624.2009.01281.x>

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>

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & 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>

Blaži Ostojić, A. (2023). Reading comprehension processes: A review based on theoretical models and research methodology. *Hrvatska revija za rehabilitacijska istraživanja*, 59(1), 122–143. <https://doi.org/10.31299/hrri.59.1.9>

Bley-Vroman, R. (2002). Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition*, 24(2), 209–213.

<https://doi.org/10.1017/S0272263102002048>

Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Livsedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 643–662). Oxford University Press.

Blythe, H. I., Livsedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, 49(2), 1583–1591.

Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology*, 39(3), 206–219.

Botting, N., & Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *International Journal of Language & Communication Disorders*, 40(1), 49–66. <https://doi.org/10.1080/13682820410001723390>

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, 49(2), 229–240.

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 155–172). Blackwell.

<https://doi.org/10.1002/9780470757642.ch9>

- Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, 75(1), 189–201. <https://doi.org/10.1348/000709904X22674>
- Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., & Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing*, 29, 435–451. <https://doi.org/10.1007/s11145-015-9608-6>
- Bruggink, M., Swart, N., Van der Lee, A., & Segers, E. (2022). *Putting PIRLS to use in classrooms across the globe: Evidence-based contributions for teaching reading comprehension in a multilingual context* (p. 110). Springer Nature.
- Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, 64, 545–559. <https://doi.org/10.1080/17470218.2010.503374>
- Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 441–458.
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, 8, 531–544.
- Büring, D. (2005). *Binding theory*. Cambridge University Press.
- Burns, M. K., Kwoka, H., Lim, B., Crone, M., Haegele, K., Parker, D. C., & Scholin, S. E. (2011). Minimum reading fluency necessary for reading comprehension among second-grade students. *Psychology in the Schools*, 48(2), 124–132. <https://doi.org/10.1002/pits.20531>
- Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory*, 14(5), 553–569.
- Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, 28(4), 679–694.

Cain, K. (2009). Children's reading comprehension difficulties: A consideration of the precursors and consequences. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 59–75). Routledge.

Cain, K. (2010). *Reading development and difficulties*. John Wiley & Sons.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure. *Reading and Writing*, 11(5–6), 489–503. <https://doi.org/10.1023/A:1008084120205>

Cain, K., & Oakhill, J. (2006). *Children's comprehension problems in oral and written language: A cognitive perspective*. Guilford Press.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, 44(5), 431–443. <https://doi.org/10.1177/0022219411410042>

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, 114(4), 647–662. <https://doi.org/10.4074/S0003503314004035>

Cain, K., Oakhill, J., & Bryant, P. (2004a). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31–42. <https://doi.org/10.1037/0022-0663.96.1.31>

Cain, K., Oakhill, J., & Lemmon, K. (2004b). Individual Differences in the Inference of Word Meanings From Context: The Influence of Reading Comprehension, Vocabulary Knowledge, and Memory Capacity. *Journal of Educational Psychology*, 96(4), 671–681. <https://doi.org/10.1037/0022-0663.96.4.671>

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, 90(1), 65–87.

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, 30(3), 681–694.

Cain, K., Oakhill, J. V., Barnes, M., & Bryant, P. E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory & Cognition*, 29, 850–859.

Cain, K., Compton, D. L., & Parrila, R. K. (Eds.). (2017). *Theories of reading development*. John Benjamins. <https://doi.org/10.1075/swll.15>

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102(3), 280–298.

Cao, Y., & Kim, Y. S. G. (2021). Is retell a valid measure of reading comprehension? *Educational Research Review*, 32, 100375.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, 23(6), 678–686.
<https://doi.org/10.1177/0956797611434536>

Carpenter, P. A., & Just, M. A. (2017). Cognitive processes in reading: Models based on readers' eye fixations. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 177–213). Routledge.

Carroll, J. M., & Snowling, M. J. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry*, 45(3), 631–640.
<https://doi.org/10.1111/j.1469-7610.2004.00252.x>

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19(2), 246–251.
<https://doi.org/10.1016/j.lindif.2008.10.002>

Carver, R. P. (1982). Optimal rate of reading prose. *Reading Research Quarterly*, 18(1), 56–88.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51.
<https://doi.org/10.1177/1529100618772271>

Catts, H. W. (2018). The simple view of reading: Advancements and false impressions. *Remedial and Special Education*, 39(5), 317–323.
<https://doi.org/10.1177/0741932518767563>

Catts, H. W., & Hogan, T. P. (2003). Language basis of reading disabilities and implications for early identification and remediation. *Reading Psychology, 24*(3–4), 223–246.
<https://doi.org/10.1080/02702710390227314>

Catts, H. W., & Kamhi, A. G. (1999). *Language and reading disabilities*. Allyn & Bacon.

Catts, H. W., & Kamhi, A. G. (Eds.). (2005). *The connections between language and reading disabilities*. Psychology Press.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*(6), 1378–1396. [https://doi.org/10.1044/1092-4388\(2005/096\)](https://doi.org/10.1044/1092-4388(2005/096))

Catts, H. W., Adlof, S. M., & Ellis Weismer, S. (2006). Language deficits in poor comprehenders: a case for the simple view of reading. *Journal of speech, language, and hearing research : JSLHR, 49*(2), 278–293. [https://doi.org/10.1044/1092-4388\(2006/023\)](https://doi.org/10.1044/1092-4388(2006/023))

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. [https://doi.org/10.1044/1092-4388\(2002/093\)](https://doi.org/10.1044/1092-4388(2002/093))

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*(2), 151–164. <https://doi.org/10.1177/002221940303600208>

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale, 59*(3), 209–217.
<https://doi.org/10.1037/h0087476>

Cho, E., Capin, P., Roberts, G., Roberts, G. J., & Vaughn, S. (2019). Examining Sources and Mechanisms of Reading Comprehension Difficulties: Comparing English Learners and Non-English Learners within the Simple View of Reading. *Journal of educational psychology, 111*(6), 982–1000. <https://doi.org/10.1037/edu0000332>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. <https://psycnet.apa.org/doi/10.1037/1040-3590.6.4.284>

- Clarke, P., Hulme, C., & Snowling, M. (2005). Individual differences in RAN and reading: A response timing analysis. *Journal of Research in Reading*, 28(2), 73–86.
<https://doi.org/10.1111/j.1467-9817.2005.00255.x>
- Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21(8), 1106–1116. <https://doi.org/10.1177/0956797610375449>
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 341–372). Elsevier.
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairment (SLI). *International Journal of Language and Communication Disorders*, 43(1), 1–40.
<https://doi.org/10.1080/13682820601116485>
- Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology Section A*, 57(7), 1153–1171. <https://doi.org/10.1080/02724980443000007>
- Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Blackwell.
<https://doi.org/10.1002/9780470757642.ch1>
- Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, 4(1), 5–17. <https://doi.org/10.1080/13603110600574322>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Cook, A. E. (2014). Processing anomalous anaphors. *Memory & Cognition*, 42(7), 1171–1185. <https://doi.org/10.3758/s13421-014-0415-0>
- Cook, A. E., & Wei, W. (2019). What can eye movements tell us about higher level comprehension? *Vision*, 3(3), Article 45. <https://doi.org/10.3390/vision3030045>
- Copeland, L., Gedeon, T., & Mendis, B. S. U. (2014). Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. *Artificial Intelligence Research*, 3(3), 35–48.

- Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, 66(5), 946–972.
<https://doi.org/10.1080/17470218.2012.722660>
- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25(1), 1–53.
<https://doi.org/10.1080/01638539809545019>
- Cragg, L., & Nation, K. (2006). Exploring written narrative in children with poor reading comprehension. *Educational Psychology*, 26(1), 55–72.
<https://doi.org/10.1080/01443410500340991>
- Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell.
- Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, 18(3), 415–436. <https://doi.org/10.1080/13594320500165896>
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277–299.
https://doi.org/10.1207/s1532799xssr1003_5
- Dahl, A. C., Carlson, S. E., Renken, M., McCarthy, K. S., & Reynolds, E. (2021). Materials matter: An exploration of text complexity and its effects on middle school readers' comprehension processing. *Language, Speech, and Hearing Services in Schools*, 52(2), 702–716. https://doi.org/10.1044/2021_LSHSS-20-00117
- 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](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.
<https://doi.org/10.3758/bf03214546>
- Davies, R., Cuetos, F., & Glez-Seijas, R. M. (2007). Reading development and dyslexia in a transparent orthography: A survey of Spanish children. *Annals of Dyslexia*, 57(2), 179–198.
<https://doi.org/10.1007/s11881-007-0010-1>

- Davies, R., Rodríguez-Ferreiro, J., Suárez, P., & Cuetos, F. (2013). Lexical and sub-lexical effects on accuracy, reaction time and response duration: Impaired and typical word and pseudoword reading in a transparent orthography. *Reading and Writing*, 26(5), 721–738. <https://doi.org/10.1007/s11145-012-9388-1>
- Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1298–1338. https://doi.org/10.1037/xlm0000366?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
- Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics*, 25(2), 223–238. doi:10.1017/S0142716404001110
- Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, 18(6), 432–451. <https://doi.org/10.1080/10888438.2014.926907>
- Deacon, S. H., & Kieffer, M. (2018). Understanding how syntactic awareness contributes to reading comprehension: Evidence from mediation and longitudinal models. *Journal of Educational Psychology*, 110(1), 72–84. https://doi.org/10.1037/edu0000198?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
- de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on reading development. *Journal of Educational Psychology*, 94(3), 594–611. https://doi.org/10.1207/S1532799XSSR0601_03
- de-la-Peña, C. (2024). Eye-tracking contribution on processing of (implicit) reading comprehension. *Journal of New Approaches in Educational Research*, 13(1), 13. <https://doi.org/10.1007/s44322-024-00013-w>
- Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, 10(2), 186–202. [https://doi.org/10.1016/S0010-9452\(74\)80009-2](https://doi.org/10.1016/S0010-9452(74)80009-2)

- Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *Journal of Neuroscience*, 39(39), 7722–7736.
<https://doi.org/10.1523/jneurosci.0675-19.2019>
- Desai, R. H., Choi, W., & Henderson, J. M. (2020). Word frequency effects in naturalistic reading. *Language, Cognition and Neuroscience*, 35(5), 583–594.
<https://doi.org/10.1080/23273798.2018.1527376>
- Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods*, 41(2), 546–557.
<https://doi.org/10.3758/BRM.41.2.546>
- Ding, J. H., & Wang, L. Y. (2006). Relations between phonological loop and reading comprehension. *Acta Psychologica Sinica*, 38(5), 694–701.
- Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading*, 16(6), 504–525.
<https://doi.org/10.1080/10888438.2011.598199>
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848–856. <https://doi.org/10.1111/jcpp.12378>
- Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, 33 (2), 103–119.
<https://doi.org/10.1177/002383099003300201>
- Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 205–242). International Reading Association.
- Dujardin, E., Ecalle, J., Auphan, P., Bailloud, N., & Magnan, A. (2023). Vocabulary and reading comprehension: What are the links in 7- to 10-year-old children? *Scandinavian Journal of Psychology*, 64(5), 582–594. <https://doi.org/10.1111/sjop.12912>
- Durukan, E. (2020). Impact of speed reading training on reading speeds and comprehension skills of secondary school students. *Cypriot Journal of Educational Sciences*, 15(2), 184–193.
<https://doi.org/10.18844/cjes.v15i2.4491>

- Dunn, L. M., Dunn, L. M., Kovačević, M., Padovan, N., Hržica, G., Kuvač Kraljević, J., Mustapić, M., Dobravac, G., & Palmović, M. (2009). *Peabody slikovni test rječnika – PPVT-III-HR*. Naklada Slap.
- Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading*, 21(5), 392–407. <https://doi.org/10.1080/10888438.2017.1315424>
- Edwards, S., Letts, C., Sinka, I., Kuvač Kraljević, J., Kologranić Belić, L., Hržica, G., & Kovačević, M. (2019). *Priručnik Nove Reynell razvojne jezične ljestvice* [Manual of the New Reynell Developmental Language Scale]. Naklada Slap.
- Egusquiza, N., Navarrete, E., & Zawiszewski, A. (2016). Antecedent frequency effects on anaphoric pronoun resolution: Evidence from Spanish. *Journal of Psycholinguistic Research*, 45(1), 71–84. <https://dx.doi.org/10.1007/s10936-014-9325-3>
- Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Blackwell.
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641–655. [https://doi.org/10.1016/S0022-5371\(81\)90220-6](https://doi.org/10.1016/S0022-5371(81)90220-6)
- Ehrlich, M. F., & Remond, M. (1997). Skilled and less skilled comprehenders: French children's processing of anaphoric devices in written texts. *British Journal of Developmental Psychology*, 15(3), 291–309. <https://doi.org/10.1111/j.2044-835X.1997.tb00522.x>
- Eilers, S. (2021). *Children's processing of anaphora during reading comprehension: An eye tracking study* [Doctoral dissertation, Universität Potsdam].
- Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018). Individual differences in children's pronoun processing during reading: Detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, 173, 250–267. <https://doi.org/10.1016/j.jecp.2018.04.005>
- Elbro, C., Oakhill, J., Megherbi, H., & Seigneuric, A. (2017). Aspects of pronominal resolution as markers of reading comprehension: The role of antecedent variability. *Reading and Writing*, 30(4), 813–827. <https://doi.org/10.1007/s11145-016-9702-4>

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1–44.

<https://doi.org/10.1080/19345740802539200>

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, 115(3), 497–516. <https://doi.org/10.1016/j.jecp.2013.03.004>

Elwér, Å., Gustafson, S., Byrne, B., Olson, R. K., Keenan, J. M., & Samuelsson, S. (2015). A retrospective longitudinal study of cognitive and language skills in poor reading comprehension. *Scandinavian Journal of Psychology*, 56(2), 157–166.

<https://doi.org/10.1111/sjop.12188>

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5), 621–636.

[https://doi.org/10.1016/s0042-6989\(01\)00301-7](https://doi.org/10.1016/s0042-6989(01)00301-7)

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813.

<https://doi.org/10.1037/0033-295X.112.4.777>

Engelen, J., Bouwmeester, S., de Bruin, A., & Jolles, J. (2014). Anticipatory processing in children with good and poor reading comprehension: Evidence from eye movements.

Reading and Writing, 27(6), 1031–1051. <https://doi.org/10.1016/j.jecp.2013.09.005>

Erdeljac, V., Sekulić, M., Willer-Gold, J., Biočina, Z., Čolović, N., Dragojević, E., & Rosandić, D. (2014). Leksičko obilježje predočivosti u mentalnom leksikonu osoba s afazijom. *Govor*, 31(1), 29–47.

Falkauskas, K., & Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1607–1627.

<https://doi.org/10.1037/xlm0000132>

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. John Benjamins.

- Ferrand, L., Bonin, P., Méot, A., Augustinova, M., & New, B. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, 40(4), 1049–1054. <https://doi.org/10.3758/BRM.40.4.1049>
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 43(2), 488–496. <https://doi.org/10.3758/brm.42.2.488>
- Fitch, R. H., Miller, S., & Tallal, P. (1997). Neurobiology of speech perception. *Annual Review of Neuroscience*, 20, 331–353. <https://doi.org/10.1146/annurev.neuro.20.1.331>
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323–330.
- Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, 23(4), 553–576. <https://doi.org/10.1007/s10648-011-9175-6>
- Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, 27(4), 935–951. <https://doi.org/10.1348/026151008X397189>
- Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes*, 48(2), 119–138. <https://doi.org/10.1080/0163853X.2010.494244>
- Foorman, B. R., & Petscher, Y. (2018). Decomposing the variance in reading comprehension to reveal the unique and common effects of language and decoding. *Journal of Visualized Experiments*, 140, Article e58557. <https://doi.org/10.3791/58557>
- Foorman, B., Koon, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). Examining general and specific factors in the dimensionality of oral language and reading in 4th–10th grades. *Journal of Educational Psychology*, 107(3), 884–899. <https://doi.org/10.1037/edu0000026>

- Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences*, 63, 12–23. <https://doi.org/10.1016/j.lindif.2018.02.011>
- Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In: G. Paris, S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 387–412). Routledge.
- Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language*, 97, 135–153. <https://doi.org/10.1016/j.jml.2017.07.008>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2010). Lexical quality in the brain: ERP evidence for quality of representation. *Journal of Cognitive Neuroscience*, 22(10)
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). Erlbaum.
- Gabaj, M., & Kuvač Kraljević, J. (2019). Označavanje likova u dječjem pripovjednom diskursu. *Logopedija*, 9(2), 40–49. <https://doi.org/10.31299/log.9.2.1>
- Galeković, M. (2023). Razumijevanje gramatičkih struktura i mjere fonološke obrade kod predškolske djece. *Govor*, 40(1), 55–82. <https://doi.org/10.22210/govor.2023.40.03>
- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74–111. <https://doi.org/10.3102/0034654313499616>
- Garnett, K. (2011). Fluency in learning to read: Conceptions, misconceptions, learning disabilities, and instructional moves. In J. R. Birsh (Ed.), *Multisensory teaching of basic language skills* (pp. 293–320). Brookes Publishing.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29(3), 336–360.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Lawrence Erlbaum.

Gathercole, S. E., & Pickering, S. J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education*, 28(2), 89–97.

<https://doi.org/10.1111/1467-8527.00225>

Georgiou, G. K., Parrila, R., & Kirby, J. R. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading*, 12(4), 325–350.

https://doi.org/10.1207/s1532799xssr1002_4

Gerth, S., & Festman, J. (2021). Reading development, word length and frequency effects: An eye-tracking study with slow and fast readers. *Frontiers in Communication*, 6, Article 743113. <https://doi.org/10.3389/fcomm.2021.743113>

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum. <https://doi.org/10.4324/9780203772157>

Gerwien, J., Filip, M., & Smolík, F. (2024). Noun imageability and the processing of sensory-based information. *Quarterly Journal of Experimental Psychology*, 77(10), 2137–2150. <https://doi.org/10.1177/17470218231216304>

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading comprehension. *Journal of Research in Reading*, 35(3), 261–284. <https://doi.org/10.1007/s11145-011-9333-8>

Geva, E., & Ryan, E. B. (1985). Use of conjunctions in expository texts by skilled and less skilled readers. *Journal of Reading Behavior*, 17(4), 331–346. <https://doi.org/10.1080/10862968509547549>

Gillon, G. (2004). *Phonological Awareness: From Research to Practice*. New York, NY: Guilford Press.

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2014). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods*, 46(2), 517–525. <https://doi.org/10.3758/s13428-013-0383-5>

- Gonthier, I., Desrochers, A., Thompson, G., & Landry, D. (2009). Normes d'imagerie et de fréquence subjective pour 1,760 mots monosyllabiques de la langue française [Imagery norms and subjective frequency of 1,760 monosyllabic words in the French language]. *Canadian Journal of Experimental Psychology*, 63(2), 139–149.
<https://doi.org/10.1037/a0015386>
- Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(4), 126–135. <https://doi.org/10.1080/19388076709556976>
- Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, 6(2), 133–151.
[https://doi.org/10.1002/\(SICI\)1099-0909\(200004/06\)6:2%3C133::AID-DYS160%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1099-0909(200004/06)6:2%3C133::AID-DYS160%3E3.0.CO;2-A)
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Erlbaum.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. <https://doi.org/10.1177/074193258600700104>
- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1–13). Erlbaum.
- Gough Kenyon, S. M., Palikara, O., & Lucas, R. M. (2018). Explaining reading comprehension in children with developmental language disorder: The importance of elaborative inferencing. *Journal of Speech, Language, and Hearing Research*, 61(10), 2517–2531. https://doi.org/10.1044/2018_JSLHR-L-17-0416
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. <https://doi.org/10.1037/0033-295x.101.3.371>
- Graesser, A. C., McNamara, D. S., & Louwerse, M. (2004). Coh-Metrix: Analysis of text coherence. *Behavior Research Methods*, 36(2), 193–202.
<https://doi.org/10.3758/BF03195564>
- Graesser, A. C., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). Guilford Press.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398. <https://doi.org/10.1111/j.1756-8765.2010.01081.x>

Hall, R., Greenberg, D., Laures-Gore, J., & Pae, H. K. (2014). The relationship between expressive vocabulary knowledge and reading skills for adult struggling readers. *Journal of Research in Reading*, 37(S1), S87–S100. <https://doi.org/10.1111/j.1467-9817.2012.01537.x>

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hamilton, S. T., Freed, E. M., & Long, D. L. (2016). Word-decoding skill interacts with working memory capacity to influence inference generation during reading. *Reading Research Quarterly*, 51(4), 391–402. <https://doi.org/10.1002/rrq.148>

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing*, 16(6), 505–539. <https://doi.org/10.1023/A:1025521722900>

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes.

Hemforth, B., Mertins, B., & Fabricius-Hansen, C. (Eds.). (2014). *Psycholinguistic approaches to meaning and understanding across languages*. Springer.

Henderson, L., Snowling, M., & Clarke, P. (2013). Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*, 17(3), 177–198. <https://doi.org/10.1080/10888438.2011.652721>

Hennessey, N.W. & Kirsner, K. (1999). The role of sub-lexical orthography in naming: A performance and acoustic analysis. *Acta Psychologica*, 103, 125–148.

Herdağdelen, A., & Marelli, M. (2017). Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive Science*, 41(4), 976–995. <https://doi.org/10.1111/cogs.12392>

Hersch, J., & Andrews, S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading*, 16(3), 240–262. <https://doi.org/10.1080/10888438.2011.564244>

- Hessel, A. K., Nation, K., & Murphy, V. A. (2021). Comprehension monitoring during reading: An eye-tracking study with children learning English as an additional language. *Scientific Studies of Reading*, 25(2), 159–178.
- Hickmann, M. (2002). *Children's discourse: Person, space and time across languages*. Cambridge University Press.
- Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 686–713. <https://doi.org/10.1037/0278-7393.28.4.686>
- Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading. *Journal of Educational Psychology*, 97(2), 193–203. [https://doi.org/10.1044/0161-1461\(2005/029\)?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle](https://doi.org/10.1044/0161-1461(2005/029)?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle)
- Hogan, T. P., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing Higher Level Language Skills to Improve Reading Comprehension. *Focus on Exceptional Children*, 44(3). <https://doi.org/10.17161/foec.v44i3.6688>
- Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, 16(3), 199–207. <https://doi.org/10.3109/17549507.2014.904441>
- Hoffman, P., Jones, R. W., & Lambon Ralph, M. A. (2013). Be concrete to be comprehended: Consistent imageability effects in semantic dementia for nouns, verbs, synonyms and associates. *Cortex*, 49(5), 1206–1218. <https://doi.org/10.1016/j.cortex.2012.05.007>
- Hoover, W. A. (2024). The simple view of reading and its broad types of reading difficulties. *Reading and Writing*, 37(9), 2277–2298. <https://doi.org/10.1007/s11145-023-10471-x>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>

- Hoover, W. A., & Tunmer, W. E. (2018). The Simple View of Reading: Three Assessments of Its Adequacy. *Remedial and Special Education*, 39(5), 304–312.
<https://doi.org/10.1177/0741932518773154>
- Hoover, W. A., & Tunmer, W. E. (2020). *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning*. Springer.
<https://doi.org/10.1007/978-3-030-44195-1>
- Høien-Tengesdal, I., & Høien, T. (2012). The reading efficiency model: An extension of the componential model of reading. *Journal of Learning Disabilities*, 45(5), 467–479.
<https://doi.org/10.1177/0022219411432688>
- Hržica, G., & Kuvač Kraljević, J. (2022). Referential choice in stories with characters of one or different genders: A study of monolingual Croatian speakers. *First Language*, 42(2), 216–233. <https://doi.org/10.1177/0142723720949728>
- Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, 103, 114–126. <https://doi.org/10.1016/j.jml.2018.08.005>
- Huang, Y. (2000). *Anaphora: A cross-linguistic approach*. Oxford University Press.
- Hulme, C., & Snowling, M. J. (Eds.). (2021). *The science of reading: A handbook* (2nd ed.). Wiley-Blackwell. <https://doi.org/10.1002/9781119705116>
- Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948–2959. <https://doi.org/10.1016/j.visres.2009.09.012>
- Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1430–1440. <https://doi.org/10.1037/0278-7393.21.6.1430>
- Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading, driving and scene perception* (pp. 29–54). Elsevier.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431–439.
<https://doi.org/10.3758/bf03208203>

- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4(1), 71–115. [https://doi.org/10.1016/s0364-0213\(81\)80005-5](https://doi.org/10.1016/s0364-0213(81)80005-5)
- Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing*, 19(4), 339–361. <https://doi.org/10.1007/s11145-005-4644-2>
- Jones, M. W., Obregón, M., Kelly, M. L., & Branigan, H. P. (2008). Elucidating the component processes involved in dyslexic and non-dyslexic reading fluency: An eye-tracking study. *Cognition*, 109(3), 389–407. <https://doi.org/10.1016/j.cognition.2008.10.005>
- Joseph, H. S. S. L., & Livversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLOS ONE*, 8(1), e54141. <https://doi.org/10.1371/journal.pone.0054141>
- Joseph, H. S. S. L., Nation, K., & Livversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, 42(2), 207–222.
- Joseph, H. S., Bremner, G., Livversedge, S. P., & Nation, K. (2015). Working memory, reading ability and the effects of distance and typicality on anaphor resolution in children. *Journal of Cognitive Psychology*, 27(5), 622–639. <https://doi.org/10.1080/20445911.2015.1005095>
- Joseph, H., Wonnacott, E., & Nation, K. (2021). Online inference making and comprehension monitoring in children during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, 74(7), 1202–1224. <https://doi.org/10.1177/1747021821999007>
- Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21(2), 85–97. <https://doi.org/10.1080/02702710050084428>
- Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1312–1318. <https://doi.org/10.1037/0278-7393.29.6.1312>
- Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7–8), 846–863. <https://doi.org/10.1080/13506280544000075>

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 39–95). MIT Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>

Just, M. A., & Carpenter, P. A. (2018). Using eye fixations to study reading comprehension. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 151–182). Routledge. <https://doi.org/10.4324/9780429505379-8>

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes*, 40(3), 239–257. https://doi.org/10.1207/s15326950dp4003_4

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1561–1566. <https://doi.org/10.1037/a0020693>

Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*, 51(4), 337–347. <https://doi.org/10.1044/jshd.5104.337>

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-3](https://doi.org/10.1016/S0749-596X(03)00023-3)

Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia*, 56, 51–82. <https://doi.org/10.1007/s11881-006-0003-5>

Katičić, R. (2002). *Sintaksa hrvatskoga književnoga jezika* [Syntax of the Croatian standard language] (3rd ed.). Hrvatska akademija znanosti i umjetnosti; Nakladni zavod Globus.

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives* (pp. 169–186). Lawrence Erlbaum Associates.

- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281–300. <https://doi.org/10.1080/10888430802132279>
- Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. *L'Année Psychologique*, 114(4), 753–777. <https://doi.org/10.4074/S0003503314004072>
- Kelić, M. (2017). *Phonological representations and phonological skills in children with dyslexia in Croatian* [Unpublished doctoral dissertation]. University of Zagreb.
- Kelić, M., Zelenika, M., & Kuvač Kraljević, J. (2016). Predictors of reading words and pseudowords in highly transparent orthography. In *5th All European Dyslexia Conference* (pp. 75–75). University of Modena and Reggio Emilia.
- Kelić, M., Zelenika Zeba, M., & Kuvač Kraljević, J. (2021). Reading predictors in Croatian: Contribution of (meta)phonological variables. *Psiholgijske Teme*, 30(2), 161–184. <https://doi.org/10.31820/pt.30.2.1>
- Kelso, K., Fletcher, J., & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language & Communication Disorders*, 42(1), 39–57. <https://doi.org/10.1080/13682820600693013>
- Kelso, K., Whitworth, A., Parsons, R., & Leitão, S. (2020). Hidden reading difficulties: Identifying children who are poor comprehenders. *Learning Disability Quarterly*, 1–12. Advance online publication. <https://doi.org/10.1177/0731948720961766>
- Kempe, C., Eriksson-Gustavsson, A., & Samuelsson, S. (2011). Are there any Matthew effects in literacy and cognitive development? *Scandinavian Journal of Educational Research*, 55(2), 181–196. <https://doi.org/10.1080/00313831.2011.554699>
- Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2015). Reading comprehension and PASS theory. In T. C. Papadopoulos, R. K. Parrila, & J. R. Kirby (Eds.), *Cognition, intelligence, and achievement* (pp. 117–136). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-410388-7.00007-5c>
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language

and decoding skills. *Journal of Educational Psychology*, 101(4), 765–778.

<https://doi.org/10.1037/a0015956>

Keresteš, G., Brkovic, I., Siegel, L. S., Tjus, T., & Hjelmquist, E. (2019). Literacy development beyond early schooling: A 4-year follow-up study of Croatian. *Reading and Writing*, 32(8), 1955–1988. <https://doi.org/10.1007/s11145-018-9931-9>

Keresteš, G., Hjelmquist, E., Veisson, M., & Siegel, L. S. (2024). The importance of fluency in reading: A comparison of English, Swedish, Croatian, and Estonian. *Reading Psychology*, 45(4), 413–430. <https://doi.org/10.1080/02702711.2024.2319576>

Kershaw, S., & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing*, 25(2), 433–464. <https://doi.org/10.1007/s11145-010-9278-3>

Kim, Y. G., Quinn, J. M., & Petscher, Y. (2021). What is text reading fluency and is it a predictor or an outcome of reading comprehension? A longitudinal investigation. *Developmental Psychology*, 57(5), 718–732. <https://doi.org/10.1037/dev0001167>

Kim, Y. S. G., Park, C., & Park, Y. (2015). Dimensions of discourse level oral language skills and their relation to reading comprehension and written composition: An exploratory study. *Reading and Writing*, 28(5), 633–654. <https://doi.org/10.1007/s11145-015-9542-7>

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42(2), 75–82. <https://doi.org/10.1111/j.1741-4369.2008.00487.x>

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, 95(3), 453–464. <https://doi.org/10.1037/0022-0663.95.3.453>

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25(4), 463–492. <https://doi.org/10.1080/02643290701674851>

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 21–38). Rowman & Littlefield Education.

- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Blackwell. <https://doi.org/10.1002/9780470757642.ch12>
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1–2), 262–284. <https://doi.org/10.1080/09541440340000213>
- Kordić, S. (1996). Zamjenice u izgradnji kohezije teksta. *Radovi Zavoda za slavensku filologiju*, 30–31, 55–100.
- Koren, I. (2019). *Subject pronoun interpretation in very young Croatian monolingual children* [Master's thesis, University of Rijeka]. Repository of the Faculty of Humanities and Social Sciences Rijeka. <https://repository.ffri.uniri.hr/object/ffri:2148>
- Košutar, S. (2022). *Razrješavanje dvosmislenosti anaforične lične zamjenice* [Disambiguation of the ambiguous personal anaphoric pronoun] [Doctoral dissertation, University of Zagreb]. Repository of the Faculty of Humanities and Social Sciences Zagreb. <https://repositorij.ffzg.unizg.hr/object/ffzg:7056>
- Kraš, T. (2008). Anaphora resolution in near-native Italian grammars: Evidence from native speakers of Croatian. *EUROSLA Yearbook*, 8(1), 107–134. <https://doi.org/10.1075/eurosla.8.08kra>
- Kuvač Kraljević, J., & Olujić, M. (2018). Imageability and subjective frequency of the 500 rated nouns in the Croatian Lexical Database. *Suvremena Lingvistika*, 44(85), 73–90. <https://doi.org/10.22210/suvlin>
- Kuvač Kraljević, J., Runje, N., Ružić, V., Škorić, A. M., Lenček, M., & Štefanec, A. (2024). Predictors of reading comprehension and profiling of poor readers in Croatian: Educational and clinical perspectives. *Frontiers in Psychology*, 15, Article 1297183. <https://doi.org/10.3389/fpsyg.2024.1297183>

Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass*, 11(2), Article e12234.

<https://doi.org/10.1111/lnc3.12234>

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>

Language and Reading Research Consortium. (2015). Learning to read: Should we keep it simple? *Reading Research Quarterly*, 50(2), 151–169. <https://doi.org/10.1002/rrq.99>

Language and Reading Research Consortium, & Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education*, 39(5), 289–303.

<https://doi.org/10.1177/0741932518762055>

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye movements during reading. *Neuroscience & Biobehavioral Reviews*, 30(6), 872–884.

<https://doi.org/10.1016/j.neubiorev.2006.06.013>

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95(2), 211–224. <https://doi.org/10.1037/0022-0663.95.2.211>

Lenček, M. (1994). Diskriminativna vrijednost nekih jezičnih zadataka za teškoće čitanja [Discriminative value of some language tasks for reading difficulties]. *Defektologija*, 30(2), 115–126.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1–3), 1–22. [https://doi.org/10.1016/0010-0277\(92\)90038-j](https://doi.org/10.1016/0010-0277(92)90038-j)

Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–75.

<https://doi.org/10.1017/s0140525x99001776>

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, P. (2005). Development of and relationship between phonological and motivational processes and naming speed in

predicting word recognition in grade 1. *Scientific Studies of Reading*, 9(4), 367–399.

https://doi.org/10.1207/s1532799xssr0904_3

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly*, 47(3), 259–282.

<https://doi.org/10.1002/rrq.020>

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the developmental relationship between oral language skills and reading comprehension. *Psychological Bulletin*, 143(4), 347–375. <https://doi.org/10.1037/bul0000095>

Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of Binding and Control phenomena. *Journal of Linguistics*, 23(2), 379–434.

<https://doi.org/10.1017/S0022226700011324>

Li, J. J., Cutting, L. E., Ryan, M., Zilioli, M., Denckla, M. B., & Mahone, E. M. (2009). Response variability in rapid automatized naming predicts reading comprehension. *Journal of Clinical and Experimental Neuropsychology*, 31(7), 877–888.

<https://doi.org/10.1080/13803390802646973>

Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B. H. (2015). Norwegian words: A lexical database for clinicians and researchers. *Clinical Linguistics & Phonetics*, 29(4), 276–290. <https://doi.org/10.3109/02699206.2014.999952>

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading times. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–76). Elsevier.

Lorch, R. F. (2015). What about expository text? In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 348–361). Cambridge University Press.

<https://doi.org/10.1017/CBO9781107279186.016>

Luke, S. G., Henderson, J. M., & Ferreira, F. (2015). Children's eye-movements during reading reflect the quality of lexical representations: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1675–1683.

<https://doi.org/10.1037/xlm0000133>

- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1–14. <https://doi.org/10.1007/s11881-003-0001-9>
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions and technologies* (pp. 107–136). Psychology Press.
- Marschark, M., & Cornoldi, C. (1991). Imagery and verbal memory. In C. Cornoldi & M. A. McDaniel (Eds.), *Imagery and cognition* (pp. 133–182). Springer.
https://doi.org/10.1007/978-1-4684-6407-8_5
- Martens, V. E., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, 98(2), 140–149.
<https://doi.org/10.1016/j.bandl.2006.04.003>
- Massonnié, J., Bianco, M., Lima, L., & Bressoux, P. (2019). Longitudinal predictors of reading comprehension in French at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction*, 60, 166–179.
<https://doi.org/10.1016/j.learninstruc.2018.01.005>
- Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three- and five-year-old children. *Journal of Child Language*, 35(2), 373–402.
<https://doi.org/10.1017/S0305000907008549>
- Matić, A. (2020). *Psiholingvistički pristup strukturno-značenjskim čimbenicima u obradi odnosnih rečenica* [Psycholinguistic approach to structural-semantic factors in processing relative clauses] [Doctoral dissertation, University of Zagreb]. Repozitorij Filozofskog fakulteta Sveučilišta u Zagrebu. <https://urn.nsk.hr/urn:nbn:hr:131:711004>
- McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry*, 41(7), 869–874.
- 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., 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>
- McMaster, K. L., Espin, C. A., & van den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research & Practice*, 29(1), 17–24.
- Megherbi, H., & Ehrlich, M. F. (2005). Language impairment in less skilled comprehenders: The on-line processing of anaphoric pronouns in a listening situation. *Reading and Writing*, 18(7), 715–753. <https://doi.org/10.1007/s11145-005-8131-6>
- Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, 30(4), 551–561. <https://doi.org/10.3758/BF03194956>
- Mézière, D. C., Yu, L., Reichle, E. D., von der Malsburg, T., & McArthur, G. (2023). Using eye-tracking measures to predict reading comprehension. *Reading Research Quarterly*, 58(3), 425–449. <https://doi.org/10.1002/rrq.498>
- Mezynski, K. (1983). Issues Concerning the Acquisition of Knowledge: Effects of Vocabulary Training on Reading Comprehension. *Review of Educational Research*, 53(2), 253-279.
- Míguez-Álvarez, C., Cuevas-Alonso, M. and Saavedra, Á. (2022), Relationships Between Phonological Awareness and Reading in Spanish: A Meta-Analysis. *Language Learning*, 72: 113-157. <https://doi.org/10.1111/lang.12471>
- Milledge, S. V., & Blythe, H. I. (2019). The changing role of phonology in reading development. *Vision*, 3(2), Article 23. <https://doi.org/10.3390/vision3020023>
- Mitkov, R. (2002). *Anaphora resolution*. Longman.
- Moats, L. C. (2008). *Speech to print: Language essentials for teachers* (2nd ed.). Brookes.
- Moats, L. C., & Dakin, K. E. (2008). *Basic facts about dyslexia and other reading problems*. International Dyslexia Association.

- Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing: Evidence from the asymmetric German orthography. *Scientific Studies of Reading*, 13(1), 1–25. <https://doi.org/10.1080/10888430802631684>
- Monster, I., Tellings, A., Burk, W. J., Keuning, J., Segers, E., & Verhoeven, L. (2022). Word properties predicting children's word recognition. *Scientific Studies of Reading*, 26(5), 373–389. <https://doi.org/10.1080/10888438.2021.2020795>
- Mott, M. S., & Rutherford, A. S. (2012). Technical examination of a measure of phonological sensitivity. *SAGE Open*, 2(2). <https://doi.org/10.1177/2158244012450447>
- Müller, K., & Brady, S. (2001). Correlates of early reading performance in a transparent orthography. *Reading and Writing*, 14(7–8), 757–799. <https://doi.org/10.1023/A:1012217704834>
- Murray, L., & Coppins, P. (2013). Formal and informal assessment of aphasia. In I. Papathanasiou, P. Coppins, & C. Potagas (Eds.), *Aphasia and related neurogenic communication disorders* (pp. 66–91). Jones & Bartlett Publishers.
- Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology Section A*, 40(4), 693–718.
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40(5), 665–681. <https://doi.org/10.1037/0012-1649.40.5.665>
- Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 248–266). Blackwell.
- Nation, K. (2013). Lexical learning and lexical processing in children with developmental language impairments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), Article 20120387. <https://doi.org/10.1098/rstb.2012.0387>
- Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *npj Science of Learning*, 2, Article 3. <https://doi.org/10.1038/s41539-017-0004-7>

Nation, K., & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly*, 32(2), 154–167.
<https://doi.org/10.1598/RRQ.32.2.2>

Nation, K., & Norbury, C. F. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders*, 25(1), 21–32.
<https://doi.org/10.1097/00011363-200501000-00004>

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, 69(4), 996–1011. <https://doi.org/10.1111/j.1467-8624.1998.tb06157.x>

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, 70(1), B1–B13. [https://doi.org/10.1016/s0010-0277\(99\)00004-9](https://doi.org/10.1016/s0010-0277(99)00004-9)

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>

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language and Cognitive Processes*, 16(2–3), 241–259.
<https://doi.org/10.1080/01690960042000003>

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research*, 47(1), 199–211.
[https://doi.org/10.1044/1092-4388\(2004/017\)](https://doi.org/10.1044/1092-4388(2004/017))

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of child psychology and psychiatry, and allied disciplines*, 51(9), 1031–1039.
<https://doi.org/10.1111/j.1469-7610.2010.02254.x>

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment*. National Institute of Child Health and Human Development.

- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13(1), 45–52. <https://doi.org/10.3758/BF03193811>
- Nicolielo-Carrilho, A. P., Crenitte, P. A. P., Lopes-Herrera, S. A., & Hage, S. R. V. (2018). Relationship between phonological working memory, metacognitive skills and reading comprehension in children with learning disabilities. *Journal of applied oral science : revista FOB*, 26, e20170414. <https://doi.org/10.1590/1678-7757-2017-0414>
- Nobre, N. (2011). *Anaphora resolution* [Unpublished master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa].
- Norberg, K. A., Perfetti, C., & Helder, A. (2022). Word-to-text integration and antecedent accessibility: Eye-tracking evidence extends results of event-related potentials (ERPs). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(4), 598–617. <https://doi.org/10.1037/xlm0001127>
- 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, and allied disciplines*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, 63, 427–452. <https://doi.org/10.1146/annurev-psych-120710-100431>
- Nouwens, S., Groen, M. A., & Verhoeven, L. (2017). How working memory relates to children's reading comprehension: The importance of domain-specificity in storage and processing. *Reading and Writing*, 30(1), 105–120. <https://doi.org/10.1007/s11145-016-9665-5>
- Oakhill, J. (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology*, 73(1), 13–20. <https://doi.org/10.1111/j.2044-8295.1982.tb01785.x>
- Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54(1), 31–39. <https://doi.org/10.1111/j.2044-8279.1984.tb00842.x>

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16(2), 91–121.

<https://doi.org/10.1080/10888438.2010.529219>

Oakhill, J., & Kyle, F. (2000). The relation between phonological awareness and working memory. *Journal of Experimental Child Psychology*, 75(2), 152–164.

<https://doi.org/10.1006/jecp.1999.2529>

Ocal, T., Durgunoglu, A., & Twite, L. (2022). Reading from screen vs reading from paper: Does it really matter? *Journal of College Reading and Learning*, 52(2), 130–148.

<https://doi.org/10.1080/10790195.2022.2028593>

Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566.

<https://doi.org/10.1037/0022-0663.98.3.554>

Ozernov-Palchik, O., Centanni, T. M., Beach, S. D., May, S., Hogan, T., & Gabrieli, J. (2021). Distinct neural substrates of individual differences in components of reading comprehension in adults with or without dyslexia. *NeuroImage*, 226, Article 117570.

<https://doi.org/10.1016/j.neuroimage.2020.117570>

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76(1, Pt. 2), 1–25. <https://doi.org/10.1037/h0025327>

Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, 30(1), 1–31. <https://doi.org/10.1080/10573569.2013.758932>

Palmović, M., Matic, A., & Kovačević, M. (2018). Resolution of anaphoric expressions in children and adults: Evidence from eye movements. *Suvremena Lingvistika*, 44(85), 139–154. <https://doi.org/10.22210/suvlin.2018.085.13>

Papadopoulou, D., Peristeri, E., Plemenou, E., Marinis, T., & Tsimpli, I. M. (2015). Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua*, 155, 98–120. <https://doi.org/10.1016/j.lingua.2014.12.006>

- Pérez, A. I., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive Processing*, 15(3), 405–413.
<https://doi.org/10.1007/s10339-014-0611-7>
- Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford University Press.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C. A. (2017). Lexical quality revisited. In E. Segers & P. van den Broek (Eds.), *Developmental perspectives in written language and literacy: In honor of Ludo Verhoeven* (pp. 51–68). John Benjamins.
- Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection* (pp. 67–86). American Psychological Association.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–214). John Benjamins.
- Perfetti, C., & Helder, A. (2021). Incremental comprehension examined in event-related potentials: Word-to-text integration and structure building. *Discourse Processes*, 58(1), 2–21.
<https://doi.org/10.1080/0163853X.2020.1743806>
- Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 269–297). Routledge.
- Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Perfetti, C. A., van Dyke, J., & Hart, L. (2001). The psycholinguistics of basic literacy. *Annual Review of Applied Linguistics*, 21, 127–149.
<https://doi.org/10.1017/S0267190501000083>

Perfetti, C., Yang, C.-L., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, 22(3), 303–318.

<https://doi.org/10.1002/acp.1419>

Perfetti, C. A., Stafura, J., & Adlof, S. M. (2013). Reading comprehension and reading comprehension problems: A word-to-text integrative perspective. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 22–32). Paul Brookes Publishing.

Petersen, D. K., & Elbro, C. (2009). *Reading comprehension: Contributions from size and semantic organization of the mental lexicon*. Paper presented at the Society for the Scientific Study of Reading, Boston, MA.

Peti-Stantić, A. (2019). *Čitanjem do (spo)razumijevanja: Od čitalačke pismenosti do čitateljske sposobnosti* [The road to comprehension: Developing reading literacy]. Ljevak.

Peti-Stantić, A., Anđel, M., Keresteš, G., Ljubešić, N., Stanojević, M. M., & Tonković, M. (2018). Psiholingvističke mjere ispitivanja 3.000 riječi hrvatskoga jezika: Konkretnost i predočivost [Psycholinguistic estimates of 3000 words of Croatian: concreteness and imageability]. *Suvremena Lingvistika*, 44(85), 91–111. <https://doi.org/10.22210/suvlin>

Peti-Stantić, A., Anđel, M., Keresteš, G., Ljubešić, N., Stanojević, M.-M., Tonković, M., & Tušek, J. (2018). *Hrvatska psiholingvistička baza (HPB)* [Croatian psycholinguistic database] [Database]. MEGAHR.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, 53(4), 323–334. <https://doi.org/10.1037/h0087320>

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1252–1270. <https://doi.org/10.1037/0278-7393.30.6.1252>

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15(1), 161–167.

<https://doi.org/10.3758/pbr.15.1.161>

- Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 2, pp. 24–43). Psychology Press.
<https://doi.org/10.4324/9780203106976>
- Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 726–750. <https://doi.org/10.1037/0096-1523.34.3.726>
- Potocki, A., Ecalle, J., & Magnan, A. (2013). Narrative comprehension skills in 5-year-old children: Correlational analysis and comprehender profiles. *Journal of Educational Research*, 106, 14–26. <https://doi.org/10.1080/00220671.2012.667013>
- Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing*, 25(7), 1635–1664. <https://doi.org/10.1007/s11145-011-9336-5>
- Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, 33(3), 217–240. <https://doi.org/10.1080/02702711.2010.507626>
- Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the simple view of reading. *Reading & Writing Quarterly*, 29(2), 168–202. <https://doi.org/10.1080/10573569.2013.758569>
- Pufpaff, L. A. (2011). Adapted assessment of phonological sensitivity skills: A preliminary investigation. *Communication Disorders Quarterly*, 33(1), 13–22.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 282–308.
<https://doi.org/10.3138/cmlr.56.2.282>
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513–536. <https://doi.org/10.1111/1467-9922.00193>
- Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during

- comprehension. *Language Learning*, 61(1), 187–218. <https://doi.org/10.1111/j.1467-9922.2010.00592.x>
- Ramus, F. (2001). Outstanding questions about phonological processing in dyslexia. *Dyslexia*, 7(4), 197–216. <https://doi.org/10.1002/dys.205>
- Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13(2), 212–218. [https://doi.org/10.1016/s0959-4388\(03\)00035-7](https://doi.org/10.1016/s0959-4388(03)00035-7)
- Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences*, 27(12), 720–726. <https://doi.org/10.1016/j.tins.2004.10.004>
- Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, 29(1–2), 104–122. <https://doi.org/10.1080/02643294.2012.677420>
- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, 126(4), 841–865. <https://doi.org/10.1093/brain/awg076>
- Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, 136(2), 630–645. <https://doi.org/10.1093/brain/aws356>
- Raven, J., Raven, J. C., & Court, J. H. (1999). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford Psychologists Press.
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41(2), 211–236. [https://doi.org/10.1016/0022-0965\(86\)90037-8](https://doi.org/10.1016/0022-0965(86)90037-8)
- Rayner, K. (1998). Eye movements in reading and information processing. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201. <https://doi.org/10.3758/BF03197692>

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *WIREs Cognitive Science*, 1(6), 787–799. <https://doi.org/10.1002/wcs.68>

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4), 504–509. <https://doi.org/10.3758/BF03214555>

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 720–732. <https://doi.org/10.1037/0096-1523.30.4.720>

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2006). *Psychology of reading* (2nd ed.). Psychology Press.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 514–528. <https://doi.org/10.1037/a0020990>

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146–161. <https://doi.org/10.1017/S0267190504000078>

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, 7(3), 219–237. https://doi.org/10.1207/S1532799XSSR0703_2

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, 105(1), 125–157. <https://doi.org/10.1037/0033-295x.105.1.125>

Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, 16(3), 187–217. <https://doi.org/10.1080/10888438.2010.538780>

- Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, 17(6), 415–434. <https://doi.org/10.1080/10888438.2013.764879>
- Ricketts, J. (2011). Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry*, 52(11), 1111–1123. <https://doi.org/10.1111/j.1469-7610.2011.02438.x>
- Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *Journal of Experimental Child Psychology*, 150, 330–345. <https://doi.org/10.1016/j.jecp.2016.05.013>
- Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language*, 71(2), 267–284. <https://doi.org/10.1006/brln.1999.2166>
- Rosenblatt, L. M. (2018). The transactional theory of reading and writing. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 35–63). Routledge.
- Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A Longitudinal Analysis of the Connection Between Oral Language and Early Reading. *The Journal of Educational Research*, 95(5), 259–272. <https://doi.org/10.1080/00220670209596600>
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI* (pp. 573–603). Erlbaum.
- Rønberg, L. F.-J., & Petersen, D. K. (2016). It matters whether reading comprehension is conceptualised as rate or accuracy. *Journal of Research in Reading*, 39(2), 209–228. <https://doi.org/10.1111/1467-9817.12047>
- Sabatini, J., O'Reilly, T., & Deane, P. (2013). Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design. *ETS Research Report Series*, 2013, i–50. <https://doi.org/10.1002/j.2333-8504.2013.tb02337.x>
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2019). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, 20(4), 1–23. <https://doi.org/10.1002/9781119261407.ch1>
- Sadoski, M., & Paivio, A. (2004). *A dual coding theoretical model of reading*. Routledge.

- Safir, K. J. (2004). *The syntax of anaphora*. Oxford University Press.
- Sánchez-Vincitore, L. V., Veras, C., Mencía-Ripley, A., Ruiz-Matuk, C. B., & Cubilla-Bonnetier, D. (2022). Reading comprehension precursors: Evidence of the simple view of reading in a transparent orthography. *Frontiers in Education*, 7, Article 914414. <https://doi.org/10.3389/feduc.2022.914414>
- Sauzéon, H., Lestage, P., Raboutet, C., N'Kaoua, B., & Claverie, B. (2004). Verbal fluency output in children aged 7–16 as a function of the production criterion: Qualitative analysis of clustering, switching processes, and semantic network exploitation. *Brain and Language*, 89(1), 192–202. [https://doi.org/10.1016/S0093-934X\(03\)00367-5](https://doi.org/10.1016/S0093-934X(03)00367-5)
- Savage, R. (2001). The "simple view" of reading: Some evidence and possible implications. *Educational Psychology in Practice*, 17(1), 17–33. <https://doi.org/10.1080/02667360120039951>
- Savage, R., Pillay, V., & Melidona, S. (2007). Deconstructing rapid automatized naming: Component processes and the prediction of reading difficulties. *Learning and Individual Differences*, 17(2), 129–146. <https://doi.org/10.1016/j.lindif.2007.04.001>
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading disabilities. In S. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). Guilford Press.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26(6), 1270–1281. <https://doi.org/10.3758/BF03201199>
- Schindler, J., & Richter, T. (2018). Reading comprehension: Individual differences, disorders, and underlying cognitive processes. In E. Dattner & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 503–524). De Gruyter Mouton. <https://doi.org/10.1515/9781614514909-025>
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6(4), 147–151. <https://doi.org/10.1027/1614-2241/a000016>

- Schroeder, S., Hyönä, J., & Liversedge, S. P. (2015). Developmental eye-tracking research in reading: Introduction to the special issue. *Journal of Cognitive Psychology*, 27(5), 500–510. <https://doi.org/10.1080/20445911.2015.1046877>
- Schroyens, W., Vitu, F., Brysbaert, M., & d'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *Quarterly Journal of Experimental Psychology Section A*, 52(4), 1021–1046. <https://doi.org/10.1080/713755859>
- Schwanenflugel, P. J., Akin, C., & Luh, W. M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, 20(1), 96–104. <https://doi.org/10.3758/BF03208259>
- Sedivy, J. (2010). Using eyetracking in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 115–138). John Benjamins. <https://doi.org/10.1075/llt.27.08sed>
- Selin, C. M., Rice, M. L., Girolamo, T. M., & Wang, C. J. (2022). Work setting effects on speech-language pathology practice: Implications for identification of children with specific language impairment. *American Journal of Speech-Language Pathology*, 31(2), 854–880. https://doi.org/10.1044/2021_AJSLP-21-00024
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. <https://doi.org/10.1348/000712603321661859>
- Shankweiler, D., Lundquist, E., Dreyer, L. G., & Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing*, 8(3), 267–294. <https://doi.org/10.1007/BF00420279>
- Silić, J. (1984). *Od rečenice do teksta: Teoretsko-metodološke pretpostavke nadrečeničnog jedinstva* [From sentence to text: Theoretical and methodological assumptions of supra-sentential units]. Sveučilišna naklada Liber.
- Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, 107(2), 321–331. <https://doi.org/10.1037/a0037769>

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading*, 17(2), 108–133.

<https://doi.org/10.1080/10888438.2011.618153>

Silliman, E. R., & Berninger, V. W. (2011). Cross-disciplinary dialogue about the nature of oral and written language problems in the context of developmental, academic, and phenotypic profiles. *Topics in Language Disorders*, 31(1), 6–23.

<https://doi.org/10.1097/TLD.0b013e31820a0b5b>

Singer, M., Andruslak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition*, 20(5), 539–548.

<https://doi.org/10.3758/BF03199586>

Skarakis-Doyle, E., & Dempsey, L. (2008). The detection and monitoring of comprehension errors by preschool children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 51(5), 1227–1243. [https://doi.org/10.1044/1092-4388\(2008/07-0136\)](https://doi.org/10.1044/1092-4388(2008/07-0136))

Snow, Catherine, Reading for Understanding: Toward an R&D Program in Reading Comprehension. Santa Monica, CA: RAND Corporation, 2002.

https://www.rand.org/pubs/monograph_reports/MR1465.html.

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

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, 57(12), 1360–1369.

<https://doi.org/10.1111/jcpp.12497>

Soares, A. P., Costa, A. S., Machado, J., Comesaña, M., & Oliveira, H. M. (2017). The Minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, 49(3), 1065–1081.

<https://doi.org/10.3758/s13428-016-0767-4>

Song, M. K., Lin, F. C., Ward, S. E., & Fine, J. P. (2013). Composite variables: When and how. *Nursing Research*, 62(1), 45–49. <https://doi.org/10.1097/NNR.0b013e3182741948>

Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119(3), 460–477. <https://doi.org/10.1016/j.lingua.2008.09.008>

Spencer, M., & Wagner, R. K. (2018). The comprehension problems of children with poor reading comprehension despite adequate decoding: A meta-analysis. *Review of Educational Research*, 88(3), 366–400. <https://doi.org/10.3102/0034654317749187>

Stanovich, K. E. (1986). Matthew effects in reading. *Reading Research Quarterly*, 21(4), 360–407. <https://doi.org/10.1598/RRQ.21.4.1>

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311–327. <https://doi.org/10.1111/lnc3.12151>

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198568971.013.0019>

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://doi.org/10.1044/jslhr.4102.407>

Sukthanker, R., Poria, S., Cambria, E., & Thirunavukarasu, R. (2020). Anaphora and coreference resolution: A review. *Information Fusion*, 59, 139–162. <https://doi.org/10.1016/j.inffus.2020.01.010>

Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities*, 42(3), 260–287. <https://doi.org/10.1177/0022219409331958>

Swart, N. M., Muijselaar, M. M. L., Steenbeek-Planting, E. G., Droop, M., de Jong, P. F., & Verhoeven, L. (2017). Differential lexical predictors of reading comprehension in fourth graders. *Reading and Writing*, 30(3), 489–507. <https://doi.org/10.1007/s11145-016-9686-0>

Škoro, K. (2024). *Osjetljivost PPVT-a i TROG-a na razvojni jezični poremećaj* [Sensitivity of PPVT and TROG to developmental language disorder] [Master's thesis, Edukacijsko-rehabilitacijski fakultet, Sveučilište u Zagrebu].

Tallal, P. (2003). Language learning disabilities: Integrating research approaches. *Current Directions in Psychological Science*, 12(6), 206–211. <https://doi.org/10.1046/j.0963-7214.2003.01262.x>

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10(4), 381–398. https://doi.org/10.1207/s1532799xssr1004_3

Tighe, E. L., & Schatschneider, C. (2014). A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. *Reading and Writing*, 27(1), 101–127. <https://doi.org/10.1007/s11145-013-9435-6>

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, 113, 33–43. <https://doi.org/10.1016/j.visres.2015.05.008>

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32(4), 383–401. <https://doi.org/10.1111/j.1467-9817.2009.01401.x>

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, 28(7), 939–957. <https://doi.org/10.1007/s11145-015-9556-1>

Tomasello, M. (2007). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, 49(6), 1193–1208. [https://doi.org/10.1044/1092-4388\(2006/086\)](https://doi.org/10.1044/1092-4388(2006/086))

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>

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing*, 26(3), 403–429. <https://doi.org/10.1007/s11145-012-9374-7>

- Tong, X., Tong, X., Shu, H., Chan, S., & McBride-Chang, C. (2014). Discourse-level reading comprehension in Chinese children: What is the role of syntactic awareness? *Journal of Research in Reading*, 37(S1), S48–S70. <https://doi.org/10.1111/1467-9817.12016>
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. (1999). *Test of word reading efficiency*. Pro-Ed.
- Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities*, 43(4), 308–321.
- Torppa, M., Georgiou, G., Salmi, P., Eklund, K., & Lyytinen, H. (2012). Examining the double-deficit hypothesis in an orthographically consistent language. *Scientific Studies of Reading*, 16(4), 287–315. <https://doi.org/10.1080/10888438.2011.554470>
- Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly*, 62(2), 179–206. <https://doi.org/10.13110/merrpalmquar1982.62.2.0179>
- Trtanj, I. (2015). *Jezično označavanje likova u dječjem pripovjednom diskursu* [Linguistic reference to characters in children's narrative discourse] [Doctoral dissertation, Faculty of Humanities and Social Sciences, Josip Juraj Strossmayer University of Osijek].
- Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, 22(3), 370–378. [https://doi.org/10.1076/1380-3395\(200006\)22:3;1-V;FT370](https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370)
- Tunmer, W. E., & Chapman, J. W. (2007). Language-related differences between discrepancy-defined and non-discrepancy-defined poor readers: A longitudinal study of dyslexia in New Zealand. *Dyslexia*, 13(1), 42–66. <https://doi.org/10.1002/dys.327>
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45(5), 453–466. <https://doi.org/10.1177/0022219411432685>
- Tunmer, W. E., & Greaney, K. (2010). Defining dyslexia. *Journal of Learning Disabilities*, 43(3), 229–243. <https://doi.org/10.1177/0022219409345009>

- Tunmer, W. E., & Hoover, W. A. (2019). The cognitive foundations of learning to read: A framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties*, 24(1), 75–93. <https://doi.org/10.1080/19404158.2019.1614081>
- Tunmer, W. E., Nesdale, A. R., & Wright, A. D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology*, 5(1), 25–34. <https://doi.org/10.1111/j.2044-835X.1987.tb01038.x>
- Turkyılmaz, M., Can, R., Yildirim, K., & Ateş, S. (2014). Relations among oral reading fluency, silent reading fluency, retell fluency, and reading comprehension. *Procedia: Social and Behavioral Sciences*, 116, 4030–4034. <https://doi.org/10.1016/j.sbspro.2014.01.885>
- Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, 33(2), 213–220. <https://doi.org/10.3758/BF03195310>
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39(2–3), 299–316. <https://doi.org/10.1080/0163853X.2005.9651685>
- van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review*, 41(3), 315–325. <https://doi.org/10.1080/02796015.2012.12087512>
- van der Lely, H. K. (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 der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A. H. A., & van Lieshout, E. C. D. M. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, 101(1), 21–36. <https://doi.org/10.1037/a0013382>
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Van Dyke, J. A. (2021). Introduction to the special issue: Mechanisms of variation in reading comprehension: Processes and products. *Scientific Studies of Reading*, 25(2), 93–103. <https://doi.org/10.1080/10888438.2021.1873347>

- Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *WIREs Cognitive Science*, 4(2), 125–134.
<https://doi.org/10.1002/wcs.1209>
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2–40. <https://doi.org/10.1046/j.0021-9630.2003.00305.x>
- Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11(1), 3–32. <https://doi.org/10.1080/10888430709336632>
- Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407–423.
<https://doi.org/10.1002/acp.1414>
- Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, 25(8), 1805–1818.
<https://doi.org/10.1007/s11145-011-9346-3>
- Verhoeven, L., Voeten, M., & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics*, 50, 28–36. <https://doi.org/10.1016/j.jneuroling.2018.03.002>
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217–234.
<https://doi.org/10.1017/S0142716401002041>
- Wagner, R. K., & Meros, D. (2010). Vocabulary and reading comprehension: Direct, indirect, and reciprocal influences. *Focus on Exceptional Children*, 43(1), 1–8.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192–212.
<https://doi.org/10.1037/0033-2909.101.2.192>
- Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, 85(1), 83–103. <https://doi.org/10.1037/0022-0663.85.1.83>

- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, 30(1), 73–87.
<https://doi.org/10.1037/0012-1649.30.1.73>
- Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Relation between decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology*, 111(3), 472–482. <https://doi.org/10.1037/edu0000302>
- Wang, W., Fan, L., Wang, Z., Liu, X., & Zhang, S. (2021). Effects of phonological loop on inferential processing during Chinese text reading: Evidence from a dual-task paradigm. *PsyCh Journal*, 10(4), 521–533. <https://doi.org/10.1002/pchj.451>
- Wechsler, D. (2009). *Wechsler Intelligence Scale for Children* (4th ed.). Pearson.
- Weekes, B. S., Hamilton, S., Oakhill, J. V., & Holliday, R. E. (2008). False recollection in children with reading comprehension difficulties. *Cognition*, 106(1), 222–233.
- Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology*, 71(1), 198–210.
<https://doi.org/10.1080/17470218.2017.1307862>
- Wen, Y., Snell, J., & Grainger, J. (2019). Parallel, cascaded, interactive processing of words during sentence reading. *Cognition*, 189, 221–226.
<https://doi.org/10.1016/j.cognition.2019.04.006>
- Westerveld, M. (2009). Measuring reading comprehension ability in children: Factors influencing test performance. *Acquiring Knowledge in Speech, Language and Hearing*, 11(2), 81–84.
- Westerveld, M. F., Armstrong, R. M., & Barton, G. M. (2020). Reading success. In M. F. Westerveld & G. M. Barton (Eds.), *Reading success in the primary years: An evidence-based interdisciplinary approach to guide assessment and intervention* (pp. 1–17). Springer.
- Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *Quarterly Journal of Experimental Psychology*, 67(6), 1151–1165. <https://doi.org/10.1080/17470218.2013.848216>

- Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology*, 173, 318–337.
<https://doi.org/10.1016/j.jecp.2018.03.014>
- Willingham, D. T. (2017). *The reading mind: A cognitive approach to understanding how the mind reads*. John Wiley & Sons.
- Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities*, 33(4), 387–407.
<https://doi.org/10.1177/002221940003300409>
- Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties*. Springer. <https://doi.org/10.1007/978-94-007-1174-7>
- World Health Organization. (2012). *MKB-10: Međunarodna klasifikacija bolesti i srodnih zdravstvenih problema* [ICD-10: International statistical classification of diseases and related health problems] (10th rev., 2nd ed., Vol. 1). Medicinska naklada.
- Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*, 66(3), 548–562.
<https://doi.org/10.1080/17470218.2011.625094>
- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 55–89. <https://doi.org/10.1037/0278-7393.33.1.55>
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60(4), 502–529. <https://doi.org/10.1016/j.jml.2009.02.001>
- Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin & Review*, 21(2), 526–533.
<https://doi.org/10.3758/s13423-013-0525-x>
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6, Article 72.
<https://doi.org/10.3389/fnhum.2012.00072>

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 597–613.

<https://doi.org/10.1037/xlm0000064>

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1385–1397. <https://doi.org/10.1037/0278-7393.31.6.1385>

Zaretsky, E., Kuvač Kraljević, J., Core, C., & Lenček, M. (2009). Literacy predictors and early reading and spelling skills as a factor of orthography: Cross-linguistic evidence. *Written Language & Literacy*, 12(1), 52–81. <https://doi.org/10.1075/wll.12.1.03zar>

Zhang, J., McBride-Chang, C., Tong, X., Wong, A. M. Y., Shu, H., & Fong, C. Y. C. (2012). Reading with meaning: The contributions of meaning-related variables at the word and subword levels to early Chinese reading comprehension. *Reading and Writing*, 25(9), 2183–2203. <https://doi.org/10.1007/s11145-011-9353-4>

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faisca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, 21(4), 551–559.

<https://doi.org/10.1177/0956797610363406>

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93(3), 369–373. <https://doi.org/10.1016/j.bandl.2004.10.010>

13. APPENDICES

Appendix 1. Descriptive Statistics for Eye-Movement Measures by AOI and Psycholinguistic Condition

	ANAPHORA			
Measure	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
High imageability - high frequency				
First fixation	340,78	139,46	1,85	4,38
First pass	505,46	220,37	2,18	8,77
Dwell time	992,19	469,86	2,03	5,24
Number of fixations	21,54	7,67	1,06	1,14
Number of revisits	7,19	4,07	1,04	0,73
Regression out probability	0,15	0,05	0,54	0,86
High imageability - low frequency				
First fixation	368,76	162,35	1,81	4,53
First pass	692,26	675,09	3,83	16,16
Dwell time	1356,62	864,15	1,64	2,35
Number of fixations	24,94	13,63	1,56	2,18
Number of revisits	8,23	6,50	1,52	2,17
Regression out probability	0,13	0,05	0,34	-0,58
Low imageability - high frequency				
First fixation	343,90	128,32	2,41	8,90

First pass	654,10	695,59	4,33	19,62
Dwell time	1090,88	677,71	1,84	4,00
Number of fixations	21,63	10,68	1,92	5,20
Number of revisits	6,34	4,50	1,14	0,96
Regression out probability	0,14	0,06	0,80	1,09

Low imageability - low frequency

First fixation	355,24	161,77	2,57	10,21
First pass	565,07	447,39	4,44	23,07
Dwell time	1253,70	810,57	1,84	3,75
Number of fixations	24,27	15,66	4,02	22,75
Number of revisits	7,84	7,21	3,09	14,63
Regression out probability	0,14	0,06	1,13	2,56

ANTECEDENT				
Measure	<i>M</i>	<i>SD</i>	Skewness	Kurtosis

High imageability - high frequency

First fixation	343,91	112,41	0,15	0,26
First pass	448,65	205,73	2,90	12,86
Dwell time	716,50	400,52	1,68	3,11
Number of fixations	13,49	6,92	1,94	4,71
Number of revisits	4,33	3,98	2,57	9,97
Regression out probability	0,22	0,10	1,03	2,38

High imageability - low frequency

First fixation	393,99	158,89	1,00	2,75
First pass	531,83	265,54	2,61	9,38
Dwell time	952,75	644,90	1,96	3,72
Number of fixations	15,63	7,15	1,28	1,84
Number of revisits	4,99	3,82	1,04	0,51
Regression out probability	0,20	0,11	2,53	9,97

Low imageability - high frequency

First fixation	338,34	145,35	1,04	1,44
First pass	418,68	193,92	3,26	18,08
Dwell time	718,37	468,72	2,02	4,78
Number of fixations	14,31	6,92	1,09	0,88
Number of revisits	4,59	3,98	1,39	2,12
Regression out probability	0,24	0,14	2,75	12,97

Low imageability - low frequency

First fixation	350,69	157,54	1,43	3,13
First pass	467,38	209,43	2,67	9,67
Dwell time	923,07	544,77	2,07	6,48
Number of fixations	15,50	7,05	1,85	5,84
Number of revisits	5,13	3,83	1,72	3,87
Regression out probability	0,20	0,12	4,47	28,37

Measure	WORDS PRECEDINGS			
	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
High imageability - high frequency				
First fixation	330,16	91,52	1,81	7,54
Dwell time	1317,94	693,14	1,84	3,00
Number of fixations	82,89	36,34	2,55	7,81
Number of revisits	23,23	13,93	1,85	3,97
High imageability - low frequency				
First fixation	344,50	97,96	1,02	4,39
Dwell time	1836,04	968,19	1,42	1,82
Number of fixations	104,37	48,64	2,16	5,67
Number of revisits	31,10	19,31	2,07	6,08
Low imageability - high frequency				
First fixation	327,56	97,93	0,93	2,68
Dwell time	1528,08	857,11	2,09	4,81
Number of fixations	95,21	40,81	2,15	5,60
Number of revisits	28,49	17,86	2,03	5,09
Low imageability - low frequency				
First fixation	341,46	107,30	1,18	3,49
Dwell time	1660,51	955,42	1,83	3,41
Number of fixations	97,30	45,54	2,05	4,91
Number of revisits	29,41	17,94	1,63	3,36

Appendix 2. Percentile Values Corresponding to the ± 1.25 SD Threshold by (Meta)language Variables

<i>Variables</i>	<i>Centiles</i>
TROG	3
PPVT	13
Vocabulary depth	13
NRDLS- Inferencing	3
NRDLS gramm judgement	3
Ver semantic fluency	8
Word reading errors	5
Word reading time	5
Pseudoword reading errors	10
Pseudoword reading time	10
Spoonerisms	13
Pseudoword repetition	13

Appendix 3. Constructed Stories Used in the Eye-Tracking Task (Original Croatian Texts)

1. USPJEH NA PRVENSTVU

Condition 1: HIHF

Bila je subota. Dan kao i svaki drugi u kojem ne ideš u školu nego ležiš u kući ili izlaziš s ekipom radeći stvari koje voliš. I, naravno, želiš da to traje što dulje. Jer ipak je to dan odmora. **Teo** se spremao za zabavu. **Izlazio je sretan** iz **kuće** kada ga je u **malenom hodniku** **mama** tiho pozvala. **Dala mu je kućni telefon** i rekla: "Javi se. Zove te Milička!". Teo je primio telefon i brzo otišao u svoju sobu. Pritvorio je vrata kako bi na miru razgovarao.

Condition 2: HILF

Mama je čekala sa strane ne bi li čula što se događa. Gledala je kako se kazaljka njiše na uri s jelenjim rogom. Šaptala je nijemo u kutu da je prošlo već 10 minuta otkad je Teo počeo pričati s profom. Mama se zabrinula zbog toga. Teovi koraci začuli su se po starinskoj etaži. Gnjev, rastući, gorući bio je očit dok je mrko silazio. Rastao je s nijemim hodom u dnevnu sobu. Mama je nervozno kružila oko Tea bojeći se pitati ga što se dogodilo.

Condition 3: LIHF

Teo je bio toliko ljut zbog cijele zbrke da mu se u glavi javila od jedanput tipična misao o odustanku. Činila se prilično jednostavna. No svejedno, nema više vremena za žaljenje samog sebe jer Milička već čeka sva sretna i vesela. Napokon će Teo i još neki ići na prvenstvo iz matke i cijeli dan mozgati nad rješenjem zadataka. I zaista, bilo je skoro deset sati i Milička je već bila pred poznatim gradskim parkom s očitom voljom. Bio je već toliko pun djece koja su strpljivo čekala natjecanje koje se odvijalo u školi pored parka.

Condition 4: LILF

Dok je gledao tu računsku strahotu na papiru, Teo je prolazio tešku i beskrajnu agoniju. Naglo i nemilo izazvala je silan stres u njegovom tijelu, a pred očima su mu letjele slike društva na Sljemenu, ali i njegove profe Milić koja ga hrabri: „Ti to možeš, znam da ti to možeš!“. Odjednom se oglasi školsko zvono koje javlja kraj prvenstva. Teu se činilo da je u nekom mističnom oblaku zanosna, magičnoj obmani. Vladao je nad mozgom i tijelom malog Tea. Rezultat prvenstva je bio izvrstan. Osjećaj kada je čuo: „Prvo mjesto na prvenstvu osvaja Teo Horvat“, neće nikad nestati.

Open-ended questions: (2 literal, 2 inferential)

1. Što je raslo?
2. Gdje su se djeca skupljala?
3. Zašto je Teo bio ljut?
4. Što je Teo mogao napraviti?

2. Priča o izlasku

Bila je sunčana i topla subota u Zagrebu. Ljudi su šetali gradom i pili kavu. Grički top označio je točno podne. Tena je otišla u šetnju. Izašla je poslije dugog vremena bolesti. Naime, imala je upalu grla i deset je dana provela u krevetu. Nije smjela izaći iz kuće. A sada će se napokon

naći s Markom na odličnim kolačima u kavani u gradskoj ulici. Popela se na sam vrh ljestvice najboljih mjesta za kušanje slastica.

I Marko se spremao. Bio je presretan što izlazi s Tenom. Hodajući uz cestu vidio je mirisnu, tanku brezu sjajnu od rose. Izrasla je „nijemo“ i cvala ovdje još prije tri godine, ali Marko je to tek danas vidio. Danas je vidio i ono što inače ne vidi. Bio je sretan. Izlazi s Tenom! Uf, kad se samo prisjeti, trebalo mu je junačke snage za pozivnicu na mirisne, cimet-kuglice. Predana je uz nijemi jecaj: „Molim te, pristani“. A ona je pristala kao da je samo čekala taj poziv!

Marko se namirisao i uredio. Odjenuo je svoju najbolju košulju. Bio je spreman za ovakvu priliku, tako potrebnu. Vrijedila je doista svake muke koju je proživio. Ali, što je bio bliže, u njemu se budio osjećaj straha, a misao o mogućem porazu, počela je nastajati. Činila se stvarnijom od stvarnosti. A Marko se nikako nije mogao osloboditi tih negativnih osjećaja. Kapljice znoja počele su se stvarati na njegovom čelu, a dlanovi su postali mokri.

U njegovoj glavi vladao je košmar. Zbilja je izazvala agoniju. Stvorila je nemoć viđenja onog što je stvarno. A to je bila Tena koja je stajala ispred Marka u najljepšoj plavoj traper-haljini koju je ikada vidio. Plave oči u čudu su ga gledale. Odjednom začuje i poznati, dragi glasić: „Marko, kuda si krenuo?“ Marko je bio u mističnoj obmani. Katarza se naglo zbila. Odagnala je tmuran košmar iz Markove glave i napokon je shvatio da mu se ostvario san.

Open-ended questions: (2 literal, 2 inferential)

1. Što je izraslo?
2. Što je vrijedilo svake muke?
3. Zašto je Marku trebalo hrabrosti?
4. Zašto se Marko kasnije prestrašio?

3. Pomoć u nevolji

Nikola je učenik petog razreda koji je na glasu kao najbolji razred u školi. U Nikolin peti c razred ide i Klara, cura tankog nosića. Bila je krasna djevojka. No, nije se mogla kretati ni pisati kao i ostala djeca u školi. Imala je teškoća i s govorom, pričala je mama. Bio je spor i tih i uvijek joj je trebalo više vremena i snage da nešto kaže. Svi su u razredu voljeli Klaru, ali Nikola ju je nekako najviše pazio. I Klara je voljela Nikolu. Voljela je s njim pričati i smijati se dok bi ju on zezao, a tako je bilo i jednog utorka.

Klara je stigla u školu i sretna sjela na svoje mjesto. Nalikovala je na malu, zatočenu, nježnu damu, u vunenom plaštu. Smjestila se mrvicu trzajući rukom u klupu ispred Nikole i

pomalo se nespretno nagnula prema njemu. Jedva je dočekala da mu pokaže svoj rad. Napravila je na gipsanoj plohi ukras, kuglicu od plastelina. Izrađen žarom i snagom koja nije bila mala, ako znamo da se radi o djevojčici s problemom kretanja. Kad se nagnula, rukom je srušila s klupe pernicu i olovke su se rasule po cijelom razredu.

U Nikolinoj glavi pojavila se samo jedna misao, jedini mogući način. Postala je tako snažna da Nikola više nije dvojio ni časa. Ustao je sa svog mjesta i počeo skupljati Klarine stvari. Ali Nikolina je dobrota stvorila problem, na čudan način. Izazvao je čudne optužbe protiv Nikole i minus u dnevniku. Naime, profesor je smatrao da je Nikolino dizanje s mjesta ometanje nastave. Zato je uz niz riječi u smislu nemira, neposluha, ometanja nastave i slično, u rubriku „zalaganje“ Nikoli upisao opomenu.

Nikola je bio u šoku, a Klarine su oči bile pune suza. Nikolom je postupno ovladao bijes, ali i osobna savjest. Plamsao je nad prividom mirnoće koja se ogledala na njegovom licu. Nastavio je skupljati stvari s poda. Profesor je opet zaurao: „Sjedni na svoje mjesto!“. Nikolin odgovor bio je odraz njegovog ponosnog stava i plemenite časti. Neposredno i nehotice stvorila je prezir i podsmijeh u njegovom odgovoru kad je rekao: „Sva sreća da moram kupiti samo Klarine stvari jer da moram dizati Klaru, vjerojatno bih dobio čistu „jedinicu“.“

Open-ended questions: (2 literal, 2 inferential)

1. Što je Klara izradila?
2. Što je postalo snažno?
3. Zašto je Klara bila uzbuđena?
4. Zašto je Nikolina dobrota izazvala problem?

4. Otkrivena tajna

Ana je bila sedamnaestogodišnja vesela i otvorena djevojka duboke mašte. Voljela je knjige i filmove te se često stavljala u uloge likova o kojima je čitala ili ih gledala na filmu. Ali imala je još jednu tajnu. Voljela je lutke i često se s njima igrala i stavljala ih u različite uloge. Jednog je popodneva lijepu, dugu haljinu oblačila lutki. Bila je plave i zelene boje. A Ana se toliko zaniijela da nije vidjela kako ju Ivana u čudu promatra.

Ivana je stajala iza starinske, drevne ure sa sovom na pročelju. Zvonila je nježno njišući kazaljke. Ana nije imala pojma da ju netko gleda. Njena igra lutkama bila je njeno utočište za bijeg od stvarnosti, u čarobnu rajsku oazu. Poput čarobnog plašta izbavila je Anu iz okova teške stvarnosti iz koje je ponekad željela pobjeći. Ivana to nije razumjela i činilo joj se sasvim opravdanim ismijati Aninu strogo skrivanu tajnu.

Stoga je Ivani pala na pamet „predobra“ ideja, vješto razrađena. Postala je moguće rješenje za Ivaninu psinu- samo treba publiku koja bi je podržala. I našla ju je vrlo brzo. U grupi na Fejsu je objavila: „Znate li što sam vidjela? Naša se Ana igra lutkama, hahaha! Uživjela se u ulogu „dobre mame“ u žaru i zanosu. Imala je niz od nekoliko pjesmica koje im je pjevala dok ih je oblačila i svlačila, hahaha!“.

Ana je bila šokirana kada je to vidjela. Ivana, kojoj je tako vjerovala, izdala je njenu najskriveniju tajnu. Njena intima je bila ranjena, a krivnja zatočena u ravnodušnoj zbilji. Vladala je nad javom koju Ana nije mogla prihvatiti. Prva joj je misao bila sakriti se, ali gdje? Kako? U razmišljanju, daleko od svega odvela ju je njena mašta. Pobjegla je na čas lađom mistike i sjete. Otisnula se, takoreći, u zaborav na strahotu koju je doživjela.

Open-ended questions: (2 literal, 2 inferential)

1. Što je zvonilo?
2. Tko je pjevao pjesmice lutkama?
3. Zašto su Ani lutke oaza?
4. Što je Ivani palo?

Appendix 4. Results of Independent Samples t-Tests for Group Differences (TLD vs. LBD) on Phonological language, non-phonological language, and Phonological loop variables.

Table 64.

Group Differences Between TLD and LBD on (Meta)language and Cognitive Measures

Measure	TLD (n = 33) M (SD)	LBD (n = 37) M (SD)	t	df	p	95% CI
Language Comprehension						
TROG-2-HR	17.27 (2.53)	14.30 (3.79)	3.90	63.17	< .001	[1.45, 4.50]

Measure	TLD (n = 33)	LBD (n = 37)	<i>t</i>	<i>df</i>	<i>p</i>	95% CI
	M (SD)	M (SD)				
PPVT-III-HR	151.67 (10.05)	143.24 (11.63)	3.22	68	.002	[3.21, 13.64]
Vocabulary Depth	39.36 (5.56)	29.41 (6.79)	6.67	68	< .001	[6.98, 12.94]
Inferencing	9.94 (0.24)	9.24 (0.98)	4.17	40.86	< .001	[0.36, 1.03]
Grammatical Judgement*	13.94 (0.35)	13.57 (1.24)	1.75	42.30	.087	[-0.06, 0.80]
Verbal Semantic Fluency	22.45 (6.10)	17.43 (6.24)	3.40	68	.001	[2.07, 7.97]
Decoding						
Word Reading Errors	0.88 (1.08)	7.11 (7.02)	-5.33	37.91	< .001	[-8.60, -3.86]
Word Reading Time (ms)	122.88 (36.18)	226.57 (115.51)	-5.18	43.76	< .001	[-144.02, -63.36]
Pseudoword Reading Errors	1.70 (1.93)	8.59 (7.97)	-5.10	40.68	< .001	[-9.63, -4.17]
Pseudoword Reading Time (ms)	82.76 (29.89)	133.73 (64.92)	-4.29	51.85	< .001	[-74.80, -27.15]
Phonological Processing						
Spoonerisms	7.24 (1.09)	3.46 (2.26)	9.08	53.23	< .001	[2.95, 4.62]
Pseudoword Repetition	15.39 (0.97)	13.51 (2.38)	4.42	48.71	< .001	[1.03, 2.74]

Note. *M* = mean; *SD* = standard deviation; CI = confidence interval. For measures with unequal variances (Levene's test $p < .05$), results with corrected degrees of freedom are reported. Positive *t* values indicate higher scores in the TLD group; negative values indicate higher scores in the LBD group. *Grammatical Judgement did not reach statistical significance ($p = .087$).

14. CURRICULUM VITAE

Antonija Blaži Ostojić was born in Zagreb, Croatia, in 1993. She completed the Undergraduate and Graduate Study Programme in Speech and Language Pathology at the Faculty of Education and Rehabilitation Sciences (ERF), University of Zagreb, in 2017, graduating with honours (*summa cum laude*). She was awarded the City of Zagreb Scholarship for Excellence (top 3% of students) on two occasions between 2013 and 2017, and received the ERF Recognition for the Highest Grade Point Average in 2013/2014. In 2017, she enrolled in the Postgraduate Doctoral Study Programme in Speech, Language and Hearing Disorders at ERF. As part of her doctoral studies, she undertook a research visit to the LABDICOLE Laboratory at the University of Seville (2019/2020).

Since 2017, she has been employed as a speech and language pathologist and Head of Clinical and Research Activities at the Blaži Speech-Language Rehabilitation Centre. Her scientific and clinical interests include typical and atypical communication, speech and language development, language acquisition and processing in children, and online/offline research methods. She has presented at international conferences including IASCL, ESLA, IALP, and EUNM-CDI, and delivered an invited lecture at the Croatian Philological Society (2023). In 2022 she received the ERF Doctoral Study Award for a paper published in an international peer-reviewed journal with international co-authors. She has authored and co-authored two peer-reviewed journal articles, three articles currently in press, several conference proceedings papers, and numerous abstracts.

She is a member of EUNM-CDI, EURLYAID, the Croatian Logopedic Association, and the ESLA Educational Standards and Common Training Framework Task Force. She holds certifications in both research experimental methods and clinical assessment and intervention programmes, and actively engages in continuous professional development.