AN EXAMINATION OF READING, READING DEVELOPMENT AND DISORDER IN A HIGHLY TRANSPARENT ORTHOGRAPHY: THE CASE OF TURKISH

A thesis submitted for the degree of Doctor of Philosophy
by

Evren Hussein Raman

Brunel University London
College of Health, Medicine and Life Sciences
Department of Life Sciences
Division of Psychology

ABSTRACT

The primary focus of the current research program concerns visual word recognition and reading aloud as a function of orthographic transparency to inform current debates about the nature of visual word recognition. Within this thesis, this topic is explored using several different approaches with Turkish as the medium of choice. Additionally, the extreme orthographic transparency of Turkish serves as an excellent medium to test theories of visual word recognition. Any universal framework would need to account for the variation found in the writing systems of the Turkic family.

Using a computational linguistic method, Chapter 2 explores current definitions of orthographic transparency and novel means of quantifying orthography, extending this approach to Turkish. The result was the production of four language models that take into account the different phoneme inventories used in Turkish, as well as the two main strategies (Word-onset vs whole-word), employed to investigate the quantification of Turkish. The models produced stipulate that Turkish is more transparent than any other alphabetic orthography that has been quantified to date. The chapter also highlights the superiority of whole-word approaches in capturing a full range of variation within an orthography despite some of the current cross-linguistic limitations of using such a method.

Chapter 3 examines the currently available resources for Turkish psycholinguistic research and in response to the discovery of a lack of resources in the domain, has led to the creation of the Turkish Lexicon database. The new resource is a sizeable psycholinguistic database that includes several measures of word frequency, contextual diversity and orthographic neighbourhood density as well as providing lexical information such as word and syllable length, bigram and trigram frequency. The Turkish Lexicon was validated using a lexical decision task and also produced a subcorpus for use in future psycholinguistic studies with children.

Furthermore, there has been hardly any empirical research investigating linguistic, metalinguistic, and cognitive processes involved in reading the highly transparent orthography of Turkish. To address this, Chapters 4 and 5 investigate how these skills

might impact Turkish children who are learning to read and also aims to uncover how developmental dyslexia might manifest itself in Turkish. As such, the current research has the potential to add to our understanding of the cognitive mechanisms that underlie reading in alphabetic languages. It is envisaged that the findings of this study will add to the current debate concerning the distinct influence of universal principles and specific variations in writing systems on cognitive reading processes. In addition, the research will provide conceivably the most comprehensive account of typical and atypical reading development in Turkish-speaking children to date which has huge potential practical implications.

Chapter 4 examines the reading strategies of monolingual Turkish schoolchildren while they completed both a single-word naming and oral reading fluency task amongst a battery of cognitive tasks. The findings of the rapid development of phonology as well as the use of two distinct strategies in single-word reading lend support to the weak versions of the phonological and orthographic depth hypothesis of reading.

Chapter 5 continues to pursue this question by examining reading disorder, i.e., Developmental Dyslexia in Turkish children. According to Wydell and Butterworth's (1999) Hypothesis of Transparency and Granularity, transparent orthographies such as Turkish should not manifest with a high incidence of phonological dyslexia. The findings of Chapter 5 lend support to this position as well as being in line with multiple deficit models of dyslexia (Pennington, 2006).

In Chapter 6, the behavioural data of this doctoral thesis is supplemented with the development of a computational model of visual word recognition in Turkish, the first of its kind. The model builds on the recent incorporation of a self-teaching algorithm (Pritchard, 2012) in the Dual Route Cascaded model of reading aloud and word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Simulations include exposing the model to vocabularies of varying size to simulate different stages of vocabulary growth in reading development. In addition, Chapter 6 took preliminary steps in investigating the manifestation of developmental dyslexia in Turkish from a computational perspective.

ACKNOWLEDGEMENTS

ٱلْحَمْدُ لِلَّهِ رَبِّ ٱلْعَٰلَمِينَ

This doctoral work is ultimately the consequence of the assistance, guidance, and support of the multitude of people that I have been fortunate enough to work with for the last seven years. First, I would like to express my deepest gratitude to my advisor Prof. Taeko Wydell for the continuous support of my doctoral studies, for her patience, passion, and vast knowledge. Her guidance has aided me immensely throughout my doctoral studies. I am thankful to her for allowing me to explore my research ideas, for sharing her wisdom and for keeping me focussed. I feel so fortunate to have gotten to work with you for so long. Our marathon meetings were a source of inspiration and have contributed to increasing my understanding of the field of psycholinguistics beyond that which is reported in this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr Adrian Williams, Dr André Szameitat and Dr Lúcia Garrido, for their insightful feedback and encouragement, but also for ensuring that I remained grounded throughout this process. As a committee, you have given me a platform for my voice to be heard and have, in turn, offered me sound advice on each of the numerous occasions we have met. I would also like to thank all the academic and administrative staff who have supported me throughout my time at Brunel. Additionally, I want to extend my thanks to several fellow (former) doctoral colleagues Dr Lisa Kuhn, Dr Solomon Mangvwat, Dr Karol Wereszczynski and Dr Rahmi Saylik, for your support and comradery.

My sincere thanks also go to Prof. Biran Mertan and Prof. Shenel Husnu Raman, who provided me with an opportunity to join their Department in Northern Cyprus, and who gave me unlimited access to the laboratory and research facilities during data collection. Without their valued support, it would not be possible to conduct large parts of this research. In addition, I would also like to thank the research assistants involved with data collection, to the teachers who opened their schools and, of course, to all of the children and their families for taking part in this research.

Additionally, I would like to thank Prof. Özgül Salor-Durna and Dr Özgün Koşaner for providing me with access to the METUbet code and *Grafofon* web interface, respectively. A big thank you to Prof. Sylviane Valdois and her team for discussions surrounding Visual Attention Span and access to the *MAEVA* source code following our meeting at the 2016 British Dyslexia Association meeting. Additionally, thanks to Prof. Cem Bozşahin for providing access to the CLC database. Also, thanks to Dr Joana Acha and Prof. Aydın Durgunoğlu for providing me with the code and material for an earlier planned analysis of morphological awareness in Turkish. Lastly, thank you to Dr Stephen Pritchard and Dr Steven Saunders for access to the learning version of the DRC model and rich and insightful discussions surrounding the implementation and optimisation of these models.

I am immensely indebted to my parents, Mustafa and Ilhan, and my brothers, Deren and Salim, for your unconditional love and support. To my father, thank you for showing me what true work ethic looks like. Mum, you have been particularly influential in my development as a thinker and as a scientist and have paved my way into this research area through your research. To my dad and brothers, thank you for enduring our ongoing dinner time debates over the years.

Lastly and most importantly, thank you to my best friend and wife, Jale, for your love, patience, encouragement and sacrifices. You have seen my highs and lows and have never wavered in your support. Thank you for putting up with my reading-related ranting, for showing interest in my work and for reassuring me of the importance of my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Acronym	Definition
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BF	Bayes Factor
BOBYQA	Bound Optimization BY Quadratic Approximation
BOUN	Boğaziçi University
CA	Chronological Age
CD	Contextual Diversity
CDP	Connectionist Dual Process
CELEX	Dutch Centre for Lexical Information
CI	Confidence interval
CLC	Children's Language Corpus
CPU	Central Processing Unit
CRFP	Corpus du Référence du Français Parlé
CV	Consonant-Vowel
DD	Developmental Dyslexia
DELIC	Description Linguistique Informatisée sur. Corpus
DMDX	the Win32 DirectX based display system used in psychological, linguistic and other laboratories around the world
DRC	Dual Route Cascaded
DSM	Diagnostic and Statistical Manual of Mental Disorders
DV	Dependent Variable
GLMM	Generalized Linear Mixed Model
GPC	Grapheme - Phoneme Correspondence
GPCE	Grapheme - Phoneme Correspondences Extraction
HAL	Hyperspace Analogue to Language
HF	High Frequency
HGT	The hypothesis of Granularity and Transparency
IA	Interaction Activation
ICC	Intraclass Correlation Coefficient
ICD	International Classification of Diseases
IPA	International Phonetic Alphabet
IQ	Intelligence Quotient
ISI	interstimulus interval
ITU	İstanbul Teknik Üniversitesi
KOBİT	Kelime Okuma Bilgisi Testi
LD	Levenshtein distance
LDT	Lexical Decision Task
LHF	Long High-Frequency
LLF	Long Low-Frequency
LMM	Linear Mixed Model
LNW	Long Nonword
LOWAC	Low Accuracy
LRT	Likelihood Ratio Test
MD	Mixed Dyslexia
MEB	Milli Eğitim Bakanlığı
MEG	Magnetoencephalography
ML	Maximum Likelihood

Acronym	Definition
MROM	Multiple Readout Model
NLP	Natural Language Processing
OD	Orthographic Depth
ODH	Orthographic Depth Hypothesis
ON	Orthographic Neighbourhood
ORF	Oral Reading Fluency
PA	Phonological Awareness
PCA	Principal Components Analysis
PD	Phonological Dyslexia
PDH	Phonological Deficit Hypothesis
PDP	Parallel Distributed Processing
PGST	Psycholinguistic Grain Size Theory
PIRLS	Progress in International Reading Literacy Study
POS	Parts of Speech
PSTM	Phonological Short-term memory
RAN	Rapid Automatised Naming
RCPM	Raven's Coloured Progressive Matrices
RT	Reaction Time
SAMPA	Speech Assessment Method Phonetic Alphabet
SD	Standard Deviation
SE	Standard Error
SES	Socioeconomic Status
SHF	Short High-Frequency
SLD	Specific Learning Difficulties
SLF	Short Low-Frequency
SLI	Specific language impairment
SNW	Short Nonword
SOV	Subject Object Verb
SPSS	Statistical Package for the Social Sciences
ST	Self-teaching
STM	Short-term Memory
SUBTLEX	Subtitle Lexicon
SWNN	Single Word/ Nonword Naming
SWR	Single Word Reading
TD	Typically Developing
TDK	Türk Dil Kurumu
TLA	Two-Layer Associative
TNC	Turkish National Corpus
UPP	Universal Phonological Principle
VA	Visual Attention
VPT	Visual Patterns Test
VSSTM	Visual-Spatial Short-term Memory
VSTM	Visual Short-term Memory
WCPM	Words Correct per Minute
WF	Word Frequency
WISC	Wechsler Intelligence Scale for Children
WM	Working Memory
WPM	Words per Minute

CHAPTER 1: GENERAL INTRODUCTION

The advent of reading and writing approximately 5000 years ago (Harris, Graham, Brindle & Sandmel, 2009) represents a unique turning point in human history. For the first time, ideas and thoughts could be conveyed beyond verbal communication and thus, the transmission of information transcended time and space. This, coupled with the rich and diverse plethora of languages and orthographies, makes reading science a fascinating domain of research. As an example of diversity, alphabetic systems are said to correspond to distinct phonemes. In contrast, syllabic systems correspond with spoken syllables, and logographic systems use symbols to represent meaning directly with few clues to pronunciation (Ellis et al., 2004). As opposed to oral language development, learning to read is considered to be the result of sustained, systematic instruction and not the direct consequence of exposure during infancy (Wydell, 2012). Consequently, the capacity to read is considered fundamental to today's contemporary society and obtaining literacy is thought to increase prospective opportunities in education and employment; indirectly contributing to the quality of life (Snowling & Hulme, 2012). This being said, it seems almost contradictory that our brains are so efficiently adapted to the specific challenges presented by the relatively recent cultural invention of writing systems. It is precisely this paradox that makes reading science such a fascinating area of research.

Reading is a complex neurodevelopmental process that entails decoding printed symbols and abstracting meaning from written language. Over the past century, reading research has enjoyed much attention in areas such as reading development (e.g. Lyytinen et al., 2006), skilled reading (e.g. Besner & Coltheart, 1979) and reading difficulties (e.g. Landerl et al., 2013). The reading process reflects the coordination of distinct perceptual and cognitive abilities beginning with low-level visual perception (Carreiras, Armstrong, Perea, & Frost, 2014) leading to eye-movement control, phonological processing, and recognition of word forms (Norris, 2013). In addition, higher-level psycholinguistic processes are involved in the abstraction of meaning from print via the feedforward and back processing of relationships between orthography, morphological structure, and semantics (Whiting, Shtyrov, & Marslen-Wilson, 2014).

Reading represents a multi-staged procedure, and it is not within the scope of this thesis to study all aspects of reading. It is, therefore, critical to explicitly state that the aspect of reading under investigation in the present study is the visual recognition of isolated words as well as the growth of text reading fluency. Furthermore, visual word recognition and reading fluency will be explored using a Cognitive Science approach within the domains of orthographic transparency, reading development and specific reading disorder, i.e., Developmental Dyslexia. The following sections will highlight the topics of interest within the current thesis. It is important to note that there is some overlap in the literature review sections between chapters as a means to reintroduce some of the core topics discussed in this thesis.

1.1 Reading Acquisition and Development in Alphabetic Writing Systems

It is widely accepted that an emergent reader will approach the process of reading differently from a skilled reader. Reading acquisition is thought to arise from a combination of cognitive, language, and social skills that have developed from birth. It is postulated that the most significant of these skills is the emergent reader's proficiency in oral language, which is widely accepted as the essential foundation for reading (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Several theoretical stage models of reading acquisition have been proposed that outline the stages of cognitive development in which learners are suggested to transition through in order to acquire the skill of reading (e.g., Ehri, 1991; Frith, 1985; Gough & Hillinger, 1980; Marsh, Friedman, Welch, & Desberg, 1981). The majority of the models mentioned above are highly harmonious in that they share the central supposition that achieving skilled reading in alphabetic languages necessitates the use of the alphabetic principle (Rozin & Gleitman, 1977). This print-to-sound mapping process is also referred to as phonological recoding (Share, 1995). These four core stage theories will each be considered in turn.

Gough & Hilinger's Two-Stage Model (1980)

In an early iteration of stage models, Gough and Hillinger (1980) propositioned a two-stage model for learning to read, which included an early-visual association stage and a decoding-based learning stage. In the first stage, it was postulated that there is an exclusively visual process that lacks information related to decoding. Children are said to use any functional source of information, i.e. contextual and pragmatic, in order to distinguish one word from another. In the second stage, children transition to the use of letter knowledge and grapheme-to-phoneme correspondences to decode novel words or pseudowords.

Marsh, Friedman, Welch & Desberg's Four-Stage Model (1981)

Similarly, Marsh, Friedman, Welch and Desberg (1981) suggested a framework that consisted of four stages of reading development. Comparable to Gough and Hillinger (1980), they proposed that initially, children make use of visual cues in order to identify words – the linguistic guessing stage. In the second stage, the sophisticated guessing stage, the authors propose that children learn to predict words from contextual cues. Following this, children learn to orthographically decode sequentially, the sequential decoding stage. In the final stage, children are said to read by analogy to familiar words that are stored in their mental lexicons.

Frith's Three-Stage Model (1985)

Similar to Gough and Hillinger's and Marsh and collegues's first stages, Frith (1985) initiated her three-stage model with a logographic stage. During this initial stage, Frith proposed that a child can recognize individual words based on certain salient features of the word but cannot read novel words. In the second stage of the model, the alphabetic stage, children acquire the ability to read by utilizing grapheme-phoneme correspondences. Finally, in the orthographic stage, Frith (1985) postulated that children could use earlier developed alphabetic skills for the decoding of unfamiliar words as at this stage they have developed a fully mature reading system parallel to that of a skilled adult reader and thus has both whole-word and grapheme-phoneme decoding strategies available to them.

Ehri's Four-Phase Model (1992; 1997)

Ehri (1992; 1997) proposed a four-phase model, beginning again with the reliance on visual cues and patterns – the pre-alphabetic phase. The partial alphabetic phase stipulates that children exploit their partial knowledge of letter names and sounds. Subsequently, the full alphabetic phase completes connections between letters and sounds, enabling children to decipher new words. With time and practise, children begin to develop their sight-word reading and terminate the learning process with the consolidated alphabetic phase in which reading behaviour starts to emulate that of skilled readers. Additionally, Ehri (1991) proposes that learning the alphabet and not the alphabetic principle is the crucial factor in the transition to the partial alphabetic phase. Ehri concludes by stating that learning to read comprises determining complete word representations that incorporate both phonological and orthographic components. One of the significant limitations of stage/phase models is that although the models are highly informative, they ultimately serve as a theoretical outline of reading development rather than as a set of falsifiable scientific hypotheses (Beech, 2005).

Additionally, there is a corpus of research that indicates that not all children transition through the stages/phases highlighted above (Stuart & Coltheart, 1988; Wimmer & Hummer, 1990). Furthermore, the two above studies raised important questions regarding the ability of children to utilize phonological processing skills in the early stages of reading development (Stuart & Coltheart, 1988) as well as questioning the applicability of the developmental stage models across languages (Share, 2008; Wimmer & Hummer, 1990). Ultimately, developmental stage/phase models are culpable of being too descriptive in their nature as crucial aspects of each of the discrete model's workings are left undetermined. The following section will focus on providing an overview of the cognitive abilities that contribute to learning to read in alphabetic writing systems.

1.2 Cognitive and Linguistic Factors involved in literacy Acquisition

There has been an increasing shift towards exploring the cognitive and linguistic factors that are involved in the development of reading ability. Before an emergent reader can advance to the final stages of reading acquisition, there is a need to develop metalinguistic skills that will facilitate learning to read. The literature readily identifies three metalinguistic skills that are considered to contribute to the acquisition of reading and spelling which are phonological awareness, orthographic awareness, and morphological awareness (Apel, Wilson-Fowler, Brimo, & Perrin, 2012) Each will now be considered in turn.

Phonological awareness may be conceptualized as the awareness of the internal sound structure of spoken words. A developing reader needs to develop an awareness of phonological units such as words, syllables, onset-rimes, and phonemes in order to develop the ability to decode words successfully. There is a sizeable body of literature investigating how developing reader's level of phonological awareness relates to reading development (Bradley & Bryant, 1983; Muter, Hulme, Snowling, & Stevenson, 2004; Wagner & Torgesen, 1987). In general, phonological awareness is frequently identified as the strongest predictor of early reading skill (Adams, 1990; Goswami & Bryant, 1990). Additionally, a recent meta-analysis by Melby-Lervåg, Lyster and Hulme (2012) explored the relationship of three measures of phonological awareness, namely, phonemic awareness, rime awareness, and verbal short-term memory, with children's reading skills. The study determined that phonemic awareness was the strongest predictor of individual differences in reading development after controlling for the variance in verbal short-term memory and rime awareness. There has also been much debate regarding the causal link between children's phonological awareness and success in learning to read. In an argument against, Castles and Coltheart (2004) contend that there is a lack of irrefutable evidence regarding a causal link between children's phonological awareness and success in learning to read while acknowledging the relationship between phonological awareness and reading ability. Instead, Castles and Coltheart (2004) further argue that phonemic awareness may not be acquired in the absence of instruction on the relationship between phonemes and graphemes; more suggestive of the concept of graphophonemic awareness (Connelly, 2002; Ehri & Soffer, 1999). Alternatively, in a direct rebuttal of the above critique, Hulme, Snowling, Caravolas and Carroll (2005) dispute Castle and Coltheart's conclusions in that they focus on a relatively narrow theoretical perspective and state that it would be more useful to study the role of phonological skills in learning to read in a broader framework of a multi-causal model.

Orthographic awareness refers to awareness of spelling patterns within words. Orthographic awareness is thought to require the morphological rules of a given language in addition to knowledge of letter patterns and sequences (Berninger & Abbott, 1994). It has been conceptualized that orthographic processing skills (Barker, Torgesen, & Wagner, 1992) may impact children's implicit acquisition of the statistical regularities of orthographic and phonological representations of words. In addition to this, a recent study identified that orthographic processing skill estimated a substantial proportion of the variance in word recognition after phonological processing was taken into consideration (Berninger & Wolf, 2009). Moreover, spelling, orthographic awareness and reading accuracy are thought to interact and supplement each other (Berninger, Abbott, Nagy, & Carlisle, 2010). Berninger and colleagues carried out a growth curve analysis of phonological, orthographic, and morphological awareness in children in grades from one to six. They reported that at the word-level, phonological and orthographic awareness demonstrated the highest growth during the earlier grades with limited growth after this period. They conclude by recommending that all three types of metalinguistic awareness need to be coordinated and applied to literacy learning. Furthermore, there is a growing interest in morphological awareness as a central facet of reading, particularly with regards to vocabulary knowledge.

Morphological awareness is regarded as the ability to understand the structure of a word as a combination of morphemes and to have the ability to manipulate them (Weber et al., 2013). The distinct role of morphology in reading acquisition is well established in scripts that use the Latin alphabet (Elbro, 1996; Leong, 1999). After accounting for reading ability, verbal and nonverbal intelligence, and phonological awareness, Deacon and Kirby (2004) reported the significant, independent contribution of morphological awareness to English pseudoword reading and reading comprehension but failed to contribute to single word reading. In order to address the notion of whether orthographic and morphological skills are separable in their contribution to reading outcomes, Roman, Kirby, Parrila, Wade-Woolley, and Deacon, (2009) conducted a series of regression analyses and established that orthographic knowledge and morphological awareness do indeed make unique contributions to reading development.

1.3 Cognitive Skills involved in Literacy Acquisition

Apart from metalinguistic skills, several cognitive skills have also been identified as contributors to reading skill development. Visual Attention Span (VA Span), Rapid Automized Naming (RAN) and various aspects of short-term and working memory will all be discussed within the remit of this thesis. As previously discussed, the development of reading in alphabetic languages necessitates learning the relationships between sequences of visual symbols and their related units of sounds. Therefore, reading may also be conceptualized as a visual perceptual task that involves processing multi-letter sequences (Bosse & Valdois, 2009). In stark contrast to the abundant research carried out regarding the role of phonological awareness in learning to read, there is still much debate in the domain regarding the impact of visual processes on learning to read. Consequently, the role of visual attention in reading has mostly been ignored by theorists and modellers, except for a few researchers (e.g., Bosse, Tainturier, & Valdois, 2007). The process of reading is undeniably dependent on the visual processing of letter strings (Bundesen, 1998) and a comprehensive investigation of reading would need to consider this.

Visual Attention Span (VA Span)

The concept of the Visual Attention Span (VA Span; Bosse, Tainturier, & Valdois, 2007) has to date primarily been applied to children with reading difficulties (see below) and is rooted in the theoretical, computational and neuropsychological framework of visual attention (Bundesen, 1990; 1998; Bundesen, Habekost, & Kyllingsbæk, 2005). VA Span is conceptualized and built on the connectionist multitrace memory (MTM) model of polysyllabic word reading (Ans, Carbonnel, & Valdois, 1998). Concerning reading, Bosse, Tainturier, and Valdois, (2007) functionalized VA Span as the number of orthographic units in words that can be simultaneously processed at a glance. Bosse and Valdois (2009) measured the VA Span of 417 children across the first, third and fifth grades and reported that independent of phonological processing skills, VA Span contributed to reading performance across grades. The authors conclude the study by suggesting that in order for the orthographic sequence of an input word to be memorized and consolidated into sight-word memory, VA Span needs to be large enough to process all the letters of a word simultaneously. In addition, there is increasing evidence that VA Span may be modulated by orthographic transparency in both monolinguals (Awadh et al., 2016) and bilinguals (Lallier, Acha, & Carreiras, 2016).

Rapid Automatized Naming (RAN)

As stated previously, word naming is thought to be dependent on a range of linguistic and cognitive subsystems. The synergy of these subsystems is considered to lead to rapid automation of the reading process in developing readers. Speed of processing is often quantified by Rapid Automatized Naming (RAN) tasks, which measure how rapidly people can name aloud objects, pictures, colours, or symbols (letters or digits). RAN tasks are thought to be reflective of phonological access to lexical storage and were initially developed to differentiate developmental dyslexics from controls (Denckla & Rudel, 1976a; Denckla & Rudel, 1976b; Denckla, Rudel, & Broman, 1981). RAN scores consistently correlate with reading ability in children and adults and thus maintains abundant significance in the reading literature, probably second only to phonological awareness. Furthermore, a meta-analysis of 35 studies revealed that performance on the RAN task is consistently correlated with, and predictive of both sight word and nonword reading (Swanson, Trainin, Necoechea, & Hammill, 2003). The authors report that the average correlation between RAN and real-word reading/nonword reading was $r = .42$ and $r = .52$ respectively. A potential concern regarding RAN is that from a cognitive perspective, it is a poorly defined construct and that it is unclear precisely what mechanism underlies RAN's relationship with reading (Kirby, Georgiou, Martinussen, & Parrila, 2010). It has also been posited that RAN could be a broad measure that assesses several cognitive skills (Arnell, Joanisse, Klein, Busseri, & Tannock, 2009). Although the importance of RAN is acknowledged, further work is necessary to identify the theoretical nature of naming speed.

Working Memory (WM), Phonological Short-term Memory (PSTM) & Visual-Spatial Short-Term memory (VSSTM)

Ultimately when learning to read, there is a requirement for an emergent reader to establish relationships between orthographic and phonological patterns in memory (Stanovich, 1991). For this thesis, the following subsystems of memory will be further discussed: Working Memory (WM), Phonological Short-Term Memory (PSTM) and Visual-Spatial Short-Term Memory (VSSTM). Concerning the updated multi-compartment model proposed by Baddeley (2000), working memory is defined as a limited capacity store of human memory that stores and manipulates information as opposed to previously suggested Short-Term Memory (STM) that only refers to the storage of information for short durations. It is also essential to establish that working memory differs from long-term memory, which is thought to have unlimited capacity and holds information in a stable form. The model of working memory assumes a central executive attentional controller that maintains two slave subsystems known as the visuospatial sketchpad and the phonological loop.

The visuospatial sketchpad functions to create and maintain a temporary visuospatial representation of the visual world. Research carried out by Logie (1986; 1995) and Klauer and Zhao (2004) suggest that spatial tasks interfere with spatial skill, whereas a purely visual activity may interfere with the capacity to remember objects or shapes. The visuospatial sketchpad can, therefore, be further divided into separate visual, spatial, and possibly kinaesthetic components. VSSTM is hypothesized to be a temporary storage component used to process received visual information for ongoing cognitive tasks (Baddeley, Eysenck, & Anderson, 2009). Huestegge and colleagues (2012) suggest that with regards to reading, VSSTM acts as an intermediary between the fast decaying perceptual impressions of notation and the, more long-term, crystallized knowledge of language-specific orthographic patterns (Dehaene et al., 2010). The precise nature of the relationship between VSSTM and reading is poorly understood, and the conflicting evidence leaves much to be examined (Bosse, Tainturier, & Valdois, 2007; Dehaene et al., 2010; Gathercole, Alloway, Willis, & Adams, 2006).

The phonological loop processes acoustic information and is assumed to have two components comprised of a temporary acoustic store and a sub-vocal articulatory rehearsal process. The phonological loop is conceivably the most comprehensively investigated aspect of working memory (Baddeley, 1992; 2000; 2003). The presence of the phonological similarity effect (Baddeley, 1966a; 1966b), as well as the word length effect (Baddeley, Thomson & Buchanan, 1975), is the primary basis of evidence in support of a phonological store and a feedback loop mediated by rehearsal, respectively. According to Baddeley, Gathercole, & Papagno (1998), the phonological loop facilitates language acquisition by affording a temporary store for novel words while they are consolidated in long-term phonological memory. Furthermore, the phonological loop is thought to be mostly responsible for PSTM (Hummel, 2009). Therefore, PSTM can be understood to be central to the processing of auditory information (Anthony, Williams, McDonald, & Francis, 2007). Concerning reading, PSTM has been demonstrated to have a significant influence on reading accuracy but not speed across several different alphabetic orthographies (Ziegler et al., 2010). This section has briefly discussed some of the critical cognitive contributors to reading development in typically developing children, but what happens when the complex process of reading development is disrupted? The next section will describe the manifestation of cognitive and linguistic difficulties in reading within the framework of developmental dyslexia with reference to alphabetic writing systems.

1.3 Developmental Dyslexia

While the majority of children learning to read do so with relative ease, there is a significant number (~10%) that struggle to obtain the relevant skills needed to extract meaning from a written language with ease and fluency (e.g., Boets, Wouters, Van Wieringen, & Ghesquiere, 2007). Specific reading disorder, also known as Developmental Dyslexia (DD), may be conceptualized as a marked difficulty in reading words not due to sensory, neurological, or intellectual impairments and inadequate schooling. Furthermore, developmental dyslexia frequently displays co-morbidity with other neurodevelopmental disorders such as attention deficit hyperactivity disorder (ADHD) (25-40%) (Landerl & Moll, 2010), dyscalculia (17%) (Ansari & Karmiloff-Smith, 2002) and developmental coordination disorder (DCD)/ dyspraxia (50%) (Kaplan, Wilson, Dewey, & Crawford, 1998). Developmental dyslexia is widely accepted to be a neurobiological disorder with a genetic origin that primarily affects the acquisition of literacy skills, in particular, learning to read with speed and accuracy. Within the literature, there is still an ongoing debate over both the definition and subsequent diagnosis of developmental dyslexia. The foundation of this disagreement is posited to stem from the diverse neurological and cognitive accounts of the disorder (Ramus, 2003). Additionally, many of these competing theories refer to distinct theoretical approaches developed in order to explain developmental dyslexia from a causal perspective (Reid, 2001). The most dominant of these competing theories will now be considered in turn.

Phonological Processing Deficit Hypothesis of Dyslexia

The predominant phonological processing deficit hypothesis of dyslexia (Muter, Hulme, Snowling, & Taylor, 1998; Rack, 1994; Share, 1995; Stanovich, Cunningham, & Cramer, 1984; Vellutino, Fletcher, Snowling, & Scanlon, 2004; Wydell & Butterworth, 1999) states that the core skill of reading ability is phonological processing and as such, is reported to be the core deficit in developmental dyslexia (Snowling, 2001). The most frequently reported problems of dyslexic children using this framework are phonological awareness deficits that hinder with the acquisition of grapheme-phoneme correspondences as well as limitations of verbal short-term memory (Snowling, 1998). Considering the role of phonological awareness in typically developing children, there is strong evidence for this hypothesis. With reference to the Dual Route Cascaded (DRC) model of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) phonological deficits are thought to affect processing via both the lexical and the nonlexical route.

Double-Deficit Hypothesis of Dyslexia

An extension of the phonological account of reading disorder is the double-deficit hypothesis (Wolf & Bowers, 1999; 2000), which postulates that in addition to the observed phonological difficulties experienced by dyslexic children, there is a second equally important deficit in naming speed as measured by RAN tasks (Allor, 2002; Denckla & Rudel, 1976b). The double-deficit hypothesis further stipulates that regarding dyslexic children, the correlations between phonological awareness and naming speed are different though both phonological awareness deficits and naming-speed deficits are reported (Wimmer, Mayringer, & Landerl, 2000).

Domain-General Perceptual Deficit of Auditory Processing Hypothesis

In an opposing view, there is an argument that observed phonological deficits might simply be an indicator of developmental dyslexia as opposed to the central causal factor. Instead, proponents of the rapid auditory processing deficit hypothesis (P. Tallal, 1984; P. E. Tallal, Galaburda, Llinás, & von Euler, 1993) suggest a domain-general perceptual deficit of auditory processing. There is indeed a growing corpus of research that indicates that developmental dyslexics manifest a low-level deficit in rapid auditory processing resulting in a phonological deficit (Stoodley, Hill, Stein, & Bishop, 2006; Temple et al., 2000)

Visual Processing Deficit Hypothesis

Several alternative hypotheses regard developmental dyslexia as a primarily visual deficit: on the one hand, the magnocellular theory of developmental dyslexia (Stein, 1989; 2001). The differential properties of magno- and parvo-cells allow researchers to investigate the two distinct pathways with a variety of psychophysics studies. With regards to developmental dyslexia, there is an argument which contends that the dysfunctional frequency and amplitude sensitive magno-cells present in both the visual and auditory modalities negatively impact motion sensitivity (binocular instability) and rapid auditory (phonological) processing respectively (Ray, Fowler, & Stein, 2005; Sperling, Lu, Manis, & Seidenberg, 2003). In a similar but distinct line of enquiry, the visual stress theory (Wilkins, 2003) stipulates that visual stress leads to distortions of text and headaches when reading through visual stress. The theory advocates for the use of coloured lenses in order to reduce visual stress but is not considered to be a specific theoretical theory of dyslexia though some dyslexic children do show marked patterns of visual stress (Ramus et al., 2003). The visual processing deficit hypothesis attempts perhaps inadvertently to fuse several theories into a coherent framework, but the theory remains contentious at least in regard to alphabetic languages (Wang, Bi, Gao, & Wydell, 2010). On the other hand, the visual attention span deficit hypothesis (Bosse, Tainturier, & Valdois, 2007) has received increased attention as it has been established that phonological awareness and visual attention span make a unique impact to the reading performance of children with developmental dyslexia. They also report that phoneme awareness was responsible for a large amount of variance in pseudoword reading.

By extension, they were adding weight for the argument of the robust impact of phonological processing on reading skills (Ziegler et al., 2008). VA Span impairments are often manifested as a deficit in the ability to recall strings of consonants though they can identify consonants in isolation and often have preserved phonological processing abilities (Bosse & Valdois, 2009b). A recent intervention case study on a French-Spanish bilingual dyslexic girl by Valdois and colleagues (2014) found that after specific training in a VA Span task, the dyslexic child reported higher scores on the VA Span tasks. In addition, there was increased activation in her superior parietal lobes bilaterally; an area thought to be associated with the neural underpinnings of VA Span (Peyrin, Démonet, N'Guyen-Morel, Le Bas, & Valdois, 2011). Based on this finding, the authors concluded in favour of a causal relationship between VA Span and developmental dyslexia, though stated that more extensive studies need to be conducted before a true conclusion could be drawn. It is also important to highlight that processes involved in WM and STM have also received much attention in children with reading disabilities over the last 30 years (Swanson, Cooney, & McNamara, 2004). It is now becoming evident that developmental dyslexia may best be defined as a multi-faceted disorder and that a combination of factors contributes to its heterogeneous manifestation (Pennington, 2006). It can, therefore, be assumed that phonological factors (Siegel, 1990; Snowling, 1995; Stanovich, 1996), working memory (Baddeley, 1993; Rack, 1994), visual processing (Stein, 1989; Wilkins, 2003) and processing speed of information (Rack, 1994) may all play vital roles in explaining developmental dyslexia (Pneuman, 2009). Now that reading processes and dyslexia have been reviewed and defined; it is essential to reframe these findings within different orthographies and to identify the relevant universal and language-specific processes involved in reading.

1.4 The Role of Orthography and Aspects of Language Universality vs Specificity

Orthography is considered to be the realisation of a writing system to a specific language. How a particular orthography maps on to the phonology of its spoken language is thought to affect literacy acquisition in children as well as adults' cognitive processing. This growing area of research has enjoyed particular focus over the past 30 years exploring universal and language-specific processing by readers of different orthographies, as well as the relationship between orthographic transparency and the incidence of developmental disorders. Much of this area of research evolved from the Orthographic Depth Hypothesis (ODH) (Frost, Katz, & Bentin, 1987; Katz & Feldman, 1983; Katz & Frost, 1992). This seminal unidimensional hypothesis postulates that print-to-sound mappings, i.e. Grapheme-Phoneme Correspondence (GPC) in alphabetic writing systems dictate the respective orthographic transparency of a written language, which in turn influences the strategies adopted by readers.

When a particular orthography has consistent GPC, it is said to be shallow or transparent, and as such, the reader adopts a print-to-sound sequential decoding strategy. Conversely, deep or opaque orthographies are thought to utilize a more direct 'whole word' look-up strategy (referred to in the DRC model as the lexical route). In summary, proponents of the strong version of the ODH suggest that phonologically transparent words should always be insensitive to both word frequency and priming effects as these effects are said to be involved in lexical processing. Criticism of the ODH has come from several transparent languages which display both word frequency, and priming effects including Croatian (Carello, Lukatela, & Turvey, 1988), Persian (Baluch & Besner, 1991) and Spanish (Sebastián-Gallés, 1991) and word frequency effects have been reported in Japanese syllabic Kana (Wydell, 1991; Rastle, Havelka, Wydell, Coltheart & Besner, 2009) and Turkish (Raman, Baluch, & Sneddon, 1996).

In response, some researchers have proposed two distinct universal hypotheses in which it is advocated that lexical access is achieved in one universal mechanism across all orthographies independent of orthographic depth. The Universal Phonology Mediation Hypothesis (UPMH) assumes the prelexical role of phonology (Frost, 1998;

Van Orden, Pennington, & Stone, 1990) whereas the Universal Direct Access View Hypothesis (UDAVH) stipulates that lexical access occurs through the visual and direct route without the mediation of phonology between orthographic forms. Orthographies have also been approached by two-dimension theories. Wydell and Butterworth (1999) instituted the Hypothesis of Granularity and Transparency (HGT) and identified the two dimensions as transparency and granularity. The hypothesis suggests that transparent orthographies will not produce a high incidence of phonological dyslexia, regardless of the level of translation, i.e., phoneme, syllable, and character. In addition to this, they stipulate that any orthography whose smallest grain size representing sound is coarse, i.e., a whole character or whole word, should also not produce a high incidence of phonological dyslexia. Of note is that Ziegler and Goswami (2005) also indicate the consequence of grain size in order to explain developmental dyslexia across different languages and as a consequence have posited the Psycholinguistic Grain Size Theory (PGST). The major difference between these two hypotheses is that the PGST suggest that orthographic transparency is not predictive of a reduced incidence of developmental dyslexia. The influence of writing systems on reading processes has highlighted both universal principles and specific variations (Perfetti, 2003). Language-specific variations in reading have briefly been discussed above, and so attention will now focus on universal aspects of reading.

The most fundamental universal is said to be the Language Constraint on Writing Systems (Perfetti & Liu, 2005) that postulates that a reader understands the meaning of printed words within the setting of a given language and not as independent symbols that confer meaning. The second universal, the Universal Phonological Principle (UPP) (Perfetti, Zhang, & Berent, 1992; Perfetti & Harris, 2013) posits that word reading activates phonology at the lowest level of language allowed by the writing system. However, Venezky (2006) states that a truly comprehensive understanding of both the universal and language-specific features of reading remains to be established. Critically, the vast majority of research that has been carried out in reading research to date has been of Anglocentric (Share, 2008) or Eurocentric focus. Consequently, the applicability of theories of reading, reading development and disorder to non-English/ European orthographies has to be questioned. That is if a further understanding of the universal principles of reading is to be achieved, then reading research needs to take into account a more diverse scope of findings of the world's orthographies. This outstanding issue is the primary motivation of this doctoral thesis. As a final consideration, the following section will highlight recent developments concerning computational models of visual word recognition while also highlighting contemporary theoretical considerations in computational modelling.

1.5 COMPUTATIONAL MODELS OF VISUAL WORD RECOGNITION

Computational models of visual word recognition have become essential tools for investigating the cognitive phenomena associated with both normal and disordered reading. Computational modelling, as an endeavour, provides a mechanistic account of a given cognitive phenomena by offering explicit and testable predictions of how these processes function (Forstmann et al., 2011)

The following section will introduce three models of reading aloud, namely, the previously mentioned, dual-route cascaded model of visual word recognition and reading aloud (DRC) (Coltheart, Rastle, Perry, Langdon & Ziegler 2001), the connectionist dual-process family of models (CDP, CDP+, CDP++) (Perry, Ziegler & Zorzi 2007) as well as the parallel distributed processing (PDP) model family (Seidenberg & McClelland 1989; Harm & Seidenberg 1999, 2004; Plaut, McClelland, Seidenberg & Patterson 1996).

1.5.1 THE DUAL-ROUTE THEORY OF READING ALOUD

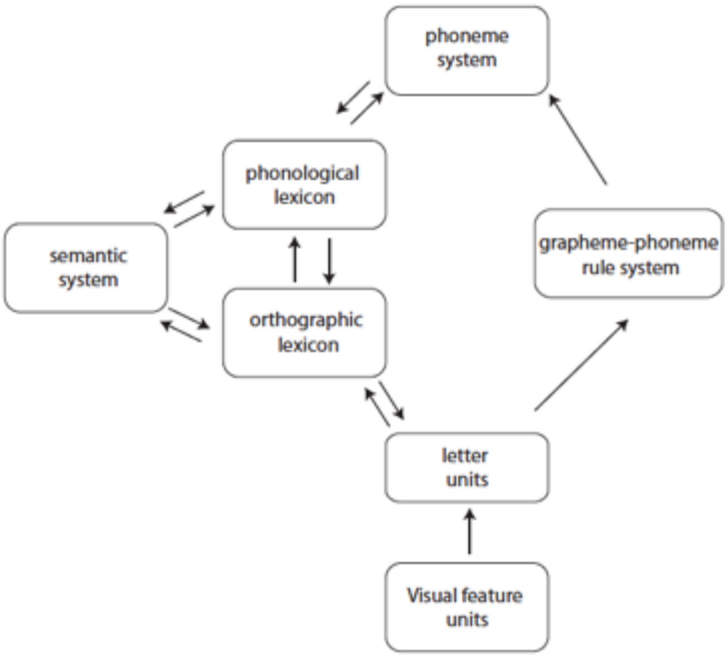

FIGURE 1: ARCHITECTURE OF THE DUAL-ROUTE CASCADED MODEL (ADAPTED FROM COLTHEART ET AL., 2001)

The dual-route cascaded (DRC) model of reading aloud and visual word recognition was implemented to simulate the cognitive mechanisms associated with skilled reading based on the dual-route theory of reading aloud (Forster & Chambers, 1973; Marshall & Newcombe, 1973). As mentioned previously, the dual-route theory postulates that two distinct routes are involved in reading aloud: a lexical and a sublexical route. The sublexical route involves serially constructing a phonological representation of a word through knowledge of how constituent parts of the word correspond to meaningful sounds. In contrast, the lexical route involves the automatic recognition of whole written words, without needing to parse the constituent parts of the word or recognize phonology beforehand. Computationally, the lexical route of the DRC is an extension of the Interactive Activation (IA) model (McClelland & Rumelhart, 1981). The lexical route can successfully identify known and irregular words, whereas the sublexical route is used to identify novel regular words and nonwords (Coltheart, 2006; Coltheart et al., 2001). In addition, the DRC model has been successful in simulating a wide variety of word reading effects observed in adult readers in English (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) as well as being implemented across a range of alphabetic orthographies including French (Ziegler, Perry, & Coltheart, 2003), German (Ziegler, Perry, & Coltheart, 2000), Greek (Kapnoula, Protopapas, Saunders, & Coltheart, 2017), Italian (Schmalz, Marinus, Coltheart, & Castles, 2015) and Russian (Ulicheva, Coltheart, Saunders, & Perry, 2016).

1.5.2 The CDP family of models

The CDP family of models (CDP: Zorzi, Houghton, & Butterworth, 1998; CDP+: Perry, Ziegler, & Zorzi, 2007; CDP++: Perry, Ziegler, & Zorzi, 2010; CDP++.parser: Perry, Ziegler, and Zorzi, 2013) represent the middle ground of computational approaches to visual word recognition. The CDP family of models maintain a dual-route architecture similar to the DRC model but retain the computational style of parallel distributed processing (PDP) models (Zorzi, 2010). Considering the architecture of the model, the lexical route of the CDP models is implemented as an IA network (similar to the DRC model). The distinction of the lexical route of the CDP models is that it contains a more extensive vocabulary than the DRC model and more recent implementations include multi-syllabic words (CDP++: Perry, Ziegler, & Zorzi, 2010). Regarding the sublexical route of the CDP models, while the DRC model uses a list of distinct rules to convert graphemes to phonemes, the CDP models use a connectionist, two-layer associative (TLA) network through the identification of statistical relationships between orthographic and phonological word parts. The TLA network, therefore, affords the CDP models properties akin to supervised learning and therefore provide a good account of developmental data on reading acquisition (Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). In addition, learning in the TLA network is based on the delta rule learning procedure (Widrow & Hoff, 1960) which is suggested to be equivalent to a classical conditioning law (the Rescorla-Wagner rule; Sutton & Barto, 1981).

1.5.3 THE PDP FAMILY OF MODELS

Taking a neurally inspired approach to understanding various aspects of cognition, the publication of the two-volume Parallel Distributed Processing (Rumelhart, McClelland, and the PDP Research Group, 1986) marked a shift away from symbolic models to what has been described as neural-network or connectionist models (Rumelhart, Smolensky, McClelland, & Hinton, 1986). The so-called triangle models of reading aloud (Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) have, along with the DRC model, motivated much of the research regarding computational modelling of reading aloud. The PDP family of models represent a distinct difference in modelling philosophy to that of previously mentioned models. In this distributed view, there is no mental lexicon for the store of word knowledge in the recognition system (Plaut, 1997; Seidenberg & McClelland, 1989, 1990) and the central difference between word and nonword stimuli is the nature of their underlying visual (orthographic), phonological and semantic statistical representations. Triangle models consist of three sets of processing units, the first is a set of units representing orthography, the second is a set of units representing phonology, and thirdly there are a set of semantic units (Powell, Plaut, & Funnell, 2006).

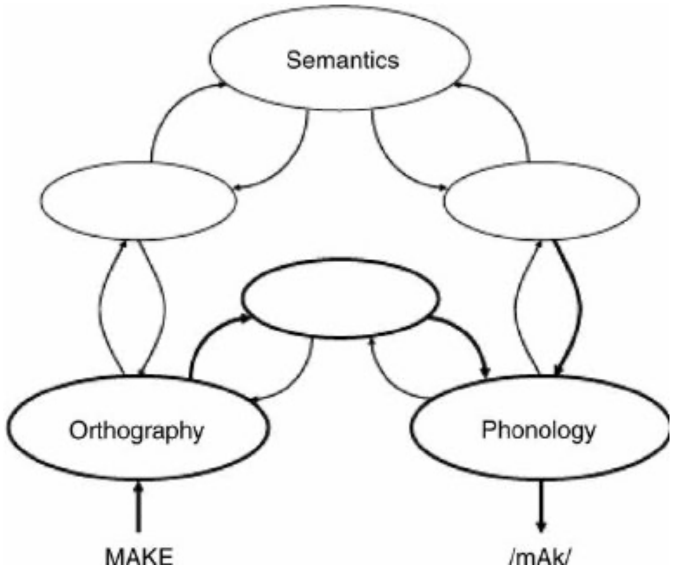

FIGURE 2: THE PDP MODEL OF READING (WOOLLAMS ET AL., 2007).

According to the triangle models, a word's pronunciation is generated by spreading activation from orthographic input units along with connections to additional phonological output units. As such the information that permits a reader to identify printed words is contained in a single set of input-to-output connections, with the pronunciation of each word being influenced by the cumulative total of all of the underlying levels of representation for that word. While an early implemented version of the triangle model (Seidenberg & McClelland, 1989) performed significantly worse than skilled readers at pronouncing nonwords (Besner, Twilley, McCann, & Seergobin, 1990), Plaut and colleagues (1996) developed a successor model that learned to produce pronunciations for both words and nonwords that simulated skilled adult reading and has subsequently been used to simulate a range of phenomena associated with reading aloud successfully. Successively, Harm and Seidenberg (1999) examined the impact of prior exposure to phonological forms of words on succeeding reading training. Their triangle model was able to accurately learn the dependencies between phonemes appearing in words and, after undergoing a full training routine, was able to replicate the findings of Plaut and colleagues (1996) concerning reading at skilled adult levels. However, it was not until the computational implementation of the semantic pathway by Harm and Seidenberg (2004) that the triangle models were considered to be complete networks. PDP models have also been extended to Chinese (Yang, McCandliss, Shu, & Zevin, 2009), German (Hutzler et al., 2004), Hebrew (Velan, Frost, Deutsch, & Plaut, 2005), Italian (Pagliuca & Monaghan, 2009) and Japanese (Ikeda et al., 2016).

While a truly comprehensive consideration of computational models of reading and reading development would demand a comparison of the above mentioned modelling approaches, this thesis focusses on the DRC model exclusively. Having reviewed the relevant literature concerning models of reading, it was decided to focus on the DRC model, as the dual-route theoretical framework has been previously employed in explaining a wide range of effects on Turkish adult single word/ nonword naming such as Lexicality (Raman, 2003) and Frequency (Raman et al., 1996). This thesis therefore represents an iterative extension on this important work in Turkish psycholinguistics.

The rationale for adopting a dual-route approach as the dominant theoretical and interpretative framework is partly driven by methodological and technical issues, highlighted in Chapter 6, and is not reflective of the author’s theoretical position of a perceived superiority of the DRC modelling approach over others.

1.6 Thesis Aim

The primary aim of this doctoral research project was, therefore, to apply contemporary cognitive, computational and psycholinguistic theories and methods to the exploration of reading, reading development and disorder using the Turkish orthography as the medium of choice. This choice was motivated by the underlying orthographic transparency of the Turkish writing system in a preliminary attempt to redress the above highlighted Anglo-centric nature of research that has been carried out in reading research to date (Share, 2008). This aim was studied in several different ways:

1. by exploring the concept of orthographic transparency using a quantitative approach (Chapter 2)
2. by highlighting the need for better psycholinguistic tools to investigate Turkish leading to the development of the Turkish Lexicon database (Chapter 3)
3. by investigating reading development in typically developing Turkish children (Chapter 4)
4. by investigating the behavioural and cognitive manifestation of reading disorder in Turkish children (Chapter 5)

5. by developing a Turkish version of the dual-route cascaded (DRC, Coltheart et al., 2001) computational model of reading aloud, and testing the newly created model against the human word and nonword reading-aloud data (Chapter 6)

It could be argued that the multidimensional nature of this thesis is either its primary strength, allowing for a multidisciplinary approach, or its central limitation, in that the scope of the current research is too broad. However, by maintaining a synergy throughout the current thesis, it is expected that any perceived limitation can be addressed. Except for Chapter 2, which was carried out to place the Turkish orthography among other alphabetic writing systems, all other chapters establish a clear link between each other. For example, the data for Chapter 4, a relatively large study carried out regarding understanding the factors that influence reading development, are also used in Chapter 5, as a benchmark against atypical reading development, as well as in Chapter 6 where children's accuracy and RT data were directly compared against newly developed computational models of reading in Turkish. The following section will outline the structure of this doctoral thesis.

1.7 Thesis Outline

To address the above-stated research aims, the current doctoral thesis contains five exploratory investigations across five chapters. Chapter 2 introduces the reader to the historical and current conceptualization of Orthographic Depth. Previous attempts at quantifying these differences have been overly crude in their approach, and all existing methodologies have limitations. The chapter focuses on recent attempts at developing measures in order to generate quantitative indices of orthographic transparency. The reader will then be introduced to the key characteristics of the Turkish writing system before being presented with the quantitative models developed for this thesis. A variety of methodologies, including onset-entropy and whole-language phonemic transcription, were explored. The models will be compared to each other as well as offering a cross-linguistic comparison to other models in different orthographies.

Chapter 3 introduces current approaches to the development of psycholinguistic resources. The chapter then provides an overview of the available resources in Turkish, which are scarce. The methodological approach adopted to create a new, widely available corpus for use in Turkish psycholinguistic database is covered. Finally, the new database was validated using a lexical decision task in adults, and the naming task used in Chapter 4 is used to validate a subcorpus that is designed for use with children.

Chapter 4 examines the limited psycholinguistic research carried out with monolingual Turkish children to date. The lack of standardized measures needed to investigate the cognitive and literacy constructs used in this thesis are addressed by the development of a battery of tasks that forms the core of the psycholinguistic component of the study. Chapter 4 will also discuss the motivations and rationale for the tasks selected in the pilot investigation before presenting the main study results and conclusions carried out with 130 Turkish-speaking children.

Chapter 5 offers an overview of reading disorder with a particular focus on reading disorder in transparent alphabetic orthographies such as Finnish, Italian and Spanish. To the best of the author's knowledge, this was the first reported study to comprehensively examine the cognitive profile of developmental dyslexia in Turkish and results are discussed using both a group and a multiple case design approach using chronological and younger typically developing controls from the Chapter 4 Turkish monolingual study.

Chapter 6 provides an overview of recent developments regarding computational modelling in visual word recognition research. The chapter will focus on current attempts to model self-learning (unsupervised). A model of reading development using a three-staged vocabulary learning paradigm will be reported. The preliminary work towards a Turkish version of the Dual Route Cascaded model will also be introduced as well as comparing the computational simulation data with the behavioural data in Chapters 4 and 5.

Finally, Chapter 7 presents a general discussion based on findings from all experiments and concludes the thesis with a summary, implications of the findings, and limitations of the study as well as possible future directions in this area of research.

CHAPTER 2: TOWARDS THE QUANTIFICATION OF TURKISH ORTHOGRAPHY

2.1 PREFACE

A fundamental question in psycholinguistic research aims to address what is universal and what varies systematically with the mapping of spoken language to writing. It is therefore of critical importance to ascertain the extent to which conclusions from reading in a particular language can be generalized to others, and equally, what specific experimental effects are restricted to the particular orthography under investigation (Frost, 2012; Share, 2008). Theoretically, exploring how and why the cognitive mechanisms of reading diverge across orthographies will shed light into how universal systems and specific properties of orthographies interact. In alphabetic orthographies, the transparency/ consistency in which letters map onto sounds have been shown to influence central aspects of both reading development (Seymour, Aro, Erskine, 2003; Landerl et al., 2013) and skilled reading (Frost, Katz, & Bentin, 1987; Ziegler, Perry, Jacobs, & Braun, 2001). The following section will consider the historical and current debate regarding orthographic transparency in alphabetic writing systems, namely the progression of theoretical frameworks and current efforts being made to quantify and model orthographic transparency.

2.2.1 ORTHOGRAPHIC DEPTH AND READING IN ALPHABETIC WRITING SYSTEMS

Orthographic Depth (OD) refers to the reliability of grapheme-phoneme correspondences (GPCs) in alphabetic writing systems (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Frost, Katz, & Bentin, 1987) and is rooted in classic dual-route models of reading (e.g. Forster and Chambers, 1973; Coltheart, 1978). For example, a completely shallow or transparent orthography like Finnish has 29 letters and 29 phonemes and is characterized by consistent one-to-one mapping of graphemes to phonemes (Seymour, Aro, Erskine, 2003), resulting in a predictable pronunciation. Different alphabetic orthographies that have near one-to-one GPCs are Hungarian, Italian, Spanish, Greek, and as will be demonstrated in this chapter, Turkish. In contrast, a deep or opaque orthography like English is characterized by multi-letter graphemes in addition to context-dependent rules and morphological effects which consequently lead to many-to-many mappings of graphemes to phonemes, and thus the effect is a large degree of ambiguity in pronunciation. In this respect, OD can be viewed as a single-dimensional continuum (Figure 1). Moreover, Katz and Frost (1992) further characterize orthographic depth into three underlying concepts, complexity, consistency, and completeness. However, Schmalz, Marinus, Coltheart and Castles, (2015) state that how these three constructs relate to each other is vague and it remains to be established whether each of them influences reading in different ways. This pertinent issue will be discussed further below.

Grapheme-to-phoneme correspondence	Ordering	Language
Transparent/shallow	1	Finnish
1 grapheme-1phoneme	2	Welsh
	3	Italian
	4	Ladin
	5	Serbo-Croatian
	6	Macedonian
	7	Spanish
	8	Catalan
	9	Portuguese
	10	Korean
	11	Hindi
	12	German
	13	Danish
	14	Dutch
	15	Lao
	16	Khmer
	17	French
	18	English
	19	Japanese
Opaque/deep	20	Chinese
1 grapheme-many phonemes	21	Arabic
Many GPC exceptions/irregular words	22	Hebrew

FIGURE 3: ORTHOGRAPHIC DEPTH OF SEVERAL ORTHOGRAPHIES (FROM FIG. 2.4 IN PERFETTI & DUNLAP, 2008)

2.2.2 The Orthographic Depth Hypothesis

Historically, initial interest into orthographies with differing systematicity of GPC rules focused on the highly transparent, biscriptal Serbo-Croatian (Lukatela, Popadić, & Turvey, 1980), the opaque, un-pointed Hebrew (Bentin, Bargai, & Katz; 1984), and mixed Persian (Baluch & Besner, 1991). The investigation of three highly contrasted orthographies gave rise to several competing views. In its strong version, the Orthographic Depth Hypothesis (ODH; Katz & Frost, 1992) posits that readers adapt their processing strategy along two different reading routes depending on the type of GPC of the language being read. As such, the strong version of the ODH proposes that transparent orthographies can be read by applying sublexical GPC procedures, in which each grapheme is sequentially mapped to its corresponding phoneme. Furthermore, the ODH proposes that deeper, more opaque orthographies such as English require the lexical procedure to access words from whole word orthographic memory. In addition, the weak version of the ODH suggests that both nonlexical and lexical routes are available to readers of all writing systems, but the relative involvement of each processing route reflects the transparency of a given orthography.

Convincing support for the strong version of the ODH comes primarily from the results of several cross-linguistic studies. Katz and Feldman (1983) report that semantic priming effects were found in lexical decision and naming tasks in English but only found the same effect during the lexical decision task in Serbo-Croatian thus suggesting that a related context does not facilitate naming relative to an unrelated context resulting in a reliance on sublexical processing. In a similar study, Frost, Katz, & Bentin (1987) report that they found greater word frequency effects and differences in reaction times in naming between words and nonwords for Hebrew in comparison to Serbo-Croatian and English. These authors conclude that the lack of frequency effects in the oral reading of Serbo-Croatian indicates a reliance on sublexical processing. However, from a critical perspective, these initial null findings were based exclusively on reading Serbo-Croatian (Besner & Smith, 1992) which has also reported semantic priming effects in other studies (e.g. Carello, Lukatela, & Turvey, 1988). Additionally, there is a corpus of research that demonstrates lexical involvement in the oral reading of many transparent orthographies such as Spanish (Sebastián-Gallés, 1991), Italian (Tabossi & Laghi, 1992), Turkish (Raman, Baluch, & Sneddon, 1996), Finnish (Wydell, Vuorinen, Helenius, & Salmelin, 2003) in a silent reading task with magnetoencephalography (MEG), and Japanese Kana (Rastle, Havelka, Wydell, Coltheart & Besner, 2009). Moreover, phonological processing has been demonstrated to be involved in word recognition in deep orthographies such as English (Ziegler, Perry, Jacobs, and Braun, 2001) or even in Japanese Kanji, a logographic/ morphographic script (Wydell, Patterson, & Humphreys, 1993) leading to a general rejection of the strong version of the ODH.

2.2.3 Alternative Universal Views of Orthographic Depth

Alternative views to the ODH stipulate that the similarities between reading processes across orthographies are greater than their differences. Two such theoretical frameworks will now be considered. One alternative view to ODH referred to as the Universal Hypothesis, argues that the lexical route is the most dominant one regardless of the transparency of the orthography (Seidenberg, 1985; Baluch & Besner, 1991; Besner & Smith, 1992; Baluch, 1993). Evidence rejecting the universal hypothesis in favour of the weaker version of the ODH comes from by Frost and colleagues (1987) (as reported above) and by Frost (1994). Frost (1994) carried out a word priming experiment using un-pointed Hebrew words and found that both word frequency and semantic priming effects were significant. However, according to Frost (1994), when the same words were pointed, both the word frequency and semantic priming effects diminished. Based on these findings, Frost argued that the pointed Hebrew script, which is highly transparent, promotes the use of a sublexical reading strategy even though participants were more accustomed to reading the words un-pointed.

The universal phonological principle (UPP) (Perfetti, Zhang, & Berent, 1992) is fashioned as a universal claim that reading involves the early use of phonology and subsequently, word reading activates phonology at the smallest unit allowed by the writing system. The UPP claims that the specific mapping differences across writing systems and orthographies affect fine-grain reading procedures within a wider universal dependence of reading on oral language. Thus, concluding that reading, irrespective of the writing system, involves phonology (Perfetti & Zhang, 1995). The UPP and weak version of the ODH are not incompatible (Perfetti, Cao, & Booth, 2013) and have recently been refined and combined to generate the Psycholinguistic Grain Size Theory (PGST) (Ziegler & Goswami, 2005) which claims that reading processes assemble phonology in accordance to the grain size of the orthography. The PGST will be discussed in greater depth in the next section.

2.2.4 The Psycholinguistic Grain Size Theory

Until recently, the ODH was the foremost theoretical framework to guide cross-linguistic reading research (Wydell & Butterworth, 1999; Perfetti, Liu, & Tan, 2005; Ziegler & Goswami, 2005; Frost, 2012). Wydell and Butterworth (1999) described the case of AS, an English-Japanese bilingual adolescent with monolingual dyslexia in English. In order to account for the behavioural dissociation discovered in this study, Wydell and Butterworth proposed the hypothesis of transparency and granularity (Figure 2), propositioning that both universal and language-specific processes exist during reading. Additionally, the authors predict that writing systems where the print-to-sound relationship is transparent should not manifest a high incidence of developmental phonological dyslexia irrespective of granularity. Likewise, writing systems where the grain size is coarse (i.e. whole character or whole word) should similarly not manifest with a high incidence of developmental phonological dyslexia.

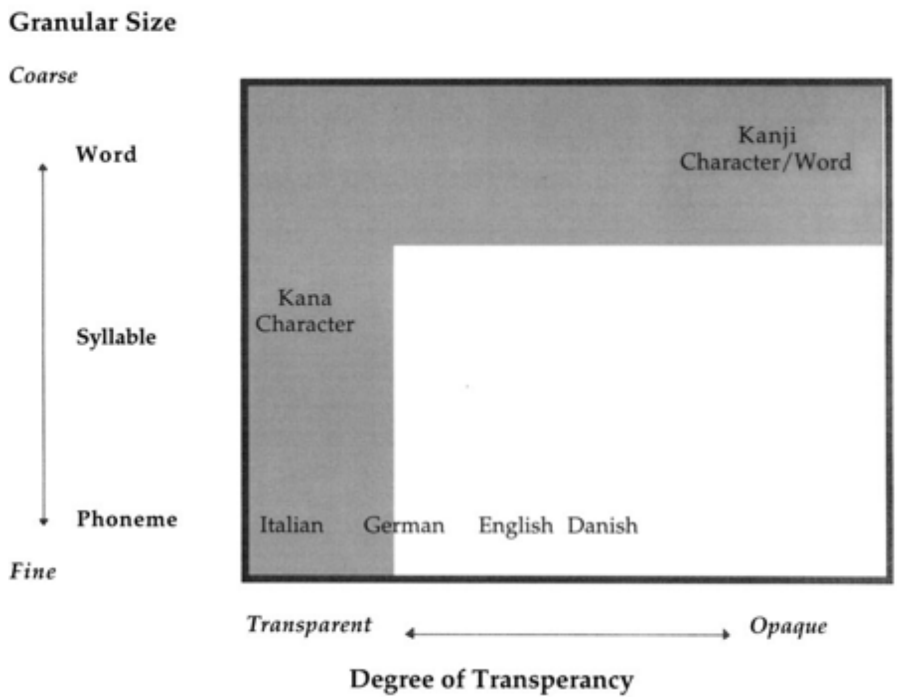

FIGURE 4: HYPOTHESIS OF GRANULARITY AND TRANSPARENCY AND ORTHOGRAPHY-TO-PHONOLOGY CORRESPONDENCE. THE SHADED AREA ON THE 'TRANSPARENCY' DIMENSION REPRESENTS ALMOST 100% TRANSPARENCY (FROM WYDELL & BUTTERWORTH, 1999).

The Psycholinguistic Grain Size Theory (PGST; Ziegler & Goswami, 2005) focuses on the influence of orthographic depth on reading acquisition and will be contemplated further below. Ziegler and Goswami (2005) posit that for phonological recoding to be effective, developing readers need to discover shared grain sizes between the orthography and phonology of their language. Furthermore, the PGST states that grain size in a given orthography is inversely proportional to the regularity of GPCs. In transparent orthographies, phonological regularity encourages sublexical decoding strategies. Divergently, in opaque orthographies, the availability of larger orthographic units, such as grapheme clusters or syllables promotes lexical, whole-word reading processing (Lallier, Carreiras, Tainturier, Savilla, & Thierry, 2013). This is mainly in line with the claims made by the weak version of the ODH with the distinction that whereas the ODH proposes a quantitative modification in the ratio of lexical-to-sublexical processing, the PGST suggests a qualitative change in the nature of sublexical processing (Schmalz, 2016).

2.2.5 Orthographic Depth and Reading Development

It has become increasingly evident over the last two decades that the characteristics of an orthography modulate reading processes and thus determining the cognitive mechanisms in which this occurs has important theoretical and practical implications (Schmalz, Marinus, Coltheart, & Castles, 2015; Spencer, 2007). The primary methodological approach thus far has been to carry out cross-linguistic investigations between English, and other (more transparent) alphabetic European orthographies. For example, Wimmer and Goswami (1994) compared German and English primary schoolchildren on the reading of digits, number names and nonwords (produced by substituting the onsets and rimes of number names). They found that nonword reading was significantly slower and inaccurate in English. Similar studies have reported the same pattern of poorer nonword reading ability in English when compared with Greek (Goswami, Porpodas, & Wheelwright, 1997) and Spanish (Goswami, Gombert, & de Barrera, 1998).

Recently, there has been a movement towards large-scale comparisons of orthographies that differ in orthographic transparency. For instance, Seymour, Aro, and Erskine (2003) carried out a seminal cross-linguistic study of the early stages of learning to read across 14 European countries. The study reported that by the end of Grade 1 reading accuracy was at ceiling level in most transparent languages (e.g., Italian, German, Greek, Spanish, and Finnish) whereas the word reading accuracy in less transparent languages (e.g., Portuguese, French, and Danish) was lower, around 80%. Interestingly, reading accuracy for English readers, the least transparent of the orthographies investigated, was 34%. These studies provide evidence that orthographic transparency regulates the ability in which children transform letter strings into a phonological code (the alphabetic principle), a process referred to as phonological recoding (Share, 1995).

Intriguingly, it has been advocated that phonological recoding provides the foundation for a self-teaching mechanism that facilitates developing readers in autonomously forming an orthographic lexicon (Bowey & Muller, 2005; Share, 1995; 1999). This, of course, also has implications on the manifestation of reading disorders, which will be covered in Chapter 5. Additionally, Moll et al. (2014) carried out an extensive study which evaluated the concurrent predictions of phonological processing and rapid automatized naming (RAN) for reading development across five orthographies with varying degrees of transparency and found that phonological processing and RAN both accounted for significant amounts of unique variance in reading development across the five orthographies (i.e. English, French, German, Hungarian and Finnish). They found that the general pattern of reading development was comparable across orthographies with RAN being the best predictor of reading speed and phonological processing being the best predictor of reading accuracy. Although a consensus concerning the orthographic transparency classification of alphabetic writing systems exists, there is a distinct lack of quantitative research regarding this topic. The following sections will focus on recent trends in redefining and quantifying orthographic transparency.

2.3 Towards the quantification of orthographies

The development of a linguistic quantification scheme would be useful in testing the claims made by both the weak version of the ODH and the PGST as detailed estimates of both transparency and granularity are deficient in the literature. Before attempting to explore quantification schemes, it is paramount that a clear definition of orthographic depth is established in order to avoid falsely attributing any behavioural differences observed between orthographies to orthographic depth *posthoc* (Schamlz et al. 2015). The sections below will consider novel reinterpretations of orthographic depth as well as both subjective and objective approaches taken in quantifying orthography.

2.3.1 Orthographic Depth: More than one construct?

As previously mentioned in this chapter, Katz and Frost (1992) extricate three concepts underlying orthographic depth: complexity, consistency, and completeness. Each will now be considered separately.

In visual word recognition, the construct of complexity refers to the existence of multi-letter rules, where several letters are needed to characterise a single phoneme (Bosch, Content, Daelemans, & de Gelder, 1994; Schmalz, Beyersmann, Cavalli, & Marinus, 2016). This may also include context-sensitive regularities wherein neighbouring letters affect a grapheme's pronunciation. There are several studies (e.g., Marinus & de Jong, 2010; Rastle & Coltheart, 1998; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998) whose findings indicate that the application of multi-letter rules produces competition between the pronunciation of the single letters and the grapheme's pronunciation, consequently inhibiting the sublexical route, and providing additional time to access lexical information via the lexical route (Katz & Frost, 1992).

The second concept is consistency, which conveys the presence of several pronunciations for a given letter string. It can be conceptualized as a measure of variability in the GPCs of a writing system (Perry, Ziegler, & Coltheart, 2002). Consistency can be defined at various levels typically using the graphemic level or of larger grain size as evidence suggests that taking into account larger parts of the syllable reduces ambiguity (e.g. Kessler & Treiman, 2001; Treiman et al., 1995; Ziegler et al., 1997). Therefore, this approach is commonly applied to inconsistent orthographies such as English. Within the modelling literature, consistency is related with connectionist models (e.g. Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996) while the concept of regularity is connected to dual-route models of reading (Coltheart et al., 2001). The regularity approach assumes that symbolic transcription procedures govern GPC and that irregular mappings are a consequence of the violation of these rules (Ziegler, Perry, & Coltheart, 2003). It is important to note that consistency and regularity are not mutually exclusive concepts and words can be classified as Consistent-Regular; Inconsistent-Regular; Inconsistent-Irregular, and Exception words (See Glushko, 1979). In the event where GPC diverges from the phonological ideal, that is, one –to one mapping, the most frequent mapping is considered to be regular, and the other mappings are said to be irregular (Zevin & Seidenberg, 2006). In line with recent developments in the field (see Schmalz, Marinus, Coltheart, & Castles, 2015 for a review), this thesis adopts a neutral view regarding the argument between dual-route and connectionist models and as such consistency/ regularity will be referred to as predictability for the remained of this chapter.

The third concept of completeness refers to the information (or lack thereof) that is conveyed by sublexical correspondences. In English, the presence of heterophonic homographs (words that are spelt the same but have distinctive pronunciations and a different meaning) is an excellent example of incomplete sublexical information (Pacht & Rayner, 1993). For instance, the word "read" has two pronunciations and context is needed to activate both the correct phonology and semantic representation of the word. Another example of incompleteness comes from abjad scripts such as Hebrew and Arabic. Within these orthographies, vowels are frequently not represented in writings (see reference to un-pointed Hebrew earlier in this chapter) and, as such, many words have indistinguishable consonant patterns (Frost & Bentin, 1992). Similar to English, lexical-semantic information is needed to identify and pronounce the word under investigation. The concept of incompleteness is not considered to be theoretically pertinent for the investigation of alphabetic scripts (Schmalz, Beyersmann, Cavalli & Marinus, 2016) and hence will not be considered any further in this thesis.

2.3.2 Methods of quantification

Now that orthographic depth has been further conceptualized, the following section will provide an overview of recent attempts to quantify these constructs. The relative strengths and weaknesses of each approach will be evaluated with a view of developing a quantification scheme for Turkish.

Complexity, as described above, has recently been quantified using the ratio of letters to phonemes (Schmalz, Beyersmann, Cavalli, & Marinus, 2016). The authors define simple words as having a letter-to-phoneme ratio of one and complex words as having a letter-to phoneme ratio above 1, indicating the occurrence of multiple letters that correspond to a single phoneme. In order to dissociate complexity from predictability, Schmalz and colleagues (2016) give the example of French, which is characterized as being highly predictable but with a high degree of complexity (van den Bosch Content, Daelemans, & De Gelder, 1994). Classifying words in a dichotomous manner of complexity, as stated above, in addition to manipulating frequency, Schmalz, Beyersmann, Cavalli, & Marinus (2016) report, using both frequentist ($\beta = -0.06$, $t = 3.5$, $p = .0005$) and Bayesian (BF = 37.9 (±1.1%)), evidence of an interaction between frequency and complexity. This finding implies that the frequency effect is stronger for words with complex correspondences. Additionally, the authors assert that this finding provides evidence that independent of predictability, complexity impairs the sublexical route which subsequently *"leads to a relative increase in the degree to which the lexical route contributes to the final output"* (p.11).

Another approach to complexity utilizes data-orientated learning algorithms in order to provide two indices of orthographic depth, namely graphemic parsing and degree of redundancy (van den Bosch, Content & Daelemans 1994). Of particular relevance is graphemic parsing, which was measured by applying a computationally obtained parsing mechanism to record all GPCs found in a test set of words. The authors found that for the orthographies under investigation, Dutch, English and French, the rate of successful parsing was 21.3%, 24.5% and 12.9% respectively. The results suggest a degree of similar complexity concerning the parsing of regular graphemes between the three orthographies. However, there are also distinct differences between the three corpora investigated with the French Grapheme- Phoneme Correspondences Extraction (GPCE) model performing worst in correctly aligned words and the English GPCE model performing the best. Additionally, the application of Decision Tree Learning to the three corpora found that English showed the least redundancy and hence can be considered to be more irregular than French and Dutch. A summary of the general findings is expressed in Figure 3.

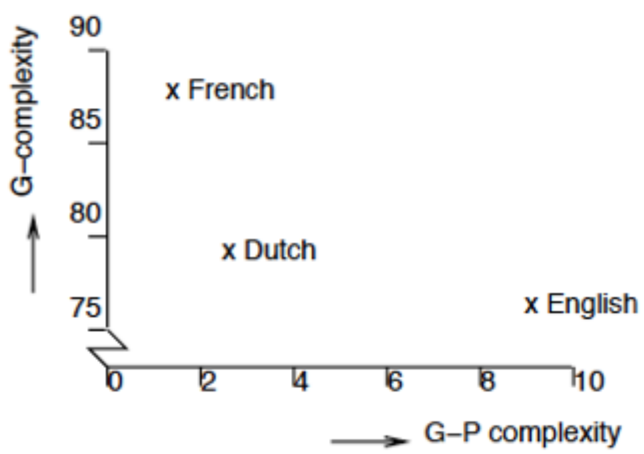

FIGURE 5: TWO- DIMENSIONAL ORTHOGRAPHIC DEPTH SPACE WITH 'X' S MARKING THE THREE CORPORA. G-COMPLEXITY STANDS FOR THE COMPLEXITY OF GRAPHEMIC PARSING, AND G-P COMPLEXITY STANDS FOR THE COMPLEXITY OF GRAPHEME-TO-PHONEME CONVERSION (FROM VAN DEN BOSCH, CONTENT & DAELEMANS 1994).

The final method of the quantification of complexity is the use of rule-based computational models such as DRC (Schamlz et al., 2015) in which the number and proportion of complex (multi-letter) GPC rules are understood to be a measure of complexity. Using this approach, the DRC has thus far been implemented in Dutch, English (Coltheart et al., 2001), French (Ziegler, Perry, & Coltheart, 2003), German (Ziegler, Perry, & Coltheart, 2000), and Italian (C. Mulatti, personal communication, 25 May 2014 as cited in Schmalz, Marinus, Coltheart, & Castles, 2015). The pattern of the number of complex rules is largely in line with previous subjective classifications of orthographic transparency, although this approach has several limitations. The DRC is limited to the computation of monosyllabic words and complexity measures derived from monosyllabic words alone are unlikely to provide reliable estimates of the GPCs of the language in question (Borgwaldt, Hellwig & Groot, 2004; Kessler & Treiman, 2001; Protopapas & Vlahou, 2009). Furthermore, the exclusive use of monosyllables in the computation of complexity limits the applicability of such an approach for cross-linguistic comparisons as languages vary significantly in the number of monosyllabic words used in a given language (Protopapas and Vlahou, 2009). Indeed, the corpus-based approach taken in this chapter demonstrates that the DRC method is redundant in orthographies such as Turkish as monosyllables represent under 1% of words in the corpus. Any such approach in quantifying an orthography such as Turkish would need to be representative of the language.

Predictability is a measure of how probable a pronunciation of a word is based on similarly spelt words. The most common approach to date has been to evaluate the DRC model's performance and therefore calculate the percentage of GPC mappings that obey pronunciation rules (predictable words). Using this rule-based approach, Ziegler and colleagues (2000) compared predictability in German and English monosyllables and reported that the percentage of correct rule application is 90.4% for German and 79.3% for English words. Like the critique of the DRC approach for the calculation of complexity, the restricted use of monosyllables for the quantification of orthographic transparency in orthographies such as Turkish is inadequate. Furthermore, the calculation of predictability as the proportion of correct mappings is limited in that such an approach is inherently unable to discriminate between cases with many and few alternative mappings. To overcome this limitation, current approaches into the predictability of alphabetic orthographies have focused on communicating variability in terms of entropy values (Borgwaldt, Hellwig, & de Groot, 2004; Martensen, Maris, & Dijkstra, 2000; Protopapas & Vlahou, 2009).

Entropy can be conceptualized as an information-theoretic notion that quantifies uncertainty using the h-index (Shannon, 1949). The h-index may be defined using the following algorithm:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i .$$

Entropy (H) is computed as the negative sum of *n* phonological (orthographic) mapping alternatives with probability P_i for the *i*th alternative multiplied by base-2 logarithms. In terms of orthographic transparency, entropy computes uncertainty in the prediction of letters/ graphemes by phonemes and vice versa. If a particular grapheme maps unambiguously to a specific phoneme, i.e. the mapping is absolutely predictable, the corresponding entropy value will be zero. Conversely, the more alternative pronunciations a grapheme has, the higher its entropy value is. For example, Borgwaldt, Hellwig and Groot. (2004; 2005) carried out analyses of entropy for English, Dutch, German, French, Hungarian, Italian, and Portuguese. In an attempt to overcome the limitations of restricting the analysis to monosyllables, as discussed above, Borgwaldt and colleagues used word-initial sound-spelling correspondences, using all the words in each language. An advantage of using word onsets is that languages with diverse orthographic and phonological properties can be objectively compared since all languages have word onsets. Thus, onset entropy offers an objective and assumption-free index of orthographic transparency.

Borgwaldt, Hellwig and Groot (2004) revealed that none of the orthographies studied to date had an ideal one-to-one mapping between letters and sounds. Also, word-initial letter entropy in Italian, Dutch, and English was significantly correlated with word-naming RT. Though, one potential critique of word-initial entropy measures is that there is little evidence to suggest that such an approach characterizes an unbiased representative sample of all possible GPC mappings. To address this issue, Protopapas and Vlahou (2009) carried out a whole-word entropy analysis of Greek. The authors found that compared to word-initial single-letter mapping to phonemes, whole-word letter-to-phoneme entropy values were approximately three times greater.

Clearly, the use of whole-word approaches captures significantly more variation than word onset approaches. However, initial-letter onset offers a broader scope for cross-linguistic applicability. Given the relative strengths and benefits, both methodologies will be considered in the current analysis.

2.3.3 Other Methodological Considerations

Whilst the main approaches to the quantification of orthographic transparency have been considered above; the following section will contemplate the characteristics of the counts being used to calculate transparency indices, namely the choice to employ word form or lemma databases as well as token or type counts of frequency.

Transparency indices may be calculated based on either word form or lemma databases in which word form databases possess all of the possible morphological variants of a lemma. In contrast, lemma databases contain only the root form of a word. Most psycholinguistic studies to date have focused on lemma form (Hofmann, Stenneken, Conrad, & Jacobs, 2007) as word forms have been found to moderately activate its corresponding lemma entry in the mental lexicon (New, Brysbaert, Segui, Ferrand, & Rastle, 2004). However, it has also been argued that that lemma measures systematically over- or underestimate the frequency of sublexical units that occur in inflective morphemes (see Hofmann, Stenneken, Conrad, & Jacobs, 2007 for details). Given the highly inflective nature of agglutinative writing systems such as Turkish, and that sound changes are produced in the stem by the process of suffixation, it was decided to focus on the creation of word-form models of orthographic consistency.

In recent word-onset entropy studies, both word form types (Borgwaldt, Hellwig & Groot., 2004) and lemma types (Borgwaldt, Hellwig & Groot., 2005) have been used. It has been argued that it may be superior to measure indices of transparency using word forms (i.e., token instead of type counts) in order to account for the well-established role of word frequency effects on visual word recognition (Balota, Yap, & Cortese, 2006). Following the recommendation of Hofmann Stenneken, Conrad, & Jacobs (2007), a more conservative methodology will contemplate both type and token frequency counts (Protopapas & Vlhaou, 2009) as is the case in this current investigation.

Additionally, both type and token frequency counts are used to investigate orthographic transparency. Type counts of sublexical units comprise of the total amount of words that contain the given unit. Alternatively, token counts are calculated by summing the number of occurrences in a text corpus of the words that contain the unit. The theoretical importance of considering frequency counts stems from the rationale that in addition to the number of different pronunciations, the relative frequency of these alternative mappings contributes to the ensuing entropy value. If some of those pronunciations appear only very rarely, and if there is one truly dominant correspondence, the entropy value is lower than in the case of all pronunciations occurring with approximately the same frequency. This means that the impact of exceptional pronunciations is rather marginal. However, the methodological option to use type or token measures is contentious, since both are thought to be independently linked with lexical processing. For example, Conrad, Carreiras, and Jacobs (2008) carried out a lexical decision task and reported that syllable frequency (token-based) was related to an inhibitory effect on lexical access. Conversely, the use of a type-based measure was correlated with a facilitative effect in lexical decision. To account for the ambiguous findings, the authors suggested that both measures are related to distinctive processing points during the visual word recognition of words.

In a contrasting account, Moscoso del Prado Martín, Ernestus, and Baayen (2004) propose that both token- and type-based effects can be accounted for by a shared, token-based mechanism. To demonstrate this, they modelled Dutch past tense formation with a simple recurrent network that exhibited both word frequency effects (token-based) and word analogical effects (type-based) that corresponded well with human behavioural data. In order to circumvent the diverging evidence regarding the relative contribution of token and type frequency effects, both will be considered in the current analysis in this thesis.

Now that the methodological considerations of quantifying an orthography have been taken into consideration, the next section will focus on providing an overview of the Turkish orthography and its defining characteristics.

2.4 THE TURKISH ORTHOGRAPHY

Modern Standard Turkish (Henceforth Turkish) is the official language of the Republic of Turkey and Cyprus and by current estimation, is fluently spoken by 80-90 million people worldwide (Kuribayashi, 2012). Outside of Turkey and Cyprus, there are large Turkish-speaking populations in historical Ottoman lands such as Bulgaria, Macedonia, Iraq, Algeria, Egypt and Syria as well as recent large emigrant populations in Germany, France, the Netherlands, Austria and the United Kingdom (Jørgensen, 2003). Turkish is considered to be a member of the South-Western Turkic branch or Oğuz subdivision of the Altaic language family. Amongst the Turkic language family, Turkish is considered to be the most culturally and politically significant in addition to being the most commonly spoken (Katzner, 2002).

Turkish is considered to be the successor of Anatolian Oğuz Turkish introduced into Anatolia during the 11th Century AD by Selçuk Turks and the subsequent Ottoman Turks. Historically, the first script to be used by the Turkish people was the runic Köktürk script (Róna-Tas, 2015). Evidence for its use dates back to 688-692 AD to the Çoyren Inscription, succeeded by the Orhun Inscription in 732-733 AD (Scharlipp, 1994). Following this period, the Uygur script was adopted and used from 745 to 970 AD. The widespread adoption of Islam among the Turks from the 10th century onward saw the Turkish language come under heavy influence from Arabic and Persian. The Arabic script was modified by the introduction of diacritics to mark vowels; though the new script remained inadequate to transcribe the eight vowels used in spoken Turkish. This consequently led to problems in deriving the correct phonology from print resulting in unpredictability that is linked with consonantal scripts. As a consequence, Çapan (1989) reported that in 1927, only about 10% of the Turkish population was considered to be literate.

Turkish language reform in 1928 is often cited as one of the most significant language reform movements in the world (Lewis, 1999). It involved script reformation by transcribing the sounds of the spoken language in a modified Latin alphabet and purification of the Turkish language, leading to the elimination of foreign words and structures, mainly from Arabic and Persian. The aim was the creation of an alphabet where each spoken sound in standard Turkish was represented by a single letter in the alphabet and ultimately providing an optimal environment to rapidly increase the acquisition of literacy skills in the general population. In the next section, the characteristics of Turkish orthography are introduced.

2.4.1 The Characteristics of Turkish

The modern Turkish orthography is composed of a 29-letter alphabet of eight vowels and 21 consonants based on a modified Latin script though the Turkish writing system contains 32 graphemes; the 29 letter forms of the Turkish alphabet as well as the graphemes < â >, < î > and < û > (TDK Yazim Kilavuzu, 2013). Of interest to this thesis is that, in Turkish, the relationship between orthography and phonology is said to be near one-to-one (Raman, 2006; 2011). However, there is also much debate as to the exact number of phonemes in Turkish (Koşaner, Birant, & Aktaş, 2013). This is an issue that will be further considered in the models produced in this chapter. Additionally, grapheme-phoneme and phoneme-grapheme conversions are both regular and consistent, resulting in bi-directional transparency. The presence of homographic loan and compound words (mainly from Arabic and Persian) as well as special cases of morphophonology, such as vowel length alternations, introduces the only ambiguities in grapheme-phoneme conversion (Göksel & Kerslake, 2004; Külekçi & Oflazer, 2006).

Turkish is an agglutinating inflectional language, and the neutral word order is subject-object-verb (SOV). In Turkish, words are typically composed of long sequences of morphemes (smallest meaningful unit) with each morpheme representing one morpheme or a meaning unit. Additionally, there are several regular morphophonemic processes such as vowel harmony, consonant assimilation and elisions that condition morphemes (Oflazer & Inkelas, 2006).

Vowel harmony is a prominent feature of spoken Turkish in which vowels work in four pairs with corresponding front/back and rounded/unrounded sounds. Vowel harmony regulates the vowels of a word by conditioning each subsequent vowel by the vowel, which immediately precedes it (Durgunoğlu & Öney, 1999). This harmony is also observed in most grammatical suffixes, whereby the vowel in the suffix harmonises with the last vowel in the word though exceptions exist (Clements & Sezer, 1982). The process in which the voicing of a consonant becomes analogous to that of the neighbouring consonant is referred to as voicing assimilation and, in Turkish, is significant for consonants because consonants are distinguished in terms of voicing. The process of voicing assimilation is most apparent in stop-initial and affricate-initial suffixes. Additionally, syllable structure in Turkish is said to be relatively simple in that four simple syllabic structures (V, VC, CV, and CVC) constitute 98% of all Turkish syllables (Durgunoğlu & Öney, 1999). Furthermore, Turkish syllable structure allows both open and closed syllables but no onset clusters in native words. Turkish syllabic structure, like French, (Sprenger-Charolles & Siegel, 1998) is said to possess clear boundaries.

2.4.2 Soft G

An ongoing debate regarding the phonology and phonetics of Turkish is the status of soft ‘g’, which is represented as <ğ> in the orthography of the Turkish alphabet. Depending on position, ‘ğ’ can be described as “*a stop, a fricative, an approximant, or a vowel*” (Ünal-Logacev, Fuchs, & Żygis, 2014). Also, ‘ğ’ never occurs in the word-initial position and is always preceded by vowels; usually in the context of high vowels. For this thesis, ‘ğ' will be treated phonologically like a consonant whilst taking into consideration the lengthening of the preceding vowels.

2.4.3 A note on the circumflex and stress

There is increasing evidence that the circumflex is falling out of use in contemporary written Turkish (Inkelas, Küntay, Sprouse, & Orgun, 2000). Though this adds a degree of ambiguity to pronunciation, for example, *hala* "paternal aunt" vs *halâ* "void" vs *hâlâ* "yet", most of the ambiguity can be resolved with context. To more accurately reflect modern standard Turkish, the three graphemes < â >, < î > and < û > that contain the circumflex will not be considered any further in this thesis.

Finally, the Turkish orthography does not mark stress, and there are a limited amount of minimal pairs in Turkish diverging only in the position of stress. However, there are several items which follow neither of the regular stress placement rules (i.e. final stress, for the majority of words, and a more complex pattern of non-final stress, referred to as Sezer stress, for place and foreign names used in Turkish (see Sezer, 1981; Inkelas, 1999).

2.5 METHOD

The section below will consider the application of the methodological issues discussed in-depth above. Furthermore, a detailed description of the lemma and word form databases created for the purpose of this thesis will be outlined and analysed.

2.5.1 TEXT CORPUS SELECTION AND CREATION

Due to the productive morphology and parsing ambiguity in agglutinative languages such as Turkish, a large corpus is needed for a robust estimation of an absolute vocabulary. The TS Corpus v2 (Sezer & Sezer, 2013; http://tscorpus.com) is a large general-purpose Turkish Corpus containing 491 million POS-Tagged tokens (including punctuation marks) and 4.9 million unique word forms that builds on and extends the BOUN Corpus (Sak, Güngör, & Saraçlar, 2007). The BOUN Corpus was created from collecting web pages from three Turkish daily newspapers (212M tokens) as well as an extensive sampling of Turkish webpages (279M tokens). In line with the view held by Baayen, Milin, and Ramscar (2016), a corpus encompassing newspapers and webpages may offer a more suitable choice than corpora that have been constructed by other means because of the seemingly higher share in the overall reading experience of the subjects.

Due to the nature of web corpora, there was a need to pre-process the raw text. A set of tools (ITU Turkish NLP Pipeline; Eryiğit, 2014; http://tools.nlp.itu.edu.tr) designed for such purposes was used and will be outlined below:

- Firstly, a normalization procedure (Torunoğlu & Eryiğit, 2014) was carried out on the raw text using a cascaded method that combined both rule-based and machine learning methods for seven different layers of normalization, namely; letter case transformation, replacement rules and lexicon lookup, proper noun detection, deasciification (Adalı & Eryiğit, 2014), vowel restoration, accent normalization and spelling correction. An overall accuracy of 78.5% represents the highest results for Turkish text normalization of social media data available to date but would still need to be manually checked for this thesis.

- Following this, the normalized text was subject to the *isTurkish* module (Şahin, Sulubacak, & Eryiğit, 2013) in which a two-level Turkish morphological analyser, using flag diacritics, was used to validate the normalized text as morphologically legal for Turkish. Initial testing of the morphological analyser reports an accuracy of 99.8%.

- During manual checking, items (4% of tokens) including any non-Turkish characters, numerals, or symbols were rejected. Additionally, to restrict the size of the resulting databases, items (words) that had an absolute frequency below 1 part per million (ppm) were rejected from the final list.

The resulting normalized text was condensed into a word-form database which contained 198,236 unique types and 91,497,080 tokens. The database was reasonably free from spelling errors and contained few idiosyncratic items, such as proper names, foreign words not quite integrated as loans in the Turkish language, or very low-frequency words unlikely to be found in contemporary use.

2.5.2 Phonetic Transcription

As mentioned previously, there is a degree of disagreement as to the number of phonemes that best represent the Turkish language. To address this from a methodological perspective, two different phonetic transcription approaches will be considered, namely *METUbet* (Salor, Pellom, Çiloğlu, Hacıoğlu, & Demirekler, 2002) and *Grafofon* (Koşaner, Birant, & Aktaş, 2013). Each phonetic transcription system will now be considered in greater depth.

METUbet (Salor, Pellom, Ciloglu, Hacioglu, & Demirekler 2002) represents a letter-to-phoneme conversion rule set grounded on the phonetic symbol set first described in Ergenç (1995). The difficulties in digitalizing IPA symbols led to the authors adopting the symbols used in the Speech Assessment Method Phonetic Alphabet (SAMPA) dictionary (Wells, 1995). However, SAMPA symbols also have poor readability since they include characters such as numbers and punctuation symbols, and as a consequence, a new simplified alphabet called METUbet was developed.

METUbet has 39 phonetic representations as opposed to the 45 phonetic SAMPA representations (Salor et al., 2002). The difference in the number of phonetic representations is due to the open-short and closed-long forms of the letters u, ü, o, ö, and i being symbolised by the same phonetic symbol in the METUbet transcription. The closed-long forms of those letters appear when they are preceding soft g, which triggers the lengthening of those letters. Using a 60ms misalignment tolerance, 99.3% of phoneme boundaries were automatically placed. Grafofon (Koşaner, Birant, & Aktaş, 2013) was initially developed as an aid for course materials for teaching Turkish as a second language. Adopting a unified approach, the authors advocate a 32-phoneme system for Turkish. A distinct feature of the Grafofon system is that it also considers allophonic variation; although the authors report 63 possible allophones, the current study found 40 and 52 allophones in the word-onset and the whole-word measures respectively. Finally, Grafofon reports an overall 96% phoneme identification accuracy. Table 1 below shows the two phoneme mapping schemes used in this thesis.

TABLE 1: PHONEME VARIATION IN TURKISH WITH MAPPINGS OF IPA, METUBET AND GRAFOFON

Uppercase	Lowercase	IPA	METUbet	Grafofon
A	a	ɑ	AA	ɑ
		a	A	*ɑ*
B	b	b	B	*b*
C	c	*ʤ*	C	*ʤ*
Ç	ç	*ʧ*	CH	*ʧ*
D	d	d	D	*d*
E	e	e	E	*e*
		ɛ	EE	
F	f	f	F	*f*
G	g	g	GG	*ɟ*
		ɟ	G	*g*
H	h	h	H	*h*
İ	i	i	IY	*i*

		ɪ	IY	
I	ı	ï	I	*ɯ*
J	j	ʒ	J	*ʒ*
K	k	k	KK	*k*
		c	K	*c*
L	l	l	L	*l*
		ɫ	LL	*ɫ*
M	m	m	M	*m*
N	n	n	NN	*n*
		ɳ	N	
O	o	ɔ	O	*o*
		o	O	
Ö	ö	œ	OE	*ø*
		ø	OE	
P	p	p	P	*p*
R	r	r	R	*r*
		ɾ	RR	
		ʁ	RH	
S	s	s	S	*s*
Ş	ş	ʃ	SH	*ʃ*
T	t	t	T	*t*
U	u	U	U	*u*
		u	U	
Ü	ü	Y	UE	*y*
		y	UE	
V	v	v	VV	*v*
		ʋ	V	
Y	y	j	Y	*j*
		:ɪ	Y	
Z	z	z	Z	*z*
		ʐ	ZH	
Ğ	ğ	:	GH	*:*

To compute the entropy values for Turkish, all mono- and polysyllabic words were extracted from the relative corpora. Furthermore, word-initial letters from the word-form database were extracted and used to calculate the entropy values for the different letter-phoneme correspondences (Borgwaldt, Hellwig, & De Groot, 2005). For example, using the METUbet transcription with token frequencies, the entropy of the grapheme <g> would be:

$$- [0.891 \times \log_2 (0.891) + 0.109 \times \log_2 (0.109)] = 0.361 \text{ bits}$$

2.6 Results

At the word-onset level, the phoneme inventory size reported here ranges from 31 (Grafofon without allophones) to 33 (METUbet). When allophonic variation is taken into consideration, the phoneme/allophone inventory increases to 40. Additionally, at the whole word-level, the phoneme inventory size reported here ranges from 32 (Grafofon without allophones) to 38 (METUbet). When allophonic variation is taken into consideration, the phoneme/allophone inventory increases to 52. For both METUbet and Grafofon, further grapheme-phoneme pairs are possible although they were not identified in this corpus. A summary of the statistics related to the transparency of the Turkish orthography can be found below (Table 2).

TABLE 2: STATISTICS RELATED TO THE TRANSPARENCY OF THE TURKISH ORTHOGRAPHY. WHERE GRAPH IS GRAPHEME, TOT PAIR IS THE TOTAL NUMBER OF GRAPHEME-PHONEME PAIRS, PHON(GRAF) IS GRAFOFON PHONOLOGY, AND PHON (METU) IS METUBET PHONOLOGY

Mapping					Type Counts			Token Counts		
From	To	Tot Pair	Mean Pairs	Type–Token	Total Consis (%)	Entropy–Consis	V:C	Total Consis (%	Entropy-Consis	V:C
Graph	Phon (Graf)	32	1.10	0.86	95.4	-0.981	0.441	93.3	-0.938	0.508
1st Graph	Phon (Graf)	31	1.11	0.93	97.4	-0.904	0.289	97.9	-0.860	0.281
Graph	Phon (METU)	38	1.31	0.92	91.7	-0.811	0.508	91.9	-0.751	0.607
1st Graph	Phon (METU)	33	1.18	0.95	99.3	-0.647	0.289	99.3	-0.570	0.281

The type sum of the most frequent phoneme for each grapheme divided by the total number of grapheme-phoneme pairs in the current corpus is 0.973 (Grafofon) and 0.993 (METUbet) for word-onsets and 0.917 (METUbet) and 0.954 (Grafofon) for whole-words. With regards to the token sum, using the same calculation as above, the ratio is 0.979 (Grafofon) and 0.993 (METUbet) for word-onsets and 0.919 (METUbet) and 0.933 (Grafofon) for whole-words. In line with Protopapas and Vlahou (2009), if these ratios can be considered to be single-number estimates of the consistency of grapheme-to-phoneme mapping, Turkish is then considered to be 91.9% (METUbet)/ 93.3% (Grafofon) consistent in the feedforward (reading) direction when whole-words using token frequencies are considered.

2.6.1 Grapheme–Phoneme Consistency and Entropy

The Grapheme-Phoneme mappings, grouped by grapheme, and the proportion of occurrence for each phoneme are shown below Table 3 (Grafofon) and Table 4 (METUbet). The proportion of the most frequent phoneme for each grapheme is displayed first, in a separate column to the left of the smaller proportions following it. Additionally, an index of orthographic consistency was calculated as the overall onset entropy value in which entropy values for all word-initial graphemes weighted by their frequency of occurrence within the current corpus are summed (Type and Token Frequency).

TABLE 3: GRAFOFON: WORD ONSET GRAPHEME-TO-PHONEME MAPPINGS.

				Pair Proportion (%)			
Grapheme	**F (%) Type**	**F (%) Token**	**Phoneme**	**Highest**	**Other**	**Entropy (Type)**	**Entropy (Token)**
a	7.24	8.31	*ɑ*	100%		0	0
b	7.65	8.73	*b*	100%		0	0
c	1.15	0.61	*ʤ*	100%		0	0
ç	3.17	2.18	*ʧ*	100%		0	0
d	7.26	9.56	*d*	100%		0	0

e	3.66	4.80	*e*	100%		0	0
f	1.72	1.72	*f*	100%		0	0
g	5.56	6.93	*ɟ*	89%		-0.004	-0.005
			g		11%	-0.001	-0.001
h	3.51	3.38	*h*	100%		0	0
i	4.06	0.88	*i*	100%		0	0
ı	0.27	0.04	*ɯ*	100%		0	0
j	0.13	0.04	*ʒ*	100%		0	0
k	11.85	10.27	*k*	82%		-0.029	-0.023
			c		18%	-0.010	-0.012
l	0.68	0.49	*l*	57%		-0.001	-0.001
			ɫ		43%	-0.002	-0.001
m	4.38	3.67	*m*	100%		0	0
n	1.10	1.26	*n*	100%		0	0
o	2.35	4.01	*o*	100%		0	0
ö	2.08	1.88	*ø*	100%		0	0
p	2.76	2.22	*p*	100%		0	0
r	1.38	1.05	*r*	100%		0	0
s	8.59	7.47	*s*	100%		0	0
ş	1.37	1.23	*ʃ*	100%		0	0
t	5.96	4.79	*t*	100%		0	0
u	1.94	0.90	*u*	100%		0	0
ü	0.83	1.11	*y*	100%		0	0
v	1.53	2.35	*v*	100%		0	0
y	6.89	8.91	*j*	100%		0	0
z	0.92	1.22	*z*	100%		0	0

Grouped by Grapheme and Sorted by Within-Grapheme Proportions. Each line denotes an individual grapheme-phoneme mapping in the corpus. F, relative frequency (percentage of occurrences of all phoneme types and tokens) in the corpus; pair proportion, percentage of this phoneme–grapheme pair as a proportion of all occurrences (tokens) of this phoneme. The proportion of the dominant mapping is listed first, on the left; other mappings follow, on the right. Entropy values are weighted by type and token frequency.

TABLE 4: METUBET: WORD ONSET GRAPHEME-TO-PHONEME MAPPINGS

GROUPED BY GRAPHEME AND SORTED BY WITHIN-GRAPHEME PROPORTIONS. EACH LINE DENOTES AN INDIVIDUAL GRAPHEME-PHONEME MAPPING IN THE CORPUS. F, RELATIVE FREQUENCY (PERCENTAGE OF OCCURRENCES OF ALL PHONEME TYPES AND TOKENS) IN THE CORPUS; PAIR PROPORTION, PERCENTAGE OF THIS PHONEME–GRAPHEME PAIR AS A PROPORTION OF ALL OCCURRENCES (TOKENS) OF THIS PHONEME. THE PROPORTION OF THE DOMINANT MAPPING IS LISTED FIRST, ON THE LEFT; OTHER MAPPINGS FOLLOW, ON THE RIGHT. ENTROPY VALUES ARE WEIGHTED BY TYPE AND TOKEN FREQUENCY.

				Pair Proportion (%)			
Grapheme	**F (%) Type**	**F (%) Token**	**Phoneme**	**Highest**	**Other**	**Entropy (Type)**	**Entropy (Token)**
a	7.24	8.31	AA	99.75		-0.0003	-0.00005
			A		0.25	-0.000004	-0.0000002
b	7.65	8.73	B	100.00		0	0
c	1.15	0.61	C	100.00		0	0
ç	3.17	2.18	CH	100.00		0	0
d	7.26	9.56	D	100.00		0	0
e	3.66	4.80	EE	92.78		-0.003	-0.004
			E		7.22	-0.001	-0.0008
f	1.72	1.72	F	100.00		0	0
g	5.56	6.93	GG	91.76		-0.006	-0.004
			G		8.24	-0.001	-0.0005
h	3.51	3.38	H	100.00		0	0
i	4.06	0.88	IY	100.00		0	0
ı	0.27	0.04	I	100.00		0	0
j	0.13	0.04	J	100.00		0	0
k	11.85	10.27	KK	82.28		-0.022	-0.02
			K		17.72	-0.009	-0.011
l	0.68	0.49	LL	99.93		-0.00001	-0.000001

			L		0.07	-0.00000004	-0.0000000005
m	4.38	3.67	M	100.00		0	0
n	1.10	1.26	NN	100.00		0	0
o	2.35	4.01	O	100.00		0	0
ö	2.08	1.88	OE	100.00		0	0
p	2.76	2.22	P	100.00		0	0
r	1.38	1.05	RR	100.00		0	0
s	8.59	7.47	S	100.00		0	0
ş	1.37	1.23	SH	100.00		0	0
t	5.96	4.79	T	100.00		0	0
u	1.94	0.90	U	100.00		0	0
ü	0.83	1.11	UE	100.00		0	0
v	1.53	2.35	VV	100.00		0	0
y	6.89	8.91	Y	100.00		0	0
z	0.92	1.22	Z	100.00		0	0

2.6.2 Grapheme–Phoneme Consistency and Entropy – Whole-word

The Grapheme-Phoneme mappings, grouped by grapheme, and the proportion of occurrence for each phoneme are reported in Table 5 (Grafofon) and Table 6 (METUbet). The proportion of the most frequent phoneme for each grapheme is displayed first, in a separate column to the left of the smaller proportions following it.

Table 5: Grafofon: Whole Word Grapheme-to-Phoneme Mappings, Grouped by Grapheme and Sorted by Within-Grapheme Proportions

Table 5: Each line denotes an individual grapheme-phoneme mapping in the corpus. F, relative frequency (percentage of occurrences of all phoneme types and tokens) in the corpus; pair proportion, percentage of this phoneme-grapheme pair as a proportion of all occurrences (tokens) of this phoneme. The proportion of the dominant mapping is listed first, on the left; other mappings follow, on the right. Entropy values are weighted by type and token frequency.

				Pair Proportion (%)			
Grapheme	**F (%) Type**	**F (%) Token**	**Phoneme**	**Highest**	**Other**	**Entropy (Type)**	**Entropy (Token)**
a	7.09	9.75	*ɑ*	100.00		0	0
b	1.77	2.06	*b*	100.00		0	0
c	1.50	1.14	*ʤ*	100.00		0	0
ç	0.91	0.00	*ʧ*	100.00		0	0
d	4.46	5.18	*d*	100.00		0	0
e	7.27	8.56	*e*	100.00		0	0
f	0.57	0.57	*f*	100.00		0	0
g	1.10	1.60	*ɟ*	73.66		-0.003	-0.002
			g		26.34	-0.001	-0.001
h	6.56	7.16	*h*	100.00		0	0
i	5.60	6.45	*i*	100.00		0	0
ı	4.29	0.59	*ɯ*	100.00		0	0
j	0.11	0.09	*ʒ*	100.00		0	0
k	5.04	5.01	*k*	56.73		-0.013	-0.013
			c		43.27	-0.011	-0.011
l	7.84	7.18	*l*	50.77		-0.020	-0.018

			ɫ	49.23	-0.019	-0.018
m	5.31	4.19	*m*	100.00	0	0
n	6.91	6.95	*n*	100.00	0	0
o	2.21	2.80	*o*	100.00	0	0
ö	0.81	1.89	*ø*	100.00	0	0
p	1.54	1.61	*p*	100.00	0	0
r	8.04	7.55	*r*	100.00	0	0
s	3.56	3.25	*s*	100.00	0	0
ş	2.09	1.88	*ʃ*	100.00	0	0
t	4.21	3.76	*t*	100.00	0	0
u	2.00	2.06	*u*	100.00	0	0
ü	1.34	1.61	*y*	100.00	0	0
v	0.81	0.92	*v*	100.00	0	0
y	3.59	3.66	*j*	100.00	0	0
z	2.47	1.71	*z*	100.00	0	0
ğ	1.00	0.82	*ː*	100.00	0	0

Table 6: METUbet: Whole Word Grapheme-to-Phoneme Mappings, Grouped by Grapheme and Sorted by Within-Grapheme Proportions

Each line denotes an individual grapheme-phoneme mapping in the corpus. F, relative frequency (percentage of occurrences of all phoneme types and tokens) in the corpus; pair proportion, percentage of this phoneme-grapheme pair as a proportion of all occurrences (tokens) of this phoneme. The proportion of the dominant mapping is listed first, on the left; other mappings follow, on the right. Entropy values are weighted by type and token frequency.

				Pair Proportion (%)			
Grapheme	**F (%) Type**	**F (%) Token**	**Phoneme**	**Highest**	**Other**	**Entropy (Type)**	**Entropy (Token)**
a	7.09	9.75	AA	75.01		-0.021	-0.017
			A		24.99	-0.011	-0.006
b	1.77	2.06	B	100.00		0	0
c	1.50	1.14	C	100.00		0	0
ç	0.91	0.00	CH	100.00		0	0
d	4.46	5.18	D	100.00		0	0
e	7.27	8.56	EE	58.02		-0.024	-0.028
			E		41.98	-0.020	-0.027
f	0.57	0.57	F	100.00		0	0
g	1.10	1.60	GG	89.10		-0.003	-0.002
			G		10.90	-0.001	-0.0005
h	6.56	7.16	H	100.00		0	0
i	5.60	6.45	IY	100.00		0	0
ı	4.29	0.59	I	100.00		0	0
j	0.11	0.09	J	100.00		0	0
k	5.04	5.01	KK	57.12		-0.012	-0.013
			K		42.88	-0.010	-0.011
l	7.84	7.18	LL	52.94		-0.020	-0.019

			L		47.06	-0.019	-0.017
m	5.31	4.19	M	100.00		0	0
n	6.91	6.95	NN	97.88		-0.002	-0.002
			N		2.12	-0.0002	-0.0001
o	2.21	2.80	O	100.00		0	0
ö	0.81	1.89	OE	100.00		0	0
p	1.54	1.61	P	100.00		0	0
r	8.04	7.55	RR	81.55		-0.015	-0.017
			RH		18.45	-0.006	-0.010
s	3.56	3.25	S	100.00		0	0
ş	2.09	1.88	SH	100.00		0	0
t	4.21	3.76	T	100.00		0	0
u	2.00	2.06	U	100.00		0	0
ü	1.34	1.61	UE	100.00		0	0
v	0.81	0.92	VV	88.11		-0.001	-0.001
			V		11.89	-0.0003	-0.0001
y	3.59	3.66	Y	100.00		0	0
z	2.47	1.71	Z	60.19		-0.011	-0.012
			ZH		39.81	-0.013	-0.014
ğ	1.00	0.82	GH	100.00		0	0

TABLE 7: FREQUENCY WEIGHTED ENTROPY VALUES FOR THE TURKISH ORTHOGRAPHY FOR WHOLE WORDS AND WORD-INITIAL UNITS ONLY

Mapping		**Type Counts**			**Token Counts**		
From	**To**	**Total**	**V**	**C**	**Total**	**V**	**C**
Grapheme	**Phoneme (Grafofon)**	0.068	0.000	0.068	0.085	0.000	0.085
First Grapheme	**Phoneme (Grafofon)**	0.045	0.000	0.045	0.043	0.000	0.043
Grapheme	**Phoneme (METUBet)**	0.189	0.075	0.114	0.196	0.078	0.118
First Grapheme	**Phoneme (METUBet)**	0.044	0.004	0.039	0.041	0.005	0.036

2.6.3 Phoneme to Grapheme Consistency and Entropy

Though the focus of this chapter regards the feedforward (reading) mappings of graphemes and phonemes, it is also important to consider the feedback (spelling) mappings (phoneme to grapheme) in Turkish. Similar to findings in another transparent agglutinative language, namely Hungarian (Borgwaldt, Hellwig, & De Groot, 2004), Turkish manifests with a one-to-one mapping of phonemes to graphemes resulting in a total consistency of 100% and an entropy value of 0. This holds true irrespective of the phoneme inventory used and type/token frequencies. This is in line with previous results as most languages display an asymmetry between the number of letters and phonemes with all alphabetic orthographies possessing a larger phoneme inventory than letters (Borgwaldt, Hellwig, & De Groot, 2004; 2005; Protopapas & Vlahou, 2009).

2.6.4 Measures of Complexity

As previously stated, complexity may be quantified using the ratio of letters to phonemes with simple words having a letter-to-phoneme ratio of one and complex words as having a letter-to phoneme ratio above 1. Using this approach with the current corpus, the proportion of simple words in Turkish using the METUbet phoneme inventory is 100%, and the proportion of simple words in Turkish using the Grafofon phoneme inventory is 99.995%. The nine complex words identified in this corpus were then further investigated and were all found to be extremely long (23 or 24 letters) and possessed <eği> or <iği> in the penultimate syllable. In the instance when <ğ> follows a front vowel in a word-final or syllable-final position, it is pronounced as a palatal glide (Göksel & Kerslake, 2004). This can thus be interpreted as a limitation of the current phoneme inventories available in Turkish and the phonological behaviour of soft ‘g’ warrants further research (see Ünal-Logacev, Fuchs, Żygis, 2014 for a more comprehensive discussion on the topic).

2.7 Further Considerations and Discussion

From the above findings, the Turkish orthography can be characterised as both highly predictable and simple at the grapheme level although deviations from the phonological ideal exist within the feedforward (reading) direction. At the grapheme level, estimates of the predictability of individual mappings range from 57% for </> to 100% for 25 graphemes (Grafofon, Word-Onset). When considering whole-word predictability, estimates range from 50.77% (Grafofon) and 52.94% for </> (METUbet) to 100% for the majority of grapheme to phoneme mappings across both phoneme inventories. In the feedback direction, regardless of the phoneme inventory employed, predictability was 100%.

Using Grafofon: word-onset, a nonparametric comparison of the 28 grapheme consistency estimates with the 31 phoneme consistency estimates found that the asymmetry is non-significant (Mann–Whitney U = 362.5, Z = -1.27, asymptotic two-tailed p = .20054. Using METUbet: word-onset, a nonparametric comparison of the 28 grapheme consistency estimates with the 33 phoneme consistency estimates found that the asymmetry was significant (Mann–Whitney U = 280, Z = -3.26, asymptotic two-tailed p = .00112.

Using Grafofon: whole-word, a nonparametric comparison of the 29 grapheme consistency estimates with the 32 phoneme consistency estimates found that the asymmetry was non-significant (Mann–Whitney U = 377, Z = -1.25, asymptotic two-tailed p = .2113. Using METUbet: whole-word, a nonparametric comparison of the 29 grapheme consistency estimates with the 38 phoneme consistency estimates confirmed that the asymmetry was significant (Mann–Whitney U = 290, Z = -3.30, asymptotic two-tailed p = .00096.

2.7.1 Types vs. Tokens

As mentioned previously, there has been considerable debate regarding the use of types or tokens with regards to frequency measures in psycholinguistic research. The following section will explore the distinction, if any, between the two in the current corpus. Firstly, as can be seen from Table 2, the correlation between the two measures is very high ranging from 0.86 (Grafofon, Whole-word) to 0.95 (METUbet, Word-Onset). A series of paired-samples t-tests were conducted to compare the relative frequency of types and tokens.

For the word-onset measures, there was no difference in the scores for relative type ($M=3.57$, $SD=2.99$) and token ($M=3.57$, $SD=3.25$) frequencies; $t(27) = -.003$, $p=.997$). When frequency-weighted entropy was considered, the trend remains the same irrespective of the Grafofon, $t(30) = -.611$, $p=.546$ and METUbet, $t(32) = -.531$, $p=.599$ phoneme inventories. Similarly, using whole-word measures there was no difference in the scores for relative type ($M=3.45$, $SD=2.50$) and token ($M=3.45$, $SD=2.79$) frequencies; $t(28) = .0$, $p=1.0$). When frequency-weighted entropy was considered, the trend also remains the same irrespective of the Grafofon, $t(31) = -1.68$, $p=.103$ and METUbet, $t(37) = -.612$, $p=.545$ phoneme inventories. Thus, it can be concluded that regardless of mapping strategy or phoneme inventory, no significant differences between type and token frequency counts were found in this corpus. The findings here are also reflected with the whole-word quantification of the Greek orthography which also found no distinction (at least in the feedforward direction) between type and token frequencies. In line with this, Moscoso del Prado Martín, Ernestus, and Baayen (2004) postulate that a shared token-based mechanism can explain both token- and type-based effects. That is not to say that there is not a linguistic difference between the two measures (e.g., Berg, 2014; Bybee, 1995; 2001). An avenue of further exploration regarding resolving these findings would be to reformulate frequency as a three-dimensional construct in which the first level represents the number of affix types, the second level representing the number of words of which the affix types belong to, and the third level representing the textual frequency of the affixed words (Berg, 2016).

2.7.2 GRAFOFON VS. METUBET PHONEME INVENTORIES

The issue of utilising two phoneme inventories in this thesis adds a degree of complexity to the interpretation of the results. A paired-samples t-test was conducted to compare the entropy indices between the different phoneme inventories, taking the mapping strategy into account. At the word-onset level, the was no significant difference between the Grafofon ($M = .0293$, $SD = .0227$) and the METUbet phoneme inventories ($M = .0282$, $SD = .0185$; $t(5) = .577$, $p = .589$. However, when the same analysis is extended to the whole-word level, a statistical difference was found between the Grafofon ($M = .0493$, $SD = .0345$) and the METUbet phoneme inventories ($M = .1283$, $SD = .0528$); $t(5) = -5.518$, $p = .003$. These findings suggest that the METUbet phoneme inventory may be the superior of the two grapheme-phoneme mapping schemes used in this chapter as it captures the most irregularity (based on standard deviation) and thus unpredictability in quantifying orthography transparency.

2.7.3 Word-Onset vs Whole-word

In line with the approach used in the quantification of the Greek orthography (Protopapas & Vlahou, 2009), it is also important to explore the differences, if any, between the use of word-onset and whole-word measures as indicated above. A paired-samples t-test was conducted to compare the total entropy indices between the different mapping strategies taking the phoneme inventories into account. For the Grafofon phoneme inventory, there was a statistically significant difference between the onset ($M = .0293$, $SD = .0227$) and the whole-word measures ($M = .0493$, $SD = .0395$); $t\ (5) = t\ (5) = 2.6$, $p = .048$. When the METUbet phoneme inventory was considered, the trend remained the same as a statistical difference was found between the onset ($M = .0282$, $SD = .0185$) and the whole-word measures ($M = .0128$, $SD = .0528$); $t\ (5) = 6.31$, $p = .001$. The results imply that the restriction of quantification of orthography to word-initial mappings introduces systematic distortions into the model due to differences in the distribution of phonemes and graphemes across word positions. As such, although word-onset entropy measures overcome the limitations of the monosyllabic bias found in other studies (e.g., Martensen, Maris, & Dijkstra, 2000) they still fall short of capturing the true degree of variation between orthography and phonology.

2.7.4 Cross-Linguistic Comparisons

Beyond the generation of well-controlled linguistic stimuli, the real value and utility of calculating a quantitative index of transparency lies in the ability to carry out cross-linguistic comparisons. In the following section, the results reported above will be compared with previous attempts at quantification of transparency. Considering the strengths and weaknesses of different approaches, only directly comparable studies will be contemplated, and as such, studies that use monosyllables or rhyme spelling bodies (Ziegler Perry, & Coltheart, 2000) will not be considered. Additionally, Borgwaldt, Hellwig and de Groot (2005) used lemma forms instead of word forms and will not be considered in a direct comparison.

As stated above, using type frequency, the overall consistency of Turkish is 97.3% (Grafofon)/ 99.3% (METUbet) for word-onsets and 91.7% (METUbet)/ 95.4% (Grafofon) for whole-words in the feedforward direction and 100% in the feedback direction.

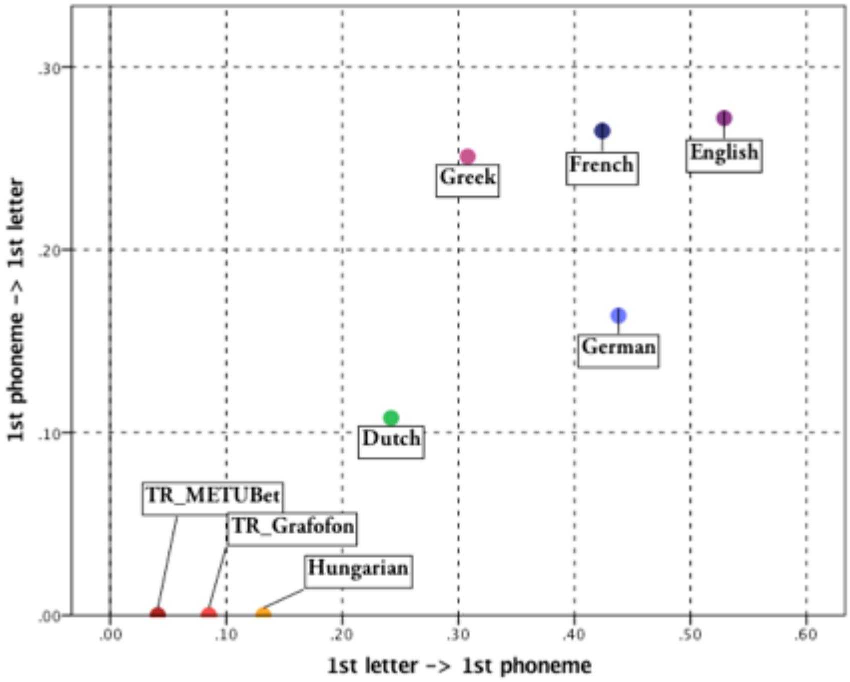

FIGURE 6: DEVIATIONS FROM ONE TO ONE MAPPING BETWEEN WORD-INITIAL LETTERS AND PHONEMES STATED IN ENTROPY VALUES. THE VALUES ON THE X-AXIS DENOTE THE DEGREE OF SPELLING-TO-SOUND AMBIGUITY, WHEREAS THE Y-AXIS DEPICTS THE DEGREE OF SOUND-TO-SPELLING AMBIGUITY

When compared to the entropy values of word-initial letter type counts (Borgwaldt, Helliwig, & de Groot, 2004; Protopapas & Vlahou, 2009), Turkish can be considered to be an example of an exceptionally consistent alphabetic orthography with only small deviations from the alphabetic principle (H = .045) in the feedforward (reading) direction and with no deviation in the feedback (H = 0) direction (see Figure 4). To this end, the Turkish orthography is similar with regards to consistency to Finnish (H = 0) and Hungarian (H = 0.13). Interestingly, the three orthographies mentioned share features such as agglutination and vowel harmony. From Figure 5, it can also be concluded that Turkish is more consistent than all other orthographies investigated using entropy counts.

Finally, in a similar line of investigation to Borgwaldt, Helliwig, and de Groot (2005), the relative contribution of vowels and consonants to orthographic transparency was also explored. Though the Borgwaldt, Helliwig, and de Groot (2005) study used lemmas instead of words, they found that the outcomes of the lemma-based study were comparable to their previous study.

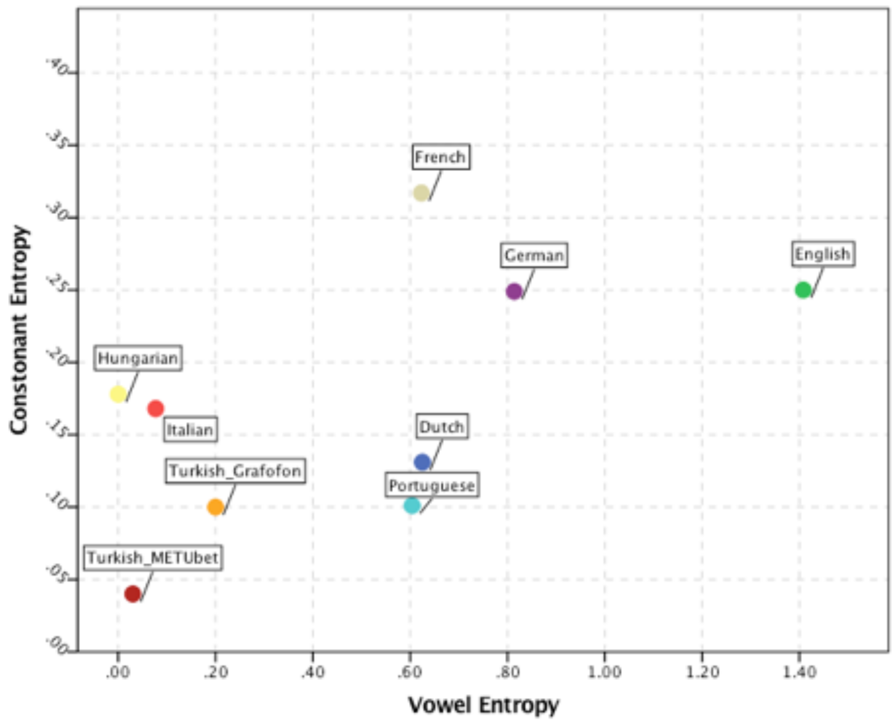

FIGURE 7: DEVIATIONS FROM 1:1 MAPPING BETWEEN WORD-INITIAL LETTERS AND PHONEMES STATED IN ENTROPY VALUES. THE VALUES ON THE X-AXIS SHOW THE DEGREE OF VOWEL AMBIGUITY, WHEREAS THE Y-AXIS DEPICTS THE DEGREE OF CONSONANT AMBIGUITY. NON-TURKISH ENTROPY VALUES TAKEN FROM BORGWALDT, HELLIWIG, & DE GROOT (2004;2005).

At the grapheme-phoneme (whole-word) level, the only meaningful comparison that can be undertaken is with Greek (Protopapas & Vlahou, 2009). At the word-level, consistency was found to be comparable in the feedforward direction: 95.7% in Greek vs 95.4% in Turkish (Grafofon) but varied considerably in the feedback direction: 82.9% in Greek vs 100% in Turkish. Thus, unlike the well-documented (e.g. Porpodas, 2006), asymmetric transparency of Greek, Turkish is an example of an orthography with bidirectional consistency (though this is not absolute). When the comparison is restricted to word-onsets, consistency indices display a difference in both the feedforward (93.8% in Greek vs 97.3% in Turkish) and the feedback direction (93.3% in Greek vs 100% in Turkish).

2.8 CONCLUSION

In conclusion, to the best of the author's knowledge, this chapter provides the first quantitative indices of orthographic transparency for Turkish. In this regard, Turkish, whilst being highly complex in terms of morphology, can be characterised as being as both highly predictable and simple at the grapheme level within both the feedforward (reading) and feedback (spelling) directions. Furthermore, the grapheme-phoneme level appears to be the most appropriate level of analysis for Turkish though further studies are needed to confirm this. Additionally, the data gathered for Turkish has been compared with similar data reported for other orthographies concluding that Turkish is the most transparent orthography that has been quantified to date.

From a methodological perspective, entropy has been presented as a more complete and informative measure of consistency than the percentage of dominant mappings, because entropy measures factor in the relative proportions of non-dominant mappings. In the case of Turkish, it would not be simple to disentangle the two measures as they are highly correlated. Furthermore, in addition to Greek (Protopapas & Vlahou, 2009), the analyses carried out in this chapter provide an interpretation of orthographic transparency using whole-word entropy measures and thus overcomes previous limitations of using unrepresentative samples of the orthography, such as monosyllabic words or word-initial letters.

Theoretically, in line with recent developments in the field (i.e. Protopapas & Vlahou, 2009; Schmalz, Marinus, Coltheart, & Castles, 2015) orthographic depth can further be conceptualised as several distinct constructs namely, degree of complexity and predictability of grapheme to phoneme mappings. However, there is also a need to develop a further framework that extends to non-alphabetic writing systems. For example, Shimron (2006) states that Hebrew depth is very different from English depth. Differences in graphic complexity (Chang, Chen, & Perfetti, 2017), for example, offer challenges for frameworks like the orthographic depth that are based exclusively on alphabets. Given the above considerations and based on the findings reported here, it is evident that the Turkish orthography provides an excellent medium for the further investigation of typical and atypical reading development in highly transparent alphabetic orthographies.

CHAPTER 3

SUBTLEX-TR: THE CREATION AND VALIDATION OF A NEW PSYCHOLINGUISTIC DATABASE FOR TURKISH

3.1 PREFACE

The need to establish reliable psycholinguistic resources for the investigation of reading processes remains a central endeavour in visual word recognition research. This chapter will introduce the topic area covering historical and current developments in the domain, specifically with regard to the recent SUBTLEX movement. Subsequently, the chapter will critically review the presently available resources for psycholinguistic research in Turkish and introduce SUBTLEX-TR, a new Turkish-word database created by taking into account frequency data from film and television subtitles. Furthermore, a new sub-corpus using subtitles for films and television shows that are appropriate for primary-school-aged children, SUBTLEX-TR-CHILD, will be introduced. Finally, SUBTLEX-TR will be validated using a Lexical Decision Task by comparing the respective variance explained by SUBTLEX-TR and TS Corpus (Sezer & Sezer, 2013) frequencies. Although lexical decision data was not available for children, SUBTLEX-TR-CHILD will be validated using the word naming data reported in Chapter 4. The chapter will conclude with a discussion of the findings and potential future developments of the new database.

3.1.1 INTRODUCTION

Visual word recognition tasks such as Single Word/Nonword Naming (SWNN) and Lexical Decision Task (LDT) represent the most widely used approaches to investigate reading (Coltheart et al., 2001). Both tasks require the participant to respond to word (or nonword) stimuli either by reading aloud (SWNN) or by indicating if the stimulus is a word or not (LDT). Thus, the stimuli for such tasks need to be carefully selected by taking into account the plethora of orthographic and phonological properties that influence lexical processing. Psycholinguistic databases enable the selection of stimuli by permitting researchers access to a collection of standardised data and the ability to manipulate and control variables that best suit their experimental parameters. For example, word frequency, the frequency in which a word occurs in a corpus, has systematically and extensively been found to be the most reliable predictor of reaction times in word recognition studies (e.g. Solomon and Postman, 1952; Forster & Chambers, 1973; Taft, 1979; Grainger, 1990). The word frequency effect indicates that high-frequency words are named faster than low-frequency words. Also, word frequency effects have been accounted for by all current computational models of visual word recognition (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996).

The work of Thorndike (1921) represents one of the first significant attempts in constructing a word frequency database by compiling a sizable amount of written texts. The seminal study of Thorndike (1921) resulted in a word frequency database of 10,000 English word types and 4.5 million-word tokens. This study was extended over two decades later by Thorndike and Lorge (1944), resulting in a word frequency database of 30,000 English word types and 18 million word tokens. Other early noteworthy frequency lists generated in different languages during this period include the *French Word Book* (Henmon, 1924; Vander Beke, 1932) and the *Spanish Word Book* (Buchanan, 1927). However, the laborious nature of manually compiling such frequency lists severely restricted the advancement of the field. The rapid development of computer technology, in general, and information processing capacity, specifically, has eased the limitations of traditional manual compilation approaches. A shift towards digitally constructed corpora has greatly benefited the fields of corpus linguistics and computational linguistics.

Kučera and Francis (1967) compiled the earliest electronic corpus, the Brown corpus, which produced the most extensively used word frequency norms in English to date. The Brown corpus consists of approximately 50K word types and 1M word tokens and is representative of written American English. Despite the widespread use of the Brown Corpus for over 40 years, there have been several significant criticisms of the frequency norms derived from the corpus. For example, Burgess and Livesay (1998) compared word frequency estimates from the Brown corpus with the Hyperspace Analogue to Language (HAL) Corpus (1995), a 97K word type and 131M word token corpus, in a word naming task. The study found that while both the Brown and HAL corpora equally predicted high-frequency words, the HAL corpus offered superior predictions for low- and medium-frequency words, i.e., the correlation between RT and

Brown corpus frequencies was low. This finding may be seen as a reflection of the limited size of the Brown corpus (Breland, 1996). More recently, the predictive validity of the Brown Corpus has again been called into question by several megastudies (lexical decision and word-naming tasks) that examined the predictive ability of word-frequency measures (Balota et al., 2004; Brysbaert & New 2009; Zevin & Seidenberg, 2002).

Brysbaert and New (2009) also state that the type of language register from which frequency measures are derived is predictive of linguistic performance of individuals in that corpora derived from everyday language use are superior to corpora generated from formal written communication. The rationale for this is that the use of formal, expressive language and refrainment from using repeated words would result in higher lexical diversity, leading to an overestimation of rare word frequencies and the underestimation of high-frequency words (Baayen 2001; New, Brysbaert, Veronis, & Pallier, 2007). Thus, the issue of language representativeness in corpora, the degree to which the corpus contains a complete range of linguistic samples that represent a language as a whole (Sinclair, 2005), should also be considered to avoid bias in the way word frequency counts are collected. As an extension of this, the number of samples used to create a corpus should also be taken into consideration. To create a lexicon that is representative of a language, Brysbaert and New (2009) proposed that the corpus should contain between 3,000 - 10,000 different texts. Statistically, obtaining word frequency measures from a more extensive corpus should permit for an improved, more accurate measure of word frequency as the standard error of the word counts varies as a function of the square root of the sample size (Lee, 2003). Additionally, a larger corpus would allow for better representation of low-frequency words as well as establishing more subtle differences between them (Burgess & Livesay, 1998). Research from the English Lexicon Project (Balota et al., 2007), the French Lexicon Project (Ferrand et al., 2010), the Dutch Lexicon Project (Keuleers, Diependaele, & Brysbaert, 2010), and the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012) show that virtually all of the known word frequency effect lies within a frequency range under ten occurrences per million words. It is also reported that the most significant word frequency effect is detected for words with a frequency between 0.1 and 1 per million words (Keuleers, Lacey, Rastle, & Brysbaert, 2012).

Thus far, this section has discussed the historical development of corpus linguistics about word frequency measures in visual word recognition tasks. While psycholinguistic resources have increased in availability, some critical issues have been identified in the domain, namely:

(1) the size of corpora in both absolute terms, i.e. the number of word types and tokens in the corpus and in the relative number of texts used to generate the corpus.
(2) Corpus representativeness with recent studies showing that corpora that make use of everyday, informal language being superior to formally written text resources.

To address these issues, New, Brysbaert, Veronis, and Pallier (2007) proposed a new type of psycholinguistic resource derived from film and television subtitles. The rationale for this new approach is that the language used in film and TV subtitles aligns more closely with everyday language, capturing language samples from "real" social situations and human interactions. This approach addresses the need for corpus representativeness identified above. Additionally, subtitles derived from film and television series are readily available from several internet websites (e.g. www.opensubtitles.org). This issue of corpus and resource availability will be considered further in the conclusion of this chapter.

New and colleagues (2007) validated a 52M word token corpus from over 9000 French films and TV series by assessing the extent to which the new subtitle word frequency measure predicted word processing times in comparison to established word frequency measures for French, i.e. the spoken *Corpus du Référence du Français Parlé* [CRFP] (Equipe DELIC, 2004) as well as the written corpus produced by New, Pallier, Brysbaert, and Ferrand (2004). The study found that film and television subtitle frequency measures accounted for about 10% more variance in lexical decision reaction times than the other corpora investigated. This study also signified the beginning of the "SUBTLEX" movement.

As stated previously, Brysbaert and New (2009) presented new frequency norms obtained from SUBTLEX-US, a 51 million word corpus based on American English films and television (TV) series subtitles. The study found that SUBTLEX-US explained a significantly higher proportion of variance in accuracy (10%) and word recognition times (6%) than the Brown corpus. This finding is not unanticipated given the limited size of the Brown corpus. Additionally, when more sizeable and contemporary corpora have been studied, the word frequency norms obtained from film and TV series subtitles show a consistent performance advantage over written-word frequency norm databases. For example, in the same study, Brysbaert and New (2009) also compared the new subtitle frequency norms to the CELEX database (17.9M words; Baayen, Piepenbrock, & van Rijn, 1993, the Hyperspace Analogue to Language (HAL) (more than 130M words; Lund & Burgess, 1996), the written British National Corpus (88M words; Leech, Rayson, & Wilson, 2001) as well as the Zeno corpus (17M words; Zeno, Ivens, Millard, & Duvvuri, 1995). The SUBTLEX-US word frequencies again manifest with a clear advantage over the other corpora investigated especially for short-words (the HAL corpus was superior for long words). Furthermore, Brysbaert, Keuleers and New (2011) compared the SUBTLEX-US word frequency measures with Google's Ngram word frequencies (131 billion-word corpus from digitised American English books (Michel et al. 2011) and found that the SUBTLEX-US corpus explained 11% more variance of lexical decision times than the Google Ngram frequencies. In conclusion, it can be stated that the observed advantage of subtitle-word frequency measures cannot be explained based solely on corpus size. Additionally, it has been observed that variance gains level off after about 30 million words (Brysbaert & New, 2009).

Since the original SUBTLEX database, comparable databases have been developed in several other languages including Chinese (SUBTLEX-CH: Cai & Brysbaert, 2010), Dutch (SUBTLEX-NL: Keuleers et al., 2010), Greek (SUBTLEX-GR: Dimitropoulou, Duñabeitia, Avilés, Corral, & Carreiras, 2010), German (SUBTLEX-DE: Brysbaert et al., 2011), Spanish (SUBTLEX-ESP: Cuetos, Glez-Nosti, Barbon, & Brysbaert, 2011), Albanian (SUBTLEX-AL: Avdyli & Cuetos, 2013), British English (SUBTLEX-UK: van Heuven, Mandera, Keuleers, & Brysbaert, 2014), Polish (SUBTLEX-PL: Mandera, Keuleers, Wodniecka, & Brysbaert, 2015) and Portuguese (SUBTLEX- PT: Soares et al., 2015). Recently, Gimenes and New, (2016) introduced Worldlex, a word frequency database built on Twitter, blog posts and newspapers for 66 languages and used a regression analysis to compare the new measures against five existing megastudies that have been conducted in English (Balota et al., 2007), French (Ferrand et al., 2010), Dutch (Keuleers, Diependaele, & Brysbaert, 2010), Malay (Yap, Liow, Jalil, & Faizal, 2010), and Chinese (Sze, Rickard, Liow, & Yap, 2014). The authors found that Worldlex frequencies predict lexical decision reaction times similar to, or better than, the frequencies that have been used to date, providing additional support for corpora that incorporate everyday spoken language use. Furthermore, the study discovered that in the five languages investigated, the blog and Twitter frequencies were more important than the newspaper frequencies. Although the Worldlex database offers a rich and significant resource for psycholinguistic research, further investigation is needed in other languages covered in the Worldlex database to validate the initial findings empirically.

In sum, along with the original SUBTLEX-US database, all subtitle word-frequency measures tested to date along with databases constructed from social media (Herdağdelen & Marelli, 2017; Gimenes & New, 2016), and blogs (Gimenes & New, 2016) have demonstrated superiority over written-word frequency norms, proposing that the linguistic style found in digital media is highly representative of the linguistic experience of young adults, principally with reference to the university student populations traditionally recruited in psycholinguistic studies. However, there is also evidence that for older adults, traditional word frequency measures based on books may be more suitable (Brysbaert & Ellis, 2016). Hence, there is a necessity to create and maintain corpora that offer more than one-word frequency measure, as Johns and colleagues (2016) suggest. That is, potentially further gains may be achieved by assembling word frequency lists that have been personalised to the participants of a study, depending on their learning histories, including reading habits.

Taking this into consideration, it is also essential to provide a brief overview of the current reading habits of Turkish-speaking children and young adults. A survey carried out by the Ministry of National Education in 1993, found that 61% of children and young adults did not read any books over the last month, whereas 13% of the population had read only one book. Furthermore, consistent with the above finding, the United Nations Human Development Index ranks Turkey as 76th in book reading amongst the 173 nations studied (Human Development Report, 2008). Relatedly, the overall level of newspaper and magazine reading has also been reported to be low (Özturk, Sevim and Eroğlu, 2006). In addition, data from the Progress in International Reading Literacy Study (PIRLS) which focuses on the reading achievement of fourth-grade primary school students found that Turkey's average score (449) was significantly lower than the international average (500) (PIRLS, 2001). Amongst university students in Turkey, there is an increasing number of reports that state the decreasing number of regular readers (e.g. Pehlivan, Serin, & Serin, 2010; Odabaş, Odabaş, & Polat, 2008). Ultimately, the transition from printed to electronic mediums regarding reading also have to be considered. For instance, Gökçearslan and Seferoglu (2016) reported that, between 2010 and 2015, the age of first internet access in Turkey fell from 5 to 2 years old on average. Also, the same study reported an increase in almost all internet activities carried out by children. Further, one of the most significant changes noted was a 45.3% to 81.4% increase in online movie and music streaming. Also, the use of social networks has seen an increase from 51.7% to 81.3% among children in Turkey. This shift towards alternate leisure activities and the increasing availability of visual media may be able to explain, at least in part, the continuous decline in reading books, magazines and newspapers. Finally, considering the above, the extraction of word frequencies from visual media such as subtitles of film and TV series appears to be an extremely valuable substitute for written-word frequencies. However, despite their relevance and availability for a growing number of languages, word frequency norms from film and TV series subtitles are still non-existent for Turkish.

As stated previously, the characteristics of Turkish make it a fascinating language for the ongoing investigation of language representation and processing. Consequently, having reliable frequency norms available for Turkish such as the ones reported in this chapter will provide an essential resource for the future development of monolingual and cross-linguistic studies that take advantage of the characteristics of the Turkish language. This study will complement and expand upon the small number of lexical databases already available in Turkish. The following section will provide a critical evaluation of currently existing psycholinguistic resources for Turkish.

3.2 OVERVIEW OF PSYCHOLINGUISTIC RESOURCES IN TURKISH

For Turkish, resources for psycholinguistic variables exist but are scarce. Conceivably the most important cause for this is the required investment of time, labour and resources necessary to develop such a resource (Cangöz, 1999). Cangöz (1999) further specified that one of the most critical and fundamental deficiencies regarding cognitive psychology studies in Turkish-speaking populations is a comprehensive word frequency resource. The fact that the rate of psychology research in Turkey is increasing also stresses the importance of creating such a resource.

Historically, Turkish word lists compiled from newspapers, books and magazines were subjected to ratings by highly literate native speakers, and after inter-rater reliability was established, a selection was used as experimental stimuli (I. Raman, 1996; 1999; 2006). Most recently, using similar methods lead to the first colour picture norms in Turkish, which also include frequency ratings (Raman, Raman, E. & Mertan, 2014). Although subjective norms can be deemed problematic, their use has nevertheless been argued in the literature to be more indicative of the dynamics of a particular language than objective frequency counts (e.g. Gernsbacher, 1984; Gordon, 1985).

With regards to frequency measures in Turkish, Göz (2003) produced a word frequency dictionary based on a 22,693-word type and 1 million- word token corpus and may represent the most substantial attempt at developing a contemporary word frequency dictionary in Turkish to date. Inspired in part by the American-English Brown corpus (Kucera and Francis, 1967), the categories of the corpus were as follows: press (35%), novel-story (20%), science (8%), popular science (9%), fine arts and biography hobby (4%), religion (3%), school book (3%) and other (10%). Thus, it can be claimed that the *Written Word Frequency Turkish Dictionary* is representative of a general domain dictionary of Turkish. However, the same criticisms that have been levied at the Brown Corpus can also be applied to the *Written Word Frequency Turkish Dictionary* (Göz, 2003). Briefly, the size of the dictionary is small in both absolute and relative terms. Additionally, the corpus fails to meet the recommended criteria that to sufficiently represent the Turkish language, a corpus should contain at least 30 million word tokens (Brysbaert & New, 2009).

The recent development of the Turkish National Corpus (TNC; Aksan et al., 2012) represents a move away from traditional corpus development approaches in Turkish. The TNC is designed to be a balanced, large-scale and general-purpose corpus for contemporary Turkish. Though not explicitly designed as a psycholinguistic database, the TNC can be used to generate word frequency (raw and parts per million) as well as contextual diversity measures for Turkish words. The TNC was constructed from 4978 documents and contains 50 million-word tokens. The main criticisms of the TNC can be roughly separated into two distinct arguments. The first is that the number of documents is relatively small meaning that there may be an overreliance on words generated from a few large sources. As stated above, this can lead to an underrepresentation of rare words and an overrepresentation of common words. More importantly, the relatively low document count would also impact on contextual diversity measures. The second is an issue of accessibility in that although the corpus is publicised as being open for research use, it does so with several restrictions. The corpus only allows up to 400 queries per day per account. This restriction thus poses severe time restrictions on extracting statistical information.

Moving on from the TNC, the TS Corpus v2 (Sezer & Sezer, 2013; http://tscorpus.com), extensively used in Chapter 2, is a large general-purpose Turkish Corpus containing 491 million POS-Tagged tokens and 4.9 million unique word forms that build on and extends the BOUN Corpus (Sak, Güngör, & Saraçlar, 2007). The BOUN Corpus was created by collecting web pages from three Turkish daily newspapers (212M tokens) as well as a general sampling of Turkish webpages (279M tokens). Furthermore, the BOUN corpus is readily accessible and is sufficiently large, making it highly appropriate for consideration in a comparison study with the new subtitle frequencies.

Feasibly the most ambitious attempt at developing a Turkish psycholinguistic database to date comes in the form of KelimetriK (Erten, Bozşahin, & Zeyrek, 2014). KelimetriK is a query-based software that provides information on word frequency, bigram and trigram frequency, orthographic neighbourhood and similarity statistics. KelimetriK can, therefore, be viewed as a valuable resource for psycholinguistic experimenters with which several lexical and sublexical variables can be controlled or manipulated depending on the research question. The N-watch software for English

(Davis, 2005), and BuscaPalabras for Spanish (Davis and Perea, 2005) represent equivalents in their respective languages. Similar to other datasets reported here, Kelimetrik also suffers from several limitations. One such issue is that Kelimetrik is limited to a stem list which does not allow querying homographic words, i.e., words with the same orthographic representation but with a distinct meaning. Additionally, the use of lemma frequencies as opposed to word frequencies would diminish the lexical diversity attributed to agglutinative orthographies. As described in Chapter 2, this issue remains contentious in the literature, although the use of word frequency measures would arguably capture more variance in such morphologically rich languages. Finally, Kelimetrik is restricted to a single word query at a time. The size of the corpus, extracted from Erten, Bozsahin, & Zeyrek, (2014) consists of only 24,414 Turkish stem words, although the word frequency measure is derived from the suitably large BOUN corpus (Sak, Güngör, & Saraçlar, 2007). In line with the current investigation, Kılıç (2008), while examining the role of vowel harmony in Turkish, compiled a list of words obtained from publicly available Turkish subtitles (Dave, 2011). More specifically, the corpus contains a word list that includes all words observed from Turkish subtitles, along with the frequency of the word's occurrence in the subtitles.

While the entirety of this section has focussed on adult corpora and databases, there is also a need to consider psycholinguistic resources designed for use with children. Until recently, Turkish psycholinguistics researchers have relied on word stimuli that they have created themselves. For example, to generate words for a one-minute word reading list, Babayiğit and Stainthrop (2007) state that due to the lack of frequency norms, they generated their own by analysing books of primary and secondary grades. The apparent lack of reproducibility by taking such an approach as well as the inefficiency of generating new stimuli for every new experiment reaffirms the need to create a widely available psycholinguistic database for use in both Turkish-speaking children and adults.

Recently, Acar, Zeyrek, Kurfali and Bozşahin (2016) produced a Child Literature Corpus (CLC) created from 535 books written for 3-12 years old Turkish-speaking children. The CLC is composed of 19,246 word types and 4,388,149 word tokens. Additionally, a 300 word subset of the CLC was used in the production of AoA and imageability measures for Turkish. While representative of a move in the right direction, the corpus is still considerably small and is subject to copyright restrictions on its distribution.

In summary, the above section highlights that although there have been essential developments with regard to the creation of psycholinguistic databases for Turkish, there is still a need to move beyond current approaches. This notion, coupled with the observation of changing reading behaviours, calls for the exploration of new avenues of corpora selection. In line with the growing SUBTLEX movement, the remainder of this chapter will focus on the development and validation of a subtitle word frequency database for Turkish considering a number of lexical and sublexical features.

3.3 The construction of SUBTLEX-TR

In this section, the creation of SUBTLEX-TR will be introduced and discussed. SUBTLEX-TR is a new word frequency measure for 924,824 Turkish words forms obtained from a 156,761,118-word corpus based on 49,220 Turkish films (14,132), and TV series (35,088) subtitles screened between 1990 and 2016. In Turkey, national film and TV production are enjoying a revival *(*Basutçu, 2008)*,* although foreign films still maintain a considerable degree of popularity and are often subtitled (as opposed to being dubbed). Therefore, amassing a subtitle corpus for Turkish is a relatively straightforward task to accomplish and the database produced in this chapter will aim to become a valuable research tool for the Turkish scientific community who utilise verbal stimuli in their experiments, especially for those who work with word reaction time data.

Overall, 158,810 files containing film and television subtitles identified as Turkish were downloaded (See Lison & Tiedemann, 2016; http://www.opensubtitles.org/) and processed using the following procedure.

i. Firstly, duplicates and corrupted files (6) were removed, resulting in 54,979 film and TV subtitle files.
ii. Following this, film and TV subtitles that were representative of pre-1990 films were filtered out of the final list resulting in 49,220 files. This step was taken to

ensure a modern language register of Turkish was being created and that words that had fallen out of contemporary use were not included in the final database.

iii. Finally, all subtitle-specific text formatting was stripped before further processing.

iv. To create the children's SUBTLEX-TR subcorpus, the additional step of filtering films labelled as "U" (Universal) and "PG" (Parental Guidance) resulted in 13,034 film and TV subtitles. The resulting subcorpus, SUBTLEX-TR-child contains 505,024 word types and 27,193,916-word tokens.

3.3.1 Word and Syllable Length

Word length effects, i.e., longer words take longer to respond to, can be considered a hallmark feature of visual word recognition processes (see Balota et al., 2004) and are thought to be reflective of sublexical processing. Furthermore, word length measures can be derived from both orthographic features (number of letters) and phonological features (number of phonemes and syllables). There is, however, some evidence of null effects of word length (e.g. Weekes, 1997) as well as several studies reporting a "U" shaped curve (e.g., New, Ferrand, Pallier, & Brysbaert, 2006) which suggests that RTs are longer for short and long words than for words that are between 5 to 8 letters. It is posited that the reverse, i.e. inhibitory length effect is partly driven by the notion that long words have fewer competitors and are, therefore, easier to identify. For Turkish, word length effects have been reported in both typically developing children (See Chapter 4) and adults (Kokten & Raman, 2007) as well as nonword repetition in Turkish-speaking children with Specific language impairment (SLI) (Topbaş, Kaçar-Kütükçü, & Kopkalli-Yavuz, 2014). Recent studies also indicate that syllabic length also contributes to visual word processing (Davies, Barbon & Cuetos, 2013). This effect has also been reported for Turkish (Öney, Peter & Katz, 1997) though see Kokten and Raman (2007) for a null finding regarding the contribution of syllabic length effects. The agglutinative nature of Turkish morphology would stipulate that word, and syllable lengths are highly correlated in Turkish. The correlation found in the current database was $r(924824) = 0.75$, $p < 0.001$. For the current database, the number of letters was calculated by summing the number of letters for each word and number of syllables was calculated using a modified version

(Appendix 1) of a syllabification algorithm for Turkish (Altınok, 2016). The frequency distribution of the number of letters and syllables can be seen in Figure 6 and Figure 7, respectively.

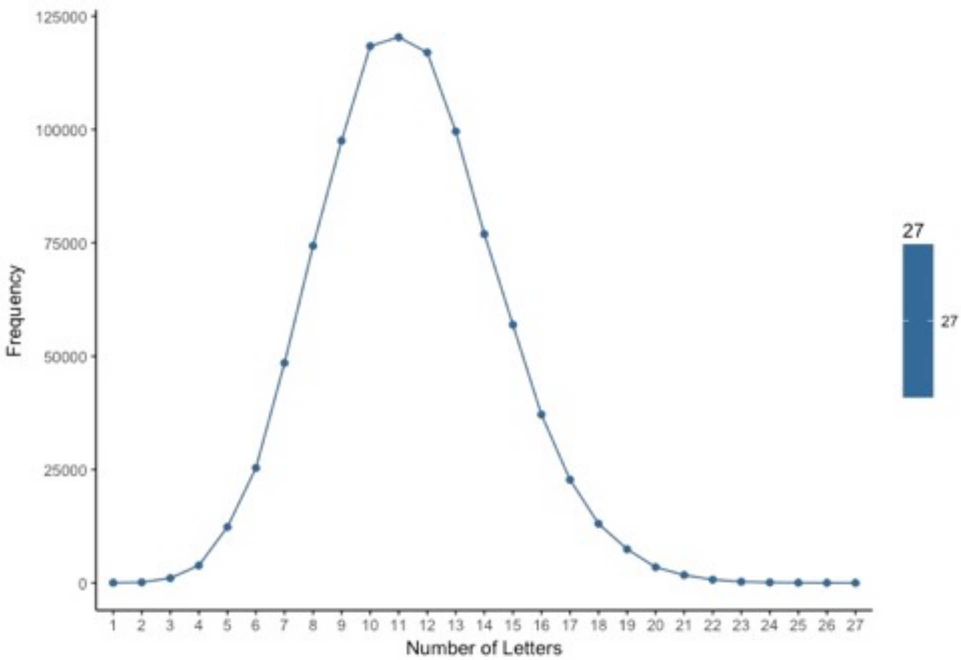

FIGURE 8: THE FREQUENCY OF WORDS BY LETTER COUNT

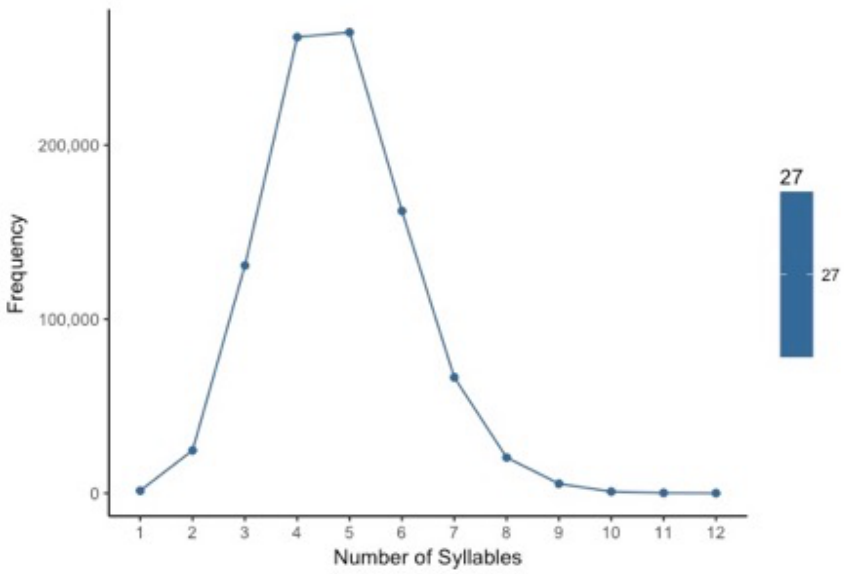

FIGURE 9: THE FREQUENCY OF WORDS BY SYLLABLE COUNT

3.3.2 WORD-FORM FREQUENCY

Similar to the finding of length effects, word frequency effects, i.e., more frequent words are named faster than less frequent words have previously been reported for

Turkish-speaking adults (Raman, Baluch & Sneddon, 1996; Raman, Baluch, & Besner, 2004). As stated previously, film and TV subtitle derived word frequency estimates are exceptionally useful at predicting human performance in behavioural tasks. In addition to raw frequency counts, it is also important to calculate standardised measures as word frequency is highly sensitive to the size of the corpus and thus corpora of varying size can be compared. The most commonly used standardised measure to date has been frequency per million (fpm) words. However, regardless of its widespread use, there are several issues with the fpm measure. For instance, in large corpora, the majority of words have a frequency of less than one fpm. Indeed, Brysbaert and New (2009) reported that for the SUBTLEX-US corpus, 76% of the 50M words were found to have a frequency of less than the intuitive starting point of 1 fpm. In the current database, which represents a sizeable 156M word corpus, the results are even more pronounced as nearly 95% of words fall below one fpm. In addition, it is currently understood that the frequency effect is compressed in that a logarithmic curve best represents it. In terms of corpora, this means that the difference in frequency between one fpm and two fpm has roughly "*the same effect on processing times as the difference between 10 fpm and 20 fpm, between 100 fpm and 200 fpm, and between 1000 fpm and 2000 fpm*" (Brysbaert & New, 2009). A recently proposed alternative will be discussed in the section below.

3.3.3 Zipf Scale

The Zipf scale (van Heuven, Mandera, Keuleers, and Brysbaert, 2014) is a recently proposed logarithmic scale and is calculated as log10 (frequency per billion words). The scale scores frequency from 1 (1 per 100 million words) to 6 (1000 per million words) with the lower half of the scale (1-3) representing low-frequency words and the upper half (4-6), high-frequency words. A particularly interesting property of the Zipf scale is that it permits unobserved words, i.e. frequency of 0 to be assigned a value by Laplace smoothing (Brysbaert and Diependaele, 2013). The addition of this transformation facilitates the comparison of corpora from different registers and with a substantial variation of word type and token counts such as the different corpora used to validate the current database. In addition to the raw frequency, fpm and the Zipf scale frequencies, this chapter also provides a frequency measure of LOG10 and LOG10[2].

3.3.4 Letter unigram and bigram frequencies

Sublexical statistical properties of words such as unigram and bigram frequencies are also important variables for consideration in psycholinguistic research. For instance, letter n-gram statistics present with uneven distribution (see Zipf, 1950) and have been shown to influence visual word recognition (see Seidenberg, 1987; Massaro & Cohen, 1994) though see Carreiras, Álvarez, & Devega (1993) for an alternative explanation of these effects. The current study found 724 letter bigrams and 20095 letter trigrams. Unigram, bigram and trigram frequency were calculated using the n-Gram (3.0.3) package in R and then summing and averaging all bi/trigram occurrences for each word. The final bigram and trigram frequencies can, therefore, be considered to be both token values as well as being position-independent.

3.3.5 Contextual Diversity

Contextual Diversity can be conceptualised as the number of documents in which a word appears and is thought to contribute to the stability of conceptual representations (Burgess & Livesay, 1998). Recently, Adelman, Brown, and Quesada (2006) stipulated that contextual diversity might be more significant than word frequency measures in explaining the variance of lexical decision/ word naming latency data. Using the number of documents (or in this case the number of films and tv subtitles) as a proxy measure of contextual diversity (CD), it is hypothesised that words of equal frequency would be differentially processed if their respective CD scores were different, i.e., words appearing in higher contexts would be processed faster. Convergently, Brysbaert and New (2009) observed that CD accounts for 1 %–3 % more variance than does word frequency. Additionally, the influence of CD has been shown in studies of word learning (e.g. Hills, Maouene, Riordan, & Smith, 2010; Perea, Soares, & Comesaña, 2013) as well as spoken word recognition (Johns, Gruenenfelder, Pisoni, & Jones, 2012) though this effect was mediated by semantic distinctiveness. To the best of the author's knowledge, no study to date has accounted for CD in Turkish visual word recognition. By creating an index of CD in this chapter, future studies in Turkish psycholinguistics will be able to explore this exciting avenue of inquiry. For the current, study CD measures were calculated using the

DocumentTermMatrix function of the tm (0.7) package in R (3.4.1) for the subtitle files of both the children and adult SUBTLEX-TR corpora. The database provides raw CD counts as well as the percentage of documents containing each word.

3.3.6 Neighbourhood Statistics

The orthographic neighbourhood (ON) of a word can be conceptualised as the number of orthographically similar words that can be created with one transformation (addition, deletion or substitution) of a given word while preserving letter order (Coltheart, Davelaar, Jonasson & Besner, 1977). The traditional measure of ON referred to in the literature as ColtheartsN, has been reported to provide evidence of a facilitative effect of ON on lexical access, though the majority of these findings appear to be restricted to low-frequency words (Andrews, 1997; Perea & Rosa, 2000). Furthermore, a limitation fo ColtheartsN is that N is calculated with words of the same length as the target word and hence longer words will always manifest with a smaller neighbourhood size. In an attempt to overcome this restriction, the measure of Orthographic Levenshtein distance 20 (OLD20) has recently been introduced (Yarkoni, Balota, & Yap, 2008). OLD20 is calculated by averaging the 20 closest words in the unit of Levenshtein distance (LD). LD is a measure that specifies the minimum number of insertions, deletions or substitutions required for converting one string to another (Levenshtein, 1966). Generally, it has been observed that words with lower OLD20 values are recognised faster than words with higher OLD20 values (Yarkoni, Balota, & Yap, 2008). However, OLD20 was found to have an inhibitory effect on a Turkish lexical decision task rather than a facilitatory effect (Erten, Bozşahin, & Zeyrek, 2014). This seemingly contradictory finding may be associated with the agglutinative morphology of Turkish in which unique suffix groups are numerous thus providing competition rather than facilitation during lexical access.

It is thus evident that any measure of ON will benefit from a large and diverse corpus. Therefore, it is imperative to provide an index of ON for Turkish using an extensive database like the one reported here. It was decided to report OLD20 scores for the words in the SUBTLEX-TR corpus for two reasons. Firstly, OLD20 is more flexible in its accommodation of longer words (which are numerous in Turkish). Second, at the time of writing this chapter and algorithm used to calculate ColtheartsN had been running continuously for over three weeks without success. OLD20 was calculated using a modified version (http://crr.ugent.be/averageLD/) of the original OLD20 measure in python3.

3.4 Validation of SUBTLEX-TR

As lexical decision reaction times are highly sensitive to word frequency measures (Balota et al., 2004), the task is currently accepted as the standardised approach to validating new word frequency measures by correlating performance in this task with word frequency estimates. In the following section, the design and implementation of a lexical decision validation task will be reported.

Method

3.4.1 Participants

Seventy-two students from the Eastern Mediterranean University in Famagusta, Cyprus participated in the experiment (37 females, 35 males; mean age= 24.03, SD = 2.82) in exchange for course credit. All students were right-handed and had normal or corrected to normal vision. Ethics approval was gained from both Eastern Mediterranean University and Brunel University London (see Appendix 2).

3.4.2 Materials

In order to compare word frequency measures generated from two or more corpora, words for which the corpora give highly divergent estimates have been recently used and have been demonstrated to be a highly efficient approach in increasing the statistical power of the lexical decision task (Mandera, Keuleers, Wodniecka & Brysbaert, 2014).

With this in mind, to make the experiment maximally informative, stimuli were selected for which the TS Corpus and SUBTLEX-TR gave highly divergent frequency estimates. In order to achieve this, linear regression was carried out on the SUBTLEX-TR Zipf frequencies, using the TS Corpus frequencies as the predictor variable. Following this, words were ordered according to their residual error, and 160 words from both extremes of the resulting lists for each corpus were selected for the experiment (320 words in total). Words at one extreme (with a substantial positive residual error value) were much more frequent in SUBTLEX-TR than would be expected on the basis of the TS Corpus, while words at the other extreme (with a substantial negative residual error value) occurred much less often in SUBTLEX-TR than would be expected on the basis of TS Corpus. In line with a recent validation lexical decision study, a further 80 words were selected at random for the experiment. Four hundred nonwords were generated using the Turkish plugin of the Wuggy application, therefore, bringing the total amount of stimuli to 800. A summary of the properties of the words and nonwords selected for this study can be found in Table 8 below.

TABLE 8: WORD/ NONWORD CHARACTERISTICS FOR LEXICAL DECISION TASK

	Words **n = 370**	**Nonwords** **n = 389**	**P**
Length	7.10 (2.57)	7.01 (2.49)	0.64
Syllable	3.03 (1.09)	2.96 (1.05)	0.36
OLD20	1.56 (.299)	1.98 (1.56)	.001*

In order to explore potential bias in the selected stimuli, the ldknn algorithm of the vwr package in R was used. The ldknn algorithm uses both k nearest neighbour classification and the Levenshtein distance metric to calculate the probability of a word response for the given stimulus based on the relative frequency of words among the nearest neighbours (see Keuleers & Brysbaert, 2011). The output of the algorithm can be seen in Figure 8 below. A logistic regression discovered that there was no significant bias for words in the current stimuli lists ($z = -0.86$, $p = 0.39$).

The mean and standard deviation (SD) in word frequency (Zipf scale) mean was 2.68 (SD = 1.16) for TS Corpus and 3.08 (SD = 1.99) for SUBTLEX-TR. The two variances were significantly different, $F(370, 370) = 169.61$, $p < .001$, and Welsch's t-test has shown significant differences in the mean frequency derived from the two corpora, $t(592) = 3.27$, $p < .001$, for this set of stimuli. With regard to the randomly selected word stimuli, mean was 2.73 (SD = 1.16) for TS Corpus and 2.79 (SD = 1.09) for SUBTLEX-TR. The difference between variances was not statistically significant, F(76, 76) = 0.84, p = .36, and the mean frequencies were not significantly different according to Welsch's t-test, t(150) = .35, p = .73.

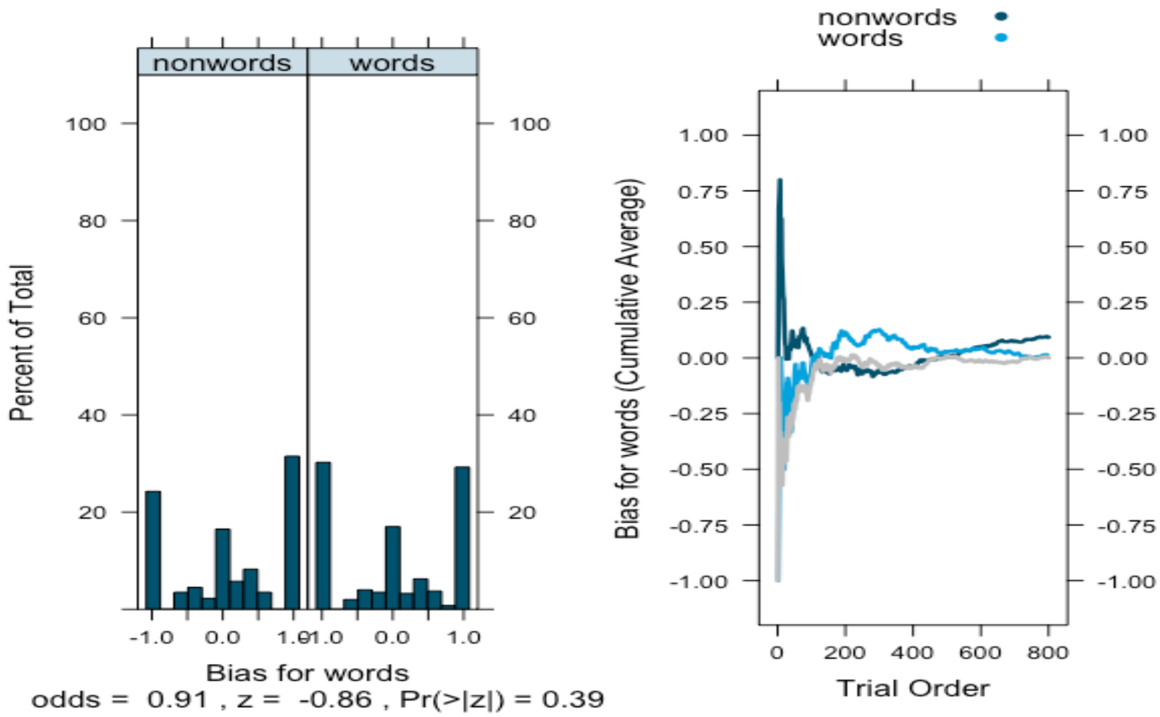

FIGURE 10: OUTPUT OF THE LDKNN ALGORITHM HIGHLIGHTING THE PROBABILITY OF A WORD RESPONSE FOR THE GIVEN STIMULUS BASED ON THE RELATIVE FREQUENCY OF WORDS AMONG THE NEAREST NEIGHBOURS

3.4.3 Procedure

Stimuli presentation and response recording were controlled by DMDX 5.1 software (Forster & Forster, 2003). Participants were requested to decide, as fast and accurately as possible, if the string of letters presented at the centre of the screen was a real word in Turkish or not. If participants considered that the letters string was a real word in Turkish, they were instructed to press the "Z" key on the keyboard ("evet" [yes] response). Equally, if they considered that the presented letter string was not a real word in Turkish, they were instructed to press the "M" key on the keyboard ("hayır" [no] response). Both speed and accuracy were stressed in the instructions.

The task comprises responses to 800 trials which were divided into eight blocks comprised of 50 words and 50 nonwords per block. First, a fixation point (+) was presented at the centre of the computer screen for 500ms. Following this, the fixation point was replaced by the stimulus (word or nonword) at the centre of the computer screen and disappeared when participants responded or until 2,500ms had passed. The order of the stimuli was randomised per block and participant. Participants were informed that several pauses (7) would occur during the experiment (each block; every 100 trials) to counteract fatigue. Before the 800 experimental trials, participants received 12 practice trials (six words and six nonwords). Each experimental session lasted approximately 45 minutes.

3.4.4 RESULTS

Trials that were under 250ms (15) or over 2500ms (388) were considered as outliers and removed from the dataset. In line with a recent SUBTLEX validation study in Polish (Mandera et al., 2014), reaction time (RT) trials that were determined to be outside of a range of whiskers of a boxplot adjusted for skewed distributions (See Hubert & Vandervieran, 2008) were also removed from the dataset. This data cleaning approach was calculated independently for each participant, in each block separately for words and nonwords. Thirty (30) words and 11 nonwords with less than one-third correct responses were then removed from the final dataset.

For the full set of 370 words, the mean RT was 798.60 (SD = 238.64), and the mean accuracy was .92 (i.e., 92%) (SD = .30). Words occurring less frequently in SUBTLEX-TR than in TS Corpus had a mean RT of 936.74 (SD=227.20) and a mean accuracy of .88 (SD = .33), while words occurring more often in SUBTLEX-TR than in TS Corpus had a mean RT of 663.85 (SD = 169.70) and a mean accuracy of .91 (SD = .28). The randomly selected words had a mean RT of 826.58 (SD = 227.82) and a mean accuracy of .91 (SD = .28). For nonwords, the mean RT was 958.20 (SD = 70.23), and the mean accuracy was .87 (SD = .09).

Before comparing the different frequency measures, four separate multiple regression analyses on accuracy and RT data were conducted to evaluate the role of various variables on lexical decision data in Turkish-speaking adults. First, the word/nonword were considered together. Tables 9 and 10, below, highlight the findings of multiple regression analysis for accuracy and RT data of the complete stimuli set, respectively.

TABLE 9: REGRESSION COEFFICIENTS OF LEXICAL DECISION TASK ACCURACY: WORDS AND NONWORDS

Accuracy			
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	10.84	10.47 – 11.24	**<0.001**
Lexicality	0.94	0.91 – 0.98	**0.001**
Length	1.47	1.42 – 1.53	**<0.001**
OLD20	0.73	0.67 – 0.80	**<0.001**
Lexicality * Length	1.19	1.15 – 1.24	**<0.001**
Lexicality * OLD20	0.67	0.61 – 0.74	**<0.001**
Length * OLD20	0.95	0.91 – 0.99	**0.009**

From Table 9, the estimated coefficients for the final model showed that lexical decision accuracy for Turkish words/nonwords was predicted by main effects of lexicality, length, and neighbourhood size (OLD20). In addition, there were significant interactions between lexicality and length, indicating stronger length effects for nonwords than words. Additionally, there was a significant lexicality by neighbourhood size interaction effect indicating stronger differential neighbourhood effects on words (inhibitory) than nonword (facilitory). Finally, there was a significant interaction between length and neighbourhood size indicating stronger neighbourhood effects on short words than longer words (inhibitory effect). No other cognitive predictors or interactions reached significance for inclusion into the final model.

TABLE 10: REGRESSION COEFFICIENTS OF LEXICAL DECISION TASK RT: WORDS AND NONWORDS

	Reaction Time (ms)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	791.36	787.96 – 794.77	**<0.001**
Lexicality	-31.95	-35.31 – -28.58	**<0.001**
Length	40.25	36.99 – 43.50	**<0.001**
OLD20	22.92	13.86 – 31.99	**<0.001**
Lexicality * Length	-3.98	-7.18 – -0.78	**0.015**
Lexicality * OLD20	29.62	20.29 – 38.96	**<0.001**

From Table 10, the estimated coefficients for the final model showed that lexical decision reaction times for Turkish words/nonwords was predicted by main effects of lexicality, length, and neighbourhood size (OLD20). In addition, there were significant interactions between lexicality and length, indicating stronger length effects for nonwords than words. Additionally, there was a significant lexicality by neighbourhood size interaction effect indicating stronger neighbourhood effects on words than nonword. Finally, there was a significant interaction between length and neighbourhood size indicating a facilitatory effect of orthographic neighborhood size for nonwords and an inhibitory neighborhood size effect for words. No other cognitive predictors or interactions reached significance. Similiarities and differences between the Turkish words/nonwords accuracy and RT models feature a large degree of overlap in their findings in terms of both main and interaction effects. The single exception to this is the significant finding of a length by neighbourhood size interaction in the accuracy data which was absent in the RT data.

Following this, two further regression analyses were carried out only considering the word data. The frequency measure of choice in this analysis was the SUBTLEX-TR. Tables 11 and 12, below, highlight the findings of multiple regression analysis for accuracy and reaction time data of the word stimuli, respectively.

TABLE 11: REGRESSION COEFFICIENTS OF LEXICAL DECISION TASK ACCURACY: WORDS

	Accuracy		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	14.51	13.43 – 15.70	**<0.001**
Length	2.03	1.87 – 2.21	**<0.001**
Frequency	2.09	1.91 – 2.28	**<0.001**
Contextual Diversity	0.94	0.82 – 1.07	0.336
OLD20	2.08	1.65 – 2.61	**<0.001**
Length* Frequency	1.06	0.97 – 1.17	0.188
Length * Contextual Diversity	0.73	0.65 – 0.83	**<0.001**
Length * OLD20	1.29	0.99 – 1.67	0.059

From Table 11, the estimated coefficients for the final model showed that lexical decision accuracy for Turkish words was predicted by main effects of frequency, length, and neighbourhood size (OLD20) but not contextual diversity. In addition, there were significant interactions between frequency and length, indicating stronger length effects for low frequency words than high frequency words. Additionally, there was a significant length by contextual diversity interaction effect indicating stronger contextual diversity effects on shorter words than longer words. No other cognitive predictors or interactions reached significance though the interaction between length and neighbourhood size approached significance.

TABLE 12: REGRESSION COEFFICIENTS OF LEXICAL DECISION TASK RT: WORDS

	Reaction Time (ms)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	729.22	716.60 – 741.84	**<0.001**
Length	21.18	16.75 – 25.62	**<0.001**
Frequency	-37.06	-48.99 – -25.13	**<0.001**
Contextual Diversity	-51.41	-77.19 – -25.63	**<0.001**
OLD20	-55.20	-73.33 – -37.08	**<0.001**
Frequency * Contextual Diversity	17.52	3.25 – 31.79	**0.016**

From Table 12, the estimated coefficients for the final model showed that lexical decision reaction times for Turkish words was predicted by main effects of frequency, length, contextual diversity and neighbourhood size (OLD20). In addition, there was a significant interaction between frequency and contextual diversity indicating stronger contextual diversity effects on lower frequency words than higher frequency words. No other cognitive predictors or interactions reached significance. Similiarities and differences between the Turkish word accuracy and RT models feature a smaller degree of overlap in their findings when compared to the word/ nonword comparison. The main effects of length, frequency and OLD20 were significant in both models. Differences were found in terms of the influence of contextual diversity which was significant in the RT model but not the accuracy model. All of the discovered significant two-way interactions were distinct between the two models.

Multiple regression analyses on accuracy and reaction time data were conducted to compare the proportion of variance accounted by the Turkish subtitle-word frequency measures (SUBTLEX-TR) with that accounted for by the written-word frequency provided by the TS Corpus frequency measures. Zipf and Zipf^2 frequencies were considered as predictors from the TS Corpus and the SUBTLEX-TR databases. For SUBTLEX-TR, LOG10 and LOG10^2 from the CD measures were also provided.

The percentage of the variance of RT and accuracy explained (Adjusted R^2) by each of the word frequency corpora as well as CD are summarised in Table 13 below.

TABLE 13: PERCENTAGES OF VARIANCE ACCOUNTED FOR BY THE VARIOUS FREQUENCY MEASURES

Model	RT (% all words)	Accuracy (%; all words)	RT (% sampled words)	Accuracy (% sampled words)
Length + OLD20 + WF(TS) + WF(TS2)	19.6	9.6	13.8	20.8
Length+ OLD20 + WF(SUB-TR) + WF(SUB-TR2)	**44.03**	**31.4**	22	**22.3**
Length + CD SUB TR + CDSUB TR2	43.4	28.1	**22.4**	22.2
Length + WF(SUM) + WF(SUM2)	34.9	1.4	21	0.3

When all 370 words were included in the analysis, the TS Corpus word frequencies explained 19.6 % of the variance in RTs and 9.6 % of the variance in accuracy. Additionally, SUBTLEX-TR frequencies explained 44.03 % of the variance in RTs and 31.4% in accuracy, which is 24.43 % more for RTs and 21.8 % more for accuracy when compared with TS Corpus frequencies. The Vuong test for nonnested models was used to test for statistical difference between models, (Vuong, 1989). The differences in performance of the two models were statistically significant for both RTs ($z = -52.72$. $p < .001$) and accuracy ($z = -5.51$, $p < .001$). When only the 77 words that were randomly sampled from the corpus were included in the analysis, the frequencies derived from the TS Corpus explained 13.8 % of the variance in RTs and

20.8% in accuracy. Word frequencies derived from the SUBTLEX-TR corpus explained 22% of the variance for RTs and 22.3 % of the variance for accuracy. The difference was significant for RTs ($z = -11.1$, $p < .001$) but not for accuracy ($z = 1.6$, $p = .11$).

For the full set of words, CD measures calculated on the basis of SUBTLEX-TR accounted for 43.4% and 28.1% of the RT and accuracy variance, respectively. The difference between the SUBTLEX-TR CD and word frequency measures was significant for RTs ($z = -23.89$, $p <.001$) and accuracy ($z = -5.28$, $p <.001$). When only randomly selected words were included in the analysis, CD explained 22.4% of the variance for RTs and 22.2 % for accuracy. This was not significantly different than the model based on subtitle word frequencies for RTs ($z = -.84$, $p =.399$) or for accuracy ($z = .30$, $p = .76$)

3.4.5 Error Analysis

To further explore the influence of lexical and sublexical word properties on reading accuracy, an additional analysis of error rates was carried out. All 759 stimuli (370 words, 389 nonwords) were reentered into the analysis. In order to satisfy conditions of normality, error scores were log-transformed after the addition of a constant of 0.01 per participant per condition and then were subjected to a one-way ANOVA. There was no significant difference between the error scores for the lexicality condition ($F<1$). However, for words, length effects were statistically significant $F(10,781) = 40.66$, $p <0.001$, $\eta^2 = 0.342$. To evaluate the nature of the difference observed, a post-hoc Bonferroni multiple comparisons were carried out. The results suggest that error rate for the 5, 6 and 7 letter conditions was significantly higher than shorter (2,3,4) and longer words (10,11,12). In a final consideration for the word data, frequency effects on error rate were examined. The one-way ANOVA revealed a significant effect of frequency on error rate $F(6,497) = 50.15$, $p<0.001$, $\eta 2 = 0.377$. To evaluate the nature of the difference observed, post-hoc Bonferroni multiple comparisons were carried out. The results suggest that the error rate for low-frequency words was significantly higher than for high-frequency words.

TABLE 14:P-VALUES OF POST-HOC MULTIPLE COMPARISONS OF LENGTH BY ERROR RATE (BONFERRONI ADJUSTED)

	2	3	4	5	6	7	8	9	10	11
3	**0.004**	-	-	-	-	-	-	-	-	-
4	**0.009**	**<0.001**	-	-	-	-	-	-	-	-
5	**<0.001**	**<0.001**	**0.009**	-	-	-	-	-	-	-
6	**<0.001**	**<0.001**	**0.045**	1.000	-	-	-	-	-	-
7	**<0.001**	**<0.001**	1.000	0.230	0.821	-	-	-	-	-
8	0.240	**<0.001**	1.000	**<0.001**	**0.001**	1.000	-	-	-	-
9	1.000	**<0.001**	1.000	**<0.001**	**<0.001**	0.807	1.000	-	-	-
10	0.058	1.000	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	-	-
11	**<0.001**	1.000	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	1.0	-
12	**0.017**	1.000	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	1.0	1.0

Bold indicates significance at 0.05

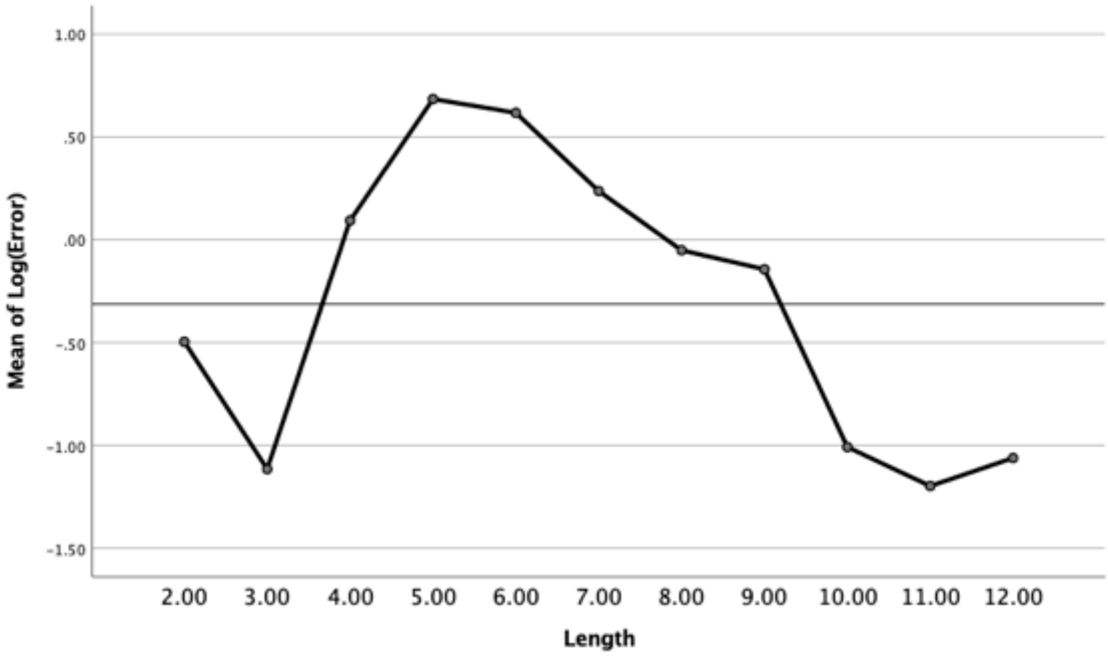

FIGURE 11: ERROR RATE (LOG TRANSFORMED) BY WORD LENGTH. HIGHER VALUES INDICATE HIGHER ERROR RATES. BLACK LINE DENOTES MEAN ACROSS CONDITIONS.

TABLE 15: P-VALUES OF POST-HOC MULTIPLE COMPARISONS OF FREQUENCY BY ERROR RATE (BONFERRONI)

	1	2	3	4	5	6
2	**<0.001**	-	-	-	-	-
3	**<0.001**	1.00	-	-	-	-
4	**<0.001**	1.00	1.00	-	-	-
5	**<0.001**	**<0.001**	**<0.001**	**<0.001**	-	-
6	**<0.001**	**<0.001**	**<0.001**	**<0.001**	1.00	-
7	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

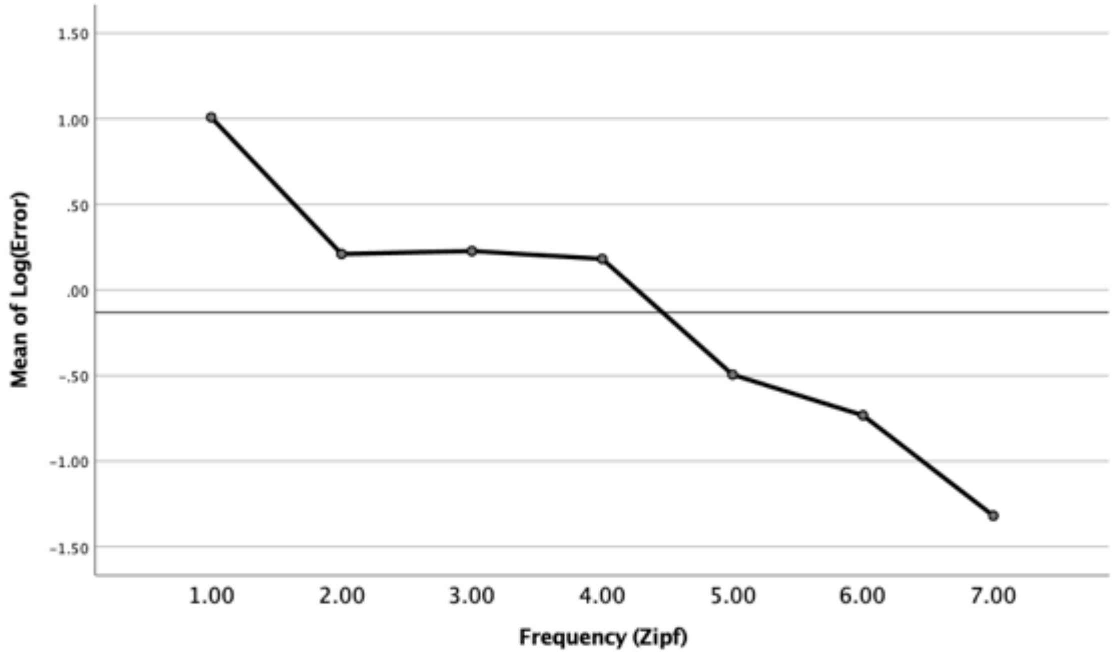

FIGURE 12: ERROR RATE (LOG TRANSFORMED) BY WORD FREQUENCY. HIGHER VALUES INDICATE HIGHER ERROR RATES. BLACK LINE DENOTES MEAN ACROSS CONDITIONS.

An identical analysis of nonword error rates was also carried out. For nonwords, length effects were statistically significant $F(10,781) = 40.66$, $p < 0.001$ $\eta^2 = 0.225$. To evaluate the nature of the difference observed, a post-hoc Bonferroni multiple comparisons were carried out. The results suggest that error rate for the 5, 6 and 7 letter conditions was significantly higher than shorter (2,3,4) and longer words (10,11,12).

TABLE 16: : P-VALUES OF POST-HOC MULTIPLE COMPARISONS OF NONWORD LENGTH BY ERROR RATE (BONFERRONI)

	2	3	4	5	6	7	8	9	10	11
3	**<0.001**	-	-	-	-	-	-	-	-	-
4	**<0.001**	1.00	-	-	-	-	-	-	-	-
5	**<0.001**	0.26	1.00	-	-	-	-	-	-	-
6	**<0.001**	1.00	1.00	0.46	-	-	-	-	-	-
7	**<0.001**	1.00	1.00	0.13	1.00	-	-	-	-	-
8	**<0.001**	1.00	1.00	**0.02**	1.00	1.00	-	-	-	-
9	**<0.001**	1.00	0.10	**<0.001**	1.00	1.00	1.00	-	-	-
10	**<0.001**	1.00	0.65	**<0.001**	1.00	1.00	1.00	1.00	-	-
11	0.11	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.05**	**<0.001**	-
12	0.22	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.02**	**<0.001**	**<0.001**

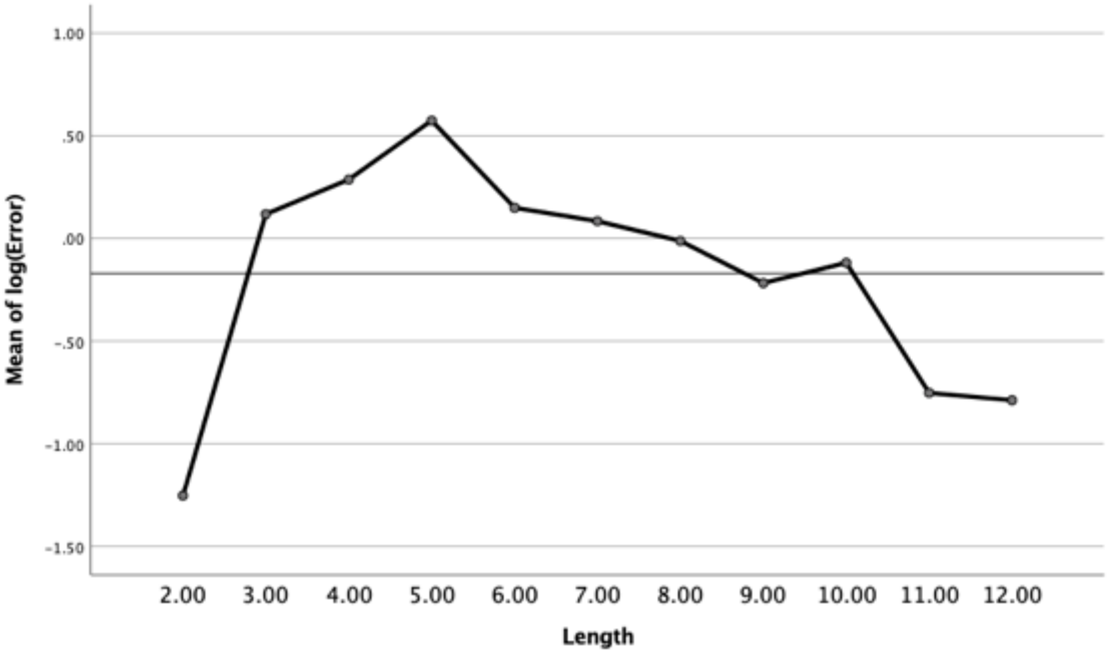

FIGURE 13: ERROR RATE (LOG TRANSFORMED) BY NONWORD LENGTH. HIGHER VALUES INDICATE HIGHER ERROR RATES. BLACK LINE DENOTES MEAN ACROSS CONDITIONS.

3.4.6 Discussion

The current validation study found a distinct advantage for SUBTLEX-TR word frequencies over word frequencies derived from the TS Corpus. The observed differences in the captured variance were greater when divergent frequency estimates were incorporated into the analysis. With reference to the RT data, the advantage of the SUBTLEX-TR corpus remained statistically significant even when only the random sampled words were used. The results further suggest that the SUBTLEX-TR derived word frequencies were the more balanced of the frequency measures in that the lexical decision RTs are more in line with the predictions from SUBTLEX-TR compared with the TS corpus. Additionally, the TS Corpus frequencies appear to underestimate RTs for words that have a much lower occurrence in SUBTLEX-TR. This finding has been proposed to be indicative that the TS Corpus has inflated frequency estimates for these words (Brysbaert & New, 2009). Given the transparent nature of the Turkish orthography, the word frequency effects reported above are rather remarkable as transparent orthographies are considered to be less sensitive to word frequency effects (Cuetos, Glez-Nosti, Barbon, & Brysbaert, 2012).

In contrast to all other studies of word frequency databases derived from subtitles, the current study found that CD measures accounted for less variance than SUBTLEX-TR word frequencies though this difference disappeared when only the randomly selected words were considered. The high degree of correlation between the two measures makes interpretation of this finding difficult though it does not diminish the importance of both word frequency and contextual diversity in word reading studies.

The analysis of error rate yielded some interesting findings. Whilst a lexicality effect was absent, there was a significant effect of length on both words and nonword error rates. In addition, a significant effect of frequency was found for words. The pattern of error rates for both word and nonword length conditions were similar in that letter strings with 4, 5 and 6 letters produced a higher number of errors than for the other length conditions. Interestingly, very long letter strings had significantly lower error rates than medium length letter strings suggesting a well-developed visual recognition system to accommodate for the agglutinative nature of Turkish orthography.

3.5 Validation of SUBTLEX-TR-Child

The creation and validation of a new normative children's database for use in Turkish represents an important and much-needed direction in the literature. Even though lexical decision data were not collected for the following validation study, the single word naming latency data reported in Chapter 4 was used to explore the potential validity and use of SUBTLEX-TR-Child. It is widely reported in the literature that single word naming studies typically account for less variance than lexical decision tasks. However, as this was a first of its kind exploration of new methods for the creation of a lexical database for use with Turkish-speaking children, it was decided to proceed with a similar analysis approach to the above adult data with caution in drawing conclusions.

3.5.1 Participants

130 primary school children's reaction time and accuracy data from the study reported in Chapter 4 was used for subcorpus validation. The mean age (in months) was 120.02 (SD = 15.18), and there were 71 females and 59 males. All children were first language

users of Turkish. Ethics approval was gained from both Eastern Mediterranean University and Brunel University London (see Appendix 3). A more in-depth overview of student demographics and recruitment strategy is offered in Chapter 4.

3.5.2 Materials

The stimuli for Turkish test words were selected from Raman, Raman and Mertan (2014) and Turkish pseudowords were generated using the Turkish plugin (Erten, Bozsahin, & Zeyrek, 2014) of Wuggy, a multilingual pseudoword generator. The 40-word stimulus list generated for real words were used as a template for generating the pseudowords. The parameters in Wuggy were set so that each real word generated ten candidate pseudowords that were matched for length and length of subsyllabic segments. Of the ten candidate pseudowords, the one that manifested the quantitatively smallest deviation from the reference word was selected for the pseudoword stimulus list. A breakdown of the stimuli used in this study can be seen below in Table 17.

TABLE 17: WORD/ NONWORD CHARACTERISTICS OF CHILDREN'S DATA

	Length	Frequency	AoA	Familiarity
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Short Low Frequency	3.30 (0.67)	14.84 (9.26)	2.25 (0.32)	4.66 (0.11)
Short High Frequency	2.70 (0.67)	144.75 (109.36)	1.80 (0.31)	4.74 (0.17)
Short Psuedoword	2.95 (0.68)	-	-	-
Long Low Frequency	6.00 (0)	10.65 (3.93)	2.42 (0.42)	4.68 (0.22)
Long High Frequency	5.90 (0.39)	58.22 (51.83)	2.04 (0.39)	4.76 (0.12)
Long Pseudoword	5.95 (0.22)	-	-	-

3.5.3 PROCEDURE

For the single word reading task, children were presented with a series of 40 words and 40 pseudowords one at a time in a randomized order. Each word appeared in the centre of a computer screen for 3000 ms with an inter-stimulus interval (ISI) of 1500 ms. The order of presentation was randomized for each child who was instructed to read aloud the words/ pseudowords as quickly as possible. A block of practice trials with five words and five pseudowords was presented for naming before the main experiment so that the children could familiarize themselves with the task, generally, and with the notion on pseudowords, specifically. Stimuli presentation and response recording were controlled by DMDX software (Forster & Forster, 2003).

3.5.4 RESULTS

Trials that were under 250ms (17) or over 3000ms (134) were considered as outliers and removed from the dataset. For the 40 words, the mean RT was 1007.94 (SD = 345.3), and the mean accuracy was .97 (SD = .05). For nonwords, the mean RT was 1234.11 (SD = 432.39), and the mean accuracy was .87 (SD = .15). In order to evaluate the internal consistency of RT and accuracy data, split-half correlations for 100 random splits of the data across participants was calculated. The resulting correlations were corrected with the Spearman-Brown prediction formula (Brown, 1910; Spearman,1910), giving average corrected reliability of .97 for RTs and .98 for accuracy.

Multiple regression analyses on accuracy and reaction time data were conducted to compare the proportion of variance accounted by the Turkish subtitle-word frequency measures (SUBTLEX-TR-child) with that accounted for by the written-word frequency provided by the CLC frequency measures. Zipf and Zipf^2 frequencies were considered as predictors from the CLC and the SUBTLEX-TR databases. For SUBTLEX-TR, LOG10 and LOG10^2 from the CD measures were also provided. The percentage of the variance of RT and accuracy explained (Adjusted R^2) by each of the word frequency corpora as well as CD are summarised in Table 18 below.

TABLE 18: PERCENTAGES OF VARIANCE ACCOUNTED FOR BY THE VARIOUS FREQUENCY MEASURES

Model	RT (% all words)	Accuracy (%; all words)
Length + WF(CLC) + WF(CLC^2)	2.7	25.2
Length+ WF(SUB-TR) + WF(SUB-TR2)	**2.8**	**26.7**
Length + CD (SUB TR) + CD (SUB TR^2)	2.7	22.6

When all 40 words were included in the analysis, the CLC word frequencies explained 2.7 % of the variance in RTs and 25.2 % of the variance in accuracy. Additionally, SUBTLEX-TR-child frequencies explained 2.8 % of the variance in RTs and 26.7 % in accuracy, which is 0.1 % more for RTs and 1.5 % more for accuracy when compared with CLC frequencies. In line with the approach adopted in the adult corpus comparison, the Vuong test for nonnested models was used to test for statistical difference between models, (Vuong, 1989). The differences in performance of the two models were not statistically significant for both RTs ($z = 0.34$ $p = .74$) and accuracy ($z = 0.28$, $p = .78$).

For the full set of words, CD measures calculated on the basis of SUBTLEX-TR-child accounted for 2.7% and 22.6% of the RT and accuracy variance, respectively. The difference between the SUBTLEX-TR-child CD and word frequency measures was nonsignificant for RTs ($z = 0.70$, $p = .48$) and accuracy ($z = 0.65$, $p = .51$).

3.5.5 DISCUSSION

In the current chapter, Turkish-speaking adult participants completed a Lexical Decision task in which they had to decide whether the displayed letter string was a Turkish word or not. The study found main effects of lexicality, length and neighbourhood size effects when looking at words and nonwords together. In addition, when only words were considered there were significant main effects of frequency, length, contextual diversity and neighbourhood size. In addition, a number of significant interaction effects were discovered and will be further discussed below.

The lexicality effect is considered to be a marker of lexical reading (Pagliuca et al., 2008) indicating that real words were recognised significantly faster than nonwords and has been previously been reported to be a reliable effect in word naming in Turkish (Raman, 2003). The findings of the current study contribute to previous findings in Turkish and underline the availability of the lexical route for reading in Turkish.

Along with word length, word frequency measures appear to be the most important variables for the investigation of lexical decision tasks in Turkish. Word length effects are particularly interesting for Turkish, given the agglutinative nature of the writing system where extremely long words are possible and frequent (Goksel & Kerslake, 2005). There appears to be an unspoken concencus amonst Turkish psycholinguistics that word length is an important variable but frequently is often used as a control variable rather than a variable of interest (e.g. Bilgin, 2016; Raman, 1999). The word length effect in Turkish has previously been reported to have a small yet significant relationship with name agreement (Raman et al., 2014). In addition, small scale Masters studies have found word length effects on fixation duration in sentence reading (Eren, 2014) and eye movement control (Bozkurt, 2017). Therefore, the present study provides the most comprehensive evidence of a word length effect in Turkish psycholinguistic research and the first concerning lexical decision tasks. The presence of a word length effect is taken as evidence of the serial nature of the sublexical route in reading . Further to this, there was evidence of a length by lexicality interaction on lexical decision accuracy data. The smaller effect of length on words than on nonwords implies that the parallel letter processes used for whole-word reading are different from those used for sublexical and that both lexical and sublexical procedures are available to Turkish adults (Weekes, 1997).

Concerning neighbourhood size, as measured by OLD20, the current study found that RTs for higher OLD20 words were shorter than lower OLD20 words. It can be thus determined that OLD20 had an inhibitory effect on Turkish lexical decision tasks (Erten et al., 2014). This is perhaps reflective of the language's rich agglutinating morphology which results in the presence of large families of words that share the same stem. Other interesting interactions were also observed that highlight the possibly mediation of lexicality and length effects by neighbourhood size in particular. In terms of the significant lexicality by neighbourhood size effect found in the current study, stronger differential neighbourhood effects on words (inhibitory) than nonword (facilitory) were found. This finding lends support to the position that challenges lexical accounts for orthographic neighborhood size effects (e.g. Fiebach et al., 2007). The significant length by neighbourhood size interaction lends further support to the above finding in that short letter strings with higher neighbourhood sizes (both words and nonwords) were recognised with lower accuracy than other conditions.

The significant word frequency effect, found in this chapter, has previously been reported in a number of studies in Turkish-speaking adults (Raman, 1999; 2003). Word frequency effects are also thought to be reflective of lexical processes and therefore adds further support to the availability of the lexical route for processing in Turkish. This finding, along with similar results in other trnsparent orthographies such as Spanish (Davies et al., 2013) and Italian (Burani et al., 2007) adds additional support to the universal hypothesis indicating that a lexical route is used, even considering the transparent nature of these orthographies. Furthmore, the significant interaction of frequency and contextual diversity was a particularly interesting and unique finding. The finding suggests that lower frequency words are particularly sensitive to the effects of contextual diversity in comparison to higher frequency words and may go some way to explain some of the null effect findings of contextual diversity in the analysis that followed.

The current validation study found a distinct advantage for SUBTLEX-TR word frequencies over word frequencies derived from the TS Corpus. The observed differences in captured variance were larger when stimuli with extremely divergent frequency estimates were incorporated into the analysis. With reference to the RT data, the advantage of the SUBTLEX-TR corpus remained statistically significant even when only the random sampled words were used. The results further suggest that the SUBTLEX-TR word frequencies are more balanced than the TS Corpus word frequencies: RTs for the three different groups of stimuli are in line with the predictions from SUBTLEX-TR. On the other hand, the TS Corpus frequencies seem to systematically underestimate RTs for words that have a much lower occurrence in SUBTLEX-TR. This could indicate that the TS Corpus has inflated frequency estimates for these words, of which most could be characterized as belonging to a very formal register. Additionally, given the transparent nature of the Turkish orthography, the word frequency effects reported above is rather remarkable as transparent orthographies are less sensitive to word frequency effects (Cuetos et al. 2012).

In contrast to all other studies of word frequency databases derived from subtitles, the current study found that CD measures accounted for less variance than SUBTLEX-TR word frequencies though this difference disappeared when only the randomly selected words were considered. The high degree of correlation between the two measures makes interpretation of this finding difficult though it does not diminish the importance of both word frequency and contextual diversity in word reading studies.

When considering the children's data, there was no advantage of the SUBTLEX-TR-child word frequencies over CLC word frequencies. This null finding also extends to children's CD measures. The variance accounted for by the accuracy data is similar between children and adults though, regarding RTs, the children's naming data accounted for significantly less variance than the adult lexical decision data. Possible interpretations of the above findings and future directions to extend the current investigations outlined in this chapter are considered below.

3.6 Conclusion and Future Directions

This chapter has introduced and validated SUBTLEX-TR, a new word frequency database for Turkish based on film and television subtitles. The lexical decision experiment carried out in this chapter has validated the usefulness of the new frequency measures by comparing them with estimates derived from TS Corpus. There was a large advantage of SUBTLEX-TR over TS Corpus when words for which estimates given by the two corpora differed most were used as stimuli. In contrast, when words were sampled randomly, the advantage became less pronounced though remained statistically significant.

The findings of this chapter, along with the growing body of research into word frequency effects on visual word recognition, highlight the complex relationship between frequency measures derived from distinct corpora and human performance on psycholinguistic tasks such as lexical decision and reading aloud. The complex nature of the word frequency effect, therefore, raises methodological issues concerning stimuli selection as even mega study approaches can introduce bias in selecting words (e.g., Keuleers et al., 2010). For example, considering the negative correlation between word frequency and word length, any word length decision would influence word frequency measures. Further studies that carry out analyses across different sets of stimuli and for different languages are needed to explore these challenges fully. Lexical decision megastudies (Balota et al., 2007; Keuleers et al.,2010; Keuleers et al.,2011) provide a useful platform for such analyses to take place. Considering the current study, even with validation using a limited set of words, the results of the experiment suggest that both SUBTLEX-TR and TS Corpus are valuable sources of word frequency estimates. The SUBTLEX-TR corpus represents the first widely available subtitle derived word database for Turkish. The database provides frequency and contextual diversity measures based on Turkish language subtitles. It is anticipated that the SUBTLEX-TR corpus will be a valuable resource in future psycholinguistic investigation in Turkish. Later iterations of the SUBTLEX-TR will contain measures of Parts-of-Speech, CV type, lemma frequencies and initial phoneme and further validation will take place in the form of a lexical decision megastudy. New releases of the opensubtitles.org subtitle data are already underway and provide the opportunity to update the current database. Additionally, Worldlex (Gimenes & New, 2016), previously mentioned in this chapter, comprises a subcorpus for Turkish containing 63.7M tokens and thus provides another excellent candidate for comparison with the new subtitle database. Another highly relevant aspect of the Worldlex database is that it also reports contextual diversity measures for each of its subcorpora which can also be used in future validation studies.

The non-significant findings of the SUBTLEX-TR-child corpus in comparison to the CLC database, raises several important methodological issues regarding validation studies in children. Firstly, the small number of words used for this sub-investigation were not selected for their highly divergent estimates of word frequency and as such were highly correlated ($r=0.69$, $p<.0001$). In addition, the low variance captured for RT across the corpora suggests that naming tasks, particularly regarding children, are less informative than lexical decision tasks for use in validation studies. This finding is in line with previous frequency measure validation studies (Cuetos et al., 2012). Furthermore, these methodological issues may be compounded by the reduced sensitivity to frequency effects in transparent orthographies. With these limitations considered, the small non-significant findings of a SUBTLEX-TR-child advantage over the CLC word frequencies warrants further exploration taking into consideration the methodological issues stated above.

CHAPTER 4: THE DEVELOPMENT OF READING IN A HIGHLY TRANSPARENT ORTHOGRAPHY

4.1 INTRODUCTION

This section will offer an overview of the existent literature on reading development, and there is some overlap with previous literature reviews. However, it is important to state that this is a necessary reintroduction rather than mere repetition. Visual word recognition is considered to be one of the fundamental skills involved in reading and according to dual-route theories of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001; Perry, Ziegler, & Zorzi, 2014), there are two available routes to reading: lexical (orthographic), or sublexical (phonological decoding). The lexical procedure is typically assessed by performance on reading high-frequency, irregular words that cannot be read using the sublexical procedure due to the presence of irregular grapheme-phoneme correspondences (GPC). Conversely, the sublexical procedure is primarily assessed by reading novel words, typically pseudowords that can only be decoded using GPC because they are absent from the mental lexicon. If the sublexical procedure is the dominant route, then effects of regularity (faster and/or more accurate processing of regular words than irregular words) and length (faster and/or more accurate processing of short letter-strings than long letter-strings) should be present. Alternatively, if the lexical procedure is the dominant route, then frequency (faster and/or more accurate processing of high-frequency words than low-frequency words) and lexicality (faster and/or more accurate processing of words than pseudowords) effects should be present. In the latter scenario, the effect of length should be less marked for words, and there should be no significant effect of regularity, at least for frequent words.

For over a century, the question of how children learn to read remains a central endeavour in psychological research (e.g. Huey, 1900). As a consequence and as mentioned previously, several theoretical stage models of reading acquisition have been proposed that outline the stages of cognitive development in which learners are suggested to transition through in order to acquire the skill of reading (Ehri, 1991; Frith, 1985; Gough & Hillinger, 1980; Marsh, Friedman, Welch, & Desberg, 1981). As highlighted in the introduction, one of the major limitations of stage/phase models is that although the models are highly informative, they ultimately serve as a theoretical framework rather than as a set of falsifiable scientific hypotheses (Beech, 2005).

Apart from a small number of learned words, it is presently accepted that during the initial stages of learning to read, children primarily make use of the sublexical procedure before a gradual transition toward the lexical procedure. Research that supports the above finding has been reported in English (Backman, Bruck, Hebert,& Seidenberg, 1984; Waters, Seidenberg, & Bruck, 1984), German (Wimmer & Hummer, 1990; Rau, Moeller, Landerl, 2014), French (Leybaert & Content, 1995; Sprenger-Charolles, Siegel, Béchennec, & Serniclaes, 2003; Sprenger-Charolles, Siegel, & Bonnet, 1998), Italian (Zoccolotti, de Luca, di Filippo, Judica, & Martelli, 2009; Zoccolotti et al., 2005) Previous studies have shown that this period is considerably short in transparent orthographies such as Turkish, German, Italian, Spanish and Greek, where learning the regularities of the orthography is comparatively straightforward due to the unambiguous GPC mappings (Avdyli, Castejón, & Cuetos, 2014; Landerl, Wimmer, & Frith, 1997; Öney & Durgunoğlu, 1997). These findings are largely in line with the Psycholinguistic Grain Size Theory (PGST; Ziegler and Goswami, 2005) in that the use of small grain sizes (i.e. phonemes) appears to be the salient feature in these languages. However, there is also a conflicting body of evidence that suggests that this is not necessarily the case. For example, Sebastian-Gallés and Parreño- Vacchiano (1995) report that 10-year-old Spanish-speaking children made more lexicalizations than adults while reading pseudowords. The authors suggest that this is because children retain orthographic regularities in their memory and construct analogies to differentiate them (Sebastián-Gallés & Parreño-Vacchiano, 1995). At that age, Arduino & Burani, (2004) propose that readers can alternate the decoding strategy with automatic access to the lexicon depending on the demands of the orthography.

Further evidence for this finding comes from Davies, Cuetos, and González-Seijas (2007) who report that both routes are available to primary school-aged children as indicated by the presence of length, frequency and neighbourhood effects. Similar conclusions have been made from a number of studies of reading development in transparent orthographies such as Italian (Bates, Burani, D’Amico, & Barca, 2001) and

Turkish (Öney & Goldman, 1984). The following section will provide an overview of previous work carried out regarding reading development in Turkish-speaking children and skilled reading in Turkish-speaking adults. The design and implementation of a pilot study will be introduced, followed by the reporting of the main study and its findings.

4.2 Reading Development in Turkish children

Studies that have been carried out examining reading development in Turkish-speaking children have thus far been limited. The majority of said studies have measured phonological awareness skills in the context of highly accurate word and pseudoword reading in Turkish. For example, in one of the first cross-linguistic and Turkish language investigations, Öney and Goldman (1984) compared the pseudoword reading ability of Turkish and American schoolchildren and found that the Turkish-speaking children were both more accurate (94% vs 59%) and faster in the 1st grade. In a longitudinal follow up, both cohorts were reported to have reached ceiling level performance for accuracy at 3rd grade, but the Turkish-speaking cohort was reported to still be more fluent. These findings have been replicated several times in similar cohorts. For example, Öney, Peter, and Katz (1997) assessed changes in phonological mediation in word recognition in Turkish and American second and 5th graders as well as adults. The authors report a greater phonological activation in Turkish than in English at all levels of reading skill as well as a stronger effect on younger than on older readers.

Furthermore, Öney and Durgunoğlu (1997) carried out a longitudinal investigation of reading development in Turkish by following 30 children through grade 1 in which they were tested three times in October, February and May. By May (end of the first grade), reading and spelling performance was stated to be at ceiling level. In addition, Öney and Durgunoğlu (1997) report that letter knowledge, i.e., the ability to recognize letters was a better predictor of reading skills than phonological awareness. These findings were understood to be reflective of the limited time duration of effect for phonological awareness in Turkish, possibly being restricted to the first few months of literacy education (Öney & Durgunoğlu, 1997). The authors also found that, by the end of the school year, reading accuracy for word and pseudoword reading was highly correlated

(.92). Both the ceiling level findings and high correlation of word and pseudoword reading have been interpreted as being reflective of the near one-to-one GPC of Turkish. The findings of this study as well as findings from several other highly transparent alphabetic orthographies; Finnish (Holopainen, Ahonen, & Lyytinen, 2001) , Greek (Porpodas, 1999) and Italian (Cossu, 1999) are suggestive of the rapid development of phonological recoding skills in contrast with opaque orthographies such as English. This, in turn, is expected to further improve the levels of phonological awareness. Conversely, Babayiğit (2006) suggests that the ceiling effect findings coupled with the lack of control of verbal short-term memory and reading skills at the beginning of the grade, complicate the interpretation of the results. This being said, the influence of letter knowledge on early reading skills is largely in line with previous research (Blaiklock, 2004; Caravolas et al., 2001; de Jong & van der Leij, 1999).

Another more recent study comparing Turkish with English found that Turkish-speaking children could manipulate syllables more accurately earlier in a syllable-tapping task than English-speaking children (Oktay & Aktan, 2002) . Oktay and Aktan (2002) suggest that this is due to Turkish words having a well-defined syllabic structure as well as a low number of possible syllable types. In a phoneme deletion task, the Turkish-speaking children performed more accurately with the authors suggesting that this is due to the presence of strong vowel harmony, which requires that morphemes change to match the nature of the preceding vowel. It is posited that such manipulation may enable Turkish-speaking children to identify individual phonemes more quickly and accurately. Moreover, Durgunoğlu and Öney (1999) reported 94% accuracy in a syllable counting task and 67% accuracy in a phoneme-counting task for Turkish pre-schoolers. The authors postulate that the well-defined syllable structure and highly suffixed morphology of Turkish permits children to manipulate syllables and as well as final phonemes, with ease (Peynircioglu, Durgunoğlu, & Öney-Kusefoglu, 1996). More recently, Babayiğit and Stainthorp (2007) assessed the role of phonological awareness and verbal STM in early reading skills of 56 Turkish-speaking preschoolers in a longitudinal study, following the children from preschool to Grade 2. Their findings lend support to the longitudinal correlation of preschool phonological short-term memory skills on subsequent reading development. However, the study failed to replicate the well-detected influence of phonological awareness on early literacy development. The authors offered two possible explanations of the unexpected findings. The first was that the extreme transparency of the Turkish orthography might not impact reading fluency by the end of Grade-1, and the second was possible methodological limitations related to the research paradigm. In a similar investigation, Babayiğit and Stainthorp (2010) examined the role of phonological and grammatical awareness as well as RAN (rapid automatized naming) and verbal STM in the development of reading and spelling skills. Grammatical awareness may be described as the ability to process the morphological and syntactic structures of a spoken language and is thought to aid contextual word recognition as opposed to single word recognition. The study found that RAN was the most powerful longitudinal predictor of reading speed and that phonological awareness was found to reliably predict spelling skills. Furthermore, Babayiğit and Stainthorp (2011) carried out a longitudinal investigation of reading fluency and comprehension using measures of phonological awareness, RAN, vocabulary, listening comprehension, and working memory. By following, children from second and fourth grades into third and fifth grades, respectively, the authors report similar findings to their previous work in that RAN was a strong predictor of reading fluency, and that phonological awareness was the strongest predictor of spelling. Although the authors use of a composite measure of reading fluency (made up of word lists and narrative text) is a valid approach to the investigation of reading in transparent orthographies, the use of a computerized, discrete trial, test method to examine visual word recognition would facilitate further fine-grained statistical analysis (Davies, Cuetos, & Glez-Seijas, 2007).

To summarize, the limited research into the development of reading skills in Turkish have been highly informative with regards to the rapid development of phonological awareness skills, which in turn enhances the development of reading ability in highly transparent orthographies. To the best of the author’s knowledge, to date, there has only been one study to incorporate speed of processing, as measured by RAN as a predictor variable of reading ability. Beyond this, there is little evidence of more comprehensive investigations into visual word recognition and reading skill development in Turkish-speaking children. Further still, there are no reports of developmental dyslexia in the literature. This gap in the literature, along with the methodological concerns of previous studies motivates the current study. Before discussing the development of the pilot investigation, the following section will aim to provide an overview of recent psycholinguistic research regarding visual word recognition studies that have been conducted in adults as an insightful contribution to reading development in Turkish.

4.3 Skilled Reading in Turkish adults

Similar to the literature in reading development in Turkish, to date, there has been limited investigation into skilled reading in Turkish adults. The doctoral research carried out by Öney (1990) is perhaps the earliest attempt at exploring the processes involved in skilled reading in Turkish. In this study, participants undertook both word naming and lexical decision tasks in which they were presented with three types of sentences (contextually consistent, inconsistent or neutral words) and had to respond to the preceding target words in order to investigate semantic priming and sentence-context effects. Öney (1990) reported that both word naming and lexical decision were facilitated by consistent, contextual words, whereas inconsistent context words inhibited naming. The overall pattern of results was suggested to be indicative that Turkish supports a substantial reliance on phonologically analytic strategy, i.e., the sublexical route in word recognition. However, in her doctoral thesis, I.Raman (1999) contended that if Turkish readers relied solely on the sublexical route, there is no reason to expect a significant semantic priming effect in naming. Furthermore, the finding of a reliable frequency effect from skilled adult readers in Turkish indicated that even readers of completely transparent orthographies make primary use of the lexical route (I.Raman, Baluch & Sneddon, 1996). As stated previously, this would suggest that during literacy development, there must be a shift from nonlexical to lexical reading strategies. In addition to this, I.Raman, Baluch, & Besner (2004) propose a model in which both lexical and nonlexical processing appears to be in parallel, interactive and equally automated for word naming in Turkish. In a similar line of work in Italian, Pagliuca, Arduino, Barca, and Burani, (2008) reported the presence of lexicality effects which is thought to be indicative of a primary reliance on the lexical route for reading aloud.

Recently, I.Raman (2011) investigated the degree to which age of acquisition (AoA) would influence dyslexic adults in word and picture naming in comparison to non-dyslexic controls in the transparent orthography of Turkish. The results of the study found that participants with dyslexia performed considerably slower than non-dyslexic controls in both the word and the picture naming tasks. One of the main findings of this investigation was that the overall error rates were not significantly different between reading impaired and typical reading participants and both groups performed close to ceiling level. With respect to this, word reading accuracy is considered a poor indicator in transparent orthographies at least where developmental dyslexia is concerned. Ultimately, very low error rates may be indicative of a speed-accuracy trade-off in which performance is compromised in favour of accuracy rather than speed (MacKay, 1982).

4.4 A NOTE OF THE TURKISH CYPRIOT EDUCATION SYSTEM

The Turkish Cypriot education system has gone through substantial changes over the past decade. The most recent iteration (2013) can be broadly defined as being made up of five parts:

1. Preschool – Children aged 4-6 (Non-compulsory)
2. Primary - Children aged 7-12
3. Secondary Junior – Adolescents aged 13-15
4. High School- Adolescents aged 16-17
5. University – 18+

Since 1995, the rate of schooling in North Cyprus has been 100%. In addition, all public primary, secondary and high schools are free, compulsory and maintained by the Ministry of National Education and Culture. According to the Education Statistical Yearbook (2019), the student-teacher ratio for public pre-primary and primary schools are 14.5 and 21.4, respectively. The total number of students across the island in pre-primary and primary education are 7360 and 19861 respectively.

In a similar vein to other studies in Turkish reading, the number of studies on early literacy education in Northern Cyprus is low. In a comprehensive investigation of teacher's views of primary literacy education in Northern Cyprus, Babayiğit and Konedralı (2007) report that nearly two-thirds (64%) of Grade 1 teachers use an eclectic method of reading instruction whereas 25% use the whole-language method. It is further reported that there was tremendous support (94%) for the use of a phonics-type approach in reading instruction. In addition, Kargın, Güldenoğlu & Ergül (2017) report that Turkish preschoolers develop a superior vocabulary than phonological awareness, letter identification and listening comprehension skills. The rationale for this, in their view, is that preschool teachers' negative beliefs about imparting early literacy skills such as letter identification or phonics teaching may be a contributing factor to the poor performance of young preschoolers on phonological awareness, and letter identification tasks. The following section will describe the theoretical foundations used in the development of a battery of tests to evaluate word decoding ability in Turkish-speaking children.

4.5 THE PILOT STUDY

The aim of the pilot study is driven, in part, by the need to address the lack of reports as well as methodological limitations in the relevant literature on Turkish reading development. The development of a battery of tests to measure the impact of several cognitive constructs on word decoding ability began as an attempt to offer a more systematic and comprehensive overview of reading development in Turkish-speaking children. Thus, the over-arching aim of the pilot will be to establish if the battery of tests and stimuli proposed in this study is appropriate for the psycholinguistic investigation of Turkish-speaking children who are in the early stages of reading development.

Although there has been important work carried out in the investigation of Turkish reading development (e.g., Durgunoğlu & Öney, 1999; Öney & Durgunoğlu, 1997)), there has been minimal research conducted on the simultaneous influence of cognitive, metalinguistic and linguistic processing skills that may influence Turkish-speaking children who are learning to read.

4.5.1 BATTERY OF TESTS

Briefly, tests were derived from existing measures found in the literature in other orthographies and were carefully constructed in order to overcome potential methodological issues and also avoid copyright infringement. The exception to this approach was with regard to non-verbal IQ, which was measured by the widely used

Raven’s Colored Progressive Matrices (Raven, 1986). The eight cognitive constructs explored by the battery were:

i) Reading accuracy and speed in single-word naming
ii) phonological awareness (PA),
iii) Rapid Automatized Naming (RAN),
iv) Visual Attention (VA) Span,
v) Non-verbal IQ
vi) Phonological short-term memory (PSTM),
vii) Visuo-spatial short-term memory (VSSTM)
viii) Working memory.

Finally, in addition to the psycholinguistic investigation carried out, the results will provide an insight into the validity of the conceptual representation of the tests developed as well as the internal reliability of the scales themselves. As described previously, the following section will provide a detailed account of the process in designing the battery of tests for the investigation of reading development in Turkish-speaking children. The tests selected for this study include:

<u>Single Word Reading (SWR) (Word/nonword naming)</u>

Visual Word Recognition is considered to be the first step of the reading process. The vast majority of studies explore the association between reading ability and reading time within the framework of accurate reading performance. Moreover, SWR tests are commonly utilized to identify children with reading impairments (Wydell & Butterworth, 1999) . In order to investigate universal and language-specific aspects of typical and atypical reading development, Ziegler and colleagues (2003a) state that it is essential to factor in theoretically critical marker effects of the reading process, such as lexicality and length effects. To this end, it was decided that lexicality (word vs pseudoword), length (short vs long) as well as frequency (low vs high) effects would be explored whilst controlling for Age-of-Acquisition (AoA) (Morrison & Ellis, 2000) and orthographic familiarity effects, in which familiar words are named faster than the unfamiliar words (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) .

Pseudowords were matched with words for length and number of syllables as well as being phonologically legal.

The stimulus for Turkish test words was selected from I.Raman, E.Raman and Mertan (2014) and Turkish pseudowords were generated using the Turkish plugin (Erten, Bozsahin, & Zeyrek, 2014) of Wuggy, a multilingual pseudoword generator (Keuleers & Brysbaert, 2010) . The justification for using the above stimulus sets is as follows: At the time of designing the experiment, the database of I.Raman, E.Raman and Mertan (2014) was the only such resource available in Turkish that had both AoA and familiarity ratings freely available. Words were selected on the basis that they were acquired early on in development as well as representing highly familiar concepts. Reliable AoA effects have been previously reported in Turkish-speaking adults as well as an adult dyslexic cohort (I.Raman, 2006; 2011). Of note, the Turkish dyslexic adults were defined by their distribution of reaction times and error rates when compared to age-matched controls as part of a wider study and may not be indicative of true dyslexic status.

With regards to pseudoword generation, previous studies in Turkish children have used manual methods of pseudoword selection by which real words are modified by one or two letters with other letters legal to the language’s rules. This potentially introduces an experimenter bias into the research design based on the implicit knowledge regarding the research’s hypothesis (Balota et al., 2007). For the purpose of this study, in order to overcome these methodological limitations, pseudowords were generated in the Turkish version of Wuggy using a semi-automated approach. A similar procedure was recently utilized on a lexical decision task in Turkish-speaking adults (Erten et al., 2014). Firstly, the 40-word stimulus list generated for real words were used as a template for generating the pseudowords. The parameters in Wuggy were set so that each real word generated ten candidate pseudowords that were matched for length and length of subsyllabic segments. Of the ten candidate pseudowords, the one that manifested the quantitatively smallest deviation from the reference word was selected for the pseudoword stimulus list. Table 19 provides an overview of the words and pseudowords used in the pilot investigation.

TABLE 19: DESCRIPTIVE CHARACTERISTICS OF THE SIX GROUPS OF WORDS AND PSUEDOWORDS CREATED BY MANIPULATIONS OF LEXICALITY, LENGTH AND FREQUENCY.

	Length	Frequency	AoA	Familiarity
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Short Low Frequency	3.30 (0.67)	14.84 (9.26)	2.25 (0.32)	4.66 (0.11)
Short High Frequency	2.70 (0.67)	144.75 (109.36)	1.80 (0.31)	4.74 (0.17)
Short Psuedoword	2.95 (0.68)	-	-	-
Long Low Frequency	6.00 (0)	10.65 (3.93)	2.42 (0.42)	4.68 (0.22)
Long High Frequency	5.90 (0.39)	58.22 (51.83)	2.04 (0.39)	4.76 (0.12)
Long Pseudoword	5.95 (0.22)	-	-	-

Methods

Stimuli & Procedure for Word Reading Task:

For the single word reading task, children were presented a series of 40 words and 40 pseudowords one at a time in a randomized order. Each stimulus appeared in the centre of a computer screen for 2500 ms with an inter-stimulus interval (ISI) of 1500ms. The order of presentation was randomized for each child who was instructed to read aloud the words/ pseudowords as quickly and accurately as possible. A block of practice trials with five words and five pseudowords was presented for naming prior to the main experiment so that the children could familiarize themselves with the task, generally, and with the notion on pseudowords, specifically.

The SuperLab 5 (Cedrus Corporation, San Pedro, CA) software package was used to control the experiment and to collect naming latencies via a voice-activated microphone (SV-1) which has a 1ms resolution. Despite the frequent use of voice-keys in psycholinguistic research, there are a number of concerns that have been voiced regarding their use (Rastle & Davis, 2002) . The major concern is that voice-keys are considered to be particularly inaccurate in detecting acoustic onsets. Another concern is that accuracy varies significantly across onset phonemes (Duyck et al., 2008; Kessler, Treiman, & Mullennix, 2002) . The resulting effect is bias data or the production of contradictory results. In order to overcome these limitations of voice-key use, a USB microphone also collected responses in order to measure accuracy and speed offline using Praat 5.4 speech signal analysis software (URL: www.praat.org)

Stimuli & Procedure for Phonological Awareness Tasks:

Phonological awareness is a complex and multifaceted skill which refers to an individual's understanding of, and ability to manipulate the sound structure of oral language and is thought to be central in reading development. Due to the weight of evidence regarding the relationship of phonological awareness with reading ability across orthographies (Castles & Friedmann, 2014; Goswami & Bryant, 1990; Janssen, Bosman, & Leseman, 2013) , it was decided that for the pilot investigation, three tests of phonological awareness would be administered. Also, the rapid development of phonological awareness in transparent orthographies, including Turkish (Durgunoğlu & Öney, 1999; Öney & Durgunoğlu, 1997) provided supplementary motivation to select three tests of phonological awareness that increased with difficulty and thus provide an opportunity to investigate the contribution of phonological awareness without the consequence of ceiling effects observed in previous studies. Finally, the simple syllabic structure of Turkish, similar to other highly transparent alphabetic orthographies of Spanish and Italian, stipulates that when syllables are CV units, onsets, rimes and phonemes share a degree of equivalency (Goswami, 2008). With this in mind, the two tests selected were at the phonemic level, and a third test was selected at the onset-rime level.

Phoneme Deletion

Phoneme deletion tasks have a long and established use within the reading development literature, particularly when considering that deficits in phonemic awareness are frequently cited as the most influential factor in an individual's probability of reading failure, especially in opaque alphabetic orthographies (e.g., Shaywitz & Shaywitz, 2003; Ziegler & Goswami, 2005) . The phoneme deletion task requested that the child pronounce a sound sequence after deleting a specified sound (e.g. say "/kedi/ meaning "cat" without /k/"). 10 words were selected from I.Raman et al., (2014) on the basis that they had not been used in the previous reading task. For the Phoneme deletion task, participants were asked to delete a specified phoneme from a spoken word. Both the initial and final phoneme locations were manipulated, i.e., 10 words (5 initial, 5 final). Furthermore, the resulting sound sequence following deletion was always a pseudoword. The number of correct responses out of 10 was scored.

Phonemic Segmentation

Phoneme segmentation is the ability to segment spoken words into constituent phonemes. With regards to development, the ability to successfully perform segmentation tasks shifts from the ability to segment at the onset-rime level to phonemic segmentation (Savage & Carless, 2005) . To this end, 20 words were selected from I.Raman, E.Raman and Mertan (2014) on the basis that they had not been used in the reading tasks. In a similar fashion to the Yopp-Singer Test of Phonemic Segmentation (Yopp, 1988), participants were instructed to sound out the letters of a given word. To confirm that a child had identified the words correctly, he/she was requested to repeat each word before segmenting it into phonemes. Before starting the experiment, the child was shown a practice item and the process of segmentation was explained and demonstrated to them. Participants were then presented with 20 such words with varying length. The raw score of this task was calculated as the number of correct responses out of 20.

Spoonerism.

A spoonerism task (e.g., Perin, 1983) requires exchanging the first phoneme of two words pronounced by the experimenter one after the other to form two new words or pseudowords (e.g., /car-park/ -> /par-cark/). Spoonerism tasks are assumed to tap into the ability to hold and manipulate phonological information. Additionally, segmentation in a spoonerism is at the level of onset and rime and not at the phonemic level (Landerl & Wimmer, 2000) . Thus, when considering tasks demands, the spoonerism task is considered to be a difficult task. To this end, 20 pairs (40 words) were selected from (I.Raman et al., 2014) on the basis that they had not been used in the reading, phoneme deletion and segmentation tasks. The experimenter provided an example before the task began in order to confirm that the child understood the task requirements. In order to score the spoonerism task, each correctly named word-pair was awarded two points with one point given for each new pseudoword correctly exchanged.

Phonological Automatization - Rapid Automatized Naming (RAN) (Rudel & Denckla (1976)

Though there is much debate within the literature, it is generally agreed that RAN tasks can be regarded as an index of the retrieval speed of phonological information from memory (Wagner & Torgesen, 1987) . Recently, a number of studies have lent support to the double-deficit hypothesis (Bowers & Wolf, 1993) in which phonological awareness and RAN make independent and differential contributions to reading ability (Cronin, 2013; Wolff, 2014). Similarly, RAN is often cited as the best predictor of reading ability in highly transparent orthographies (Landerl et al., 2013; Moll et al., 2014). For the current study, participants' responses to objects, colours, numbers, and letters (Denckla & Rudel, 1976c) were measured using 50 items arranged in 5 rows of 10 items each. None of the five different token items for each subtest appeared consecutively on the same line. The overall time taken in seconds to process each subset was reported.

Visual Attention Span (Ans, Carbonnel, & Valdois, 1998)

Visual attention Span (VA Span) is described as the quantity of distinct visual elements that can be processed in parallel in a multi-element array (Ans, Carbonnel, & Valdois, 1998). It is considered to be a measure of visual attention span capacity rather than a verbal STM task. Of interest to this thesis, is that there are no reports of VA Span in the Turkish literature. This merits further investigation as VA Span skills may be modulated by grapheme-to-phoneme consistency (Bosse & Valdois, 2009a) and as such, in accordance with the PGST (Ziegler & Goswami, 2005) , the finer grain size found in highly transparent orthographies would translate to smaller VA Spans in children learning to read such orthographies.

The methodology of the visual attention span tasks followed the exact procedure of the global report outlined by Bosse and Valdois (2009a). Briefly, stimuli were made up of 20 random five-letter strings (e.g., R H S D M) constructed from a selection of 10 consonants (B, P, T, F, L, M, D, S, R, H). Each letter appeared a total of 10 times, twice in each position. In addition, letters were never repeated in a string, and the five-consonant strings never matched the skeleton of a real word. Each trial began with the presentation of a central fixation point for 200 ms, which was followed by a blank screen for 50 ms. A horizontally centred five-letter sequence was then presented on the fixation point for 200 ms. The participant was requested to name as many letters as possible immediately after they disappeared independently of position. The total number of correct letters named out of 100 was used as the measure for VA Span.

Non-verbal IQ - Raven's Coloured Progressive Matrices (RCPM) (Raven, 1986)

In an attempt to quantify a measure of non-verbal intelligence in children, the experiment also required the completion of Raven's Coloured Progressive Matrices (RCPM) (Raven, 1986) . Though widely used in the typical and atypical reading development literature, Goswami (2003) argues against the use of nonverbal IQ tests when they are the only matching measure used between dyslexics and controls. Whilst this reservation is valid, partly due to the functional nature of nonverbal IQ tests, measures used within a wider framework in which children are excluded from membership of dyslexic groups have demonstrated usefulness (Boets et al., 2010; Zoccolotti et al., 2005). The child version of the RCPM consisted of 36 coloured items in 3 sets (A, Ab, B), with 12 items per set. The items were a sequence of perceptual and conceptual matching exercises, and children had to complete the pattern by selecting one out of six possible pattern matches (Raven, 1986). The test required minimal verbal instructions, which were given in Turkish.

Memory

Phonological STM

The phonological loop component of Baddeley's Working Memory model (Baddeley, 2003) is thought to reflect the storage and maintenance of written phonological sequences as an addition to speech stimuli. Furthermore, there is some evidence that suggests that tasks that measure short-term phonological memory are moderately correlated with individual differences associated with reading skill (Gathercole & Baddeley, 1993) . Traditionally, digit span tasks are utilized in the assessment of developmental dyslexia in order to measure a child's ability to store verbal information (Everatt et al., 2010). It has further been suggested that digit span measures of phonological short-term memory may be constrained by the pure individual capacity of the phonological loop in addition to more general working memory limitations (Wagner & Muse, 2012) . For the pilot study, participants were verbally presented with a series of random digits (e.g., '8, 3, 4'). The lists of numbers range from 2 to 8 numbers long.

Using SuperLab 5, after the pre-recorded verbal presentation of a list, the participant was required to immediately repeat them back in the order they were presented by the experimenter. If they did this successfully, they were given a second list of the same length, progressing and subsequently moving onto longer lists (e.g., '9, 2, 4, 0'). The test was terminated when the child failed to correctly repeat both of the lists of a given length. The length of the longest list a participant recalled was scored as their digit span.

Visuo-spatial STM (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999)
It is currently considered that non-verbal short-term memory is made up of discrete visual and spatial/sequential components. The visual subcomponent of the previously mentioned visuospatial sketchpad is thought to be responsible for the binding of static visuo-spatial information. With respect to this, the Visual Patterns Test (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999) is considered to be a measure of simultaneous visual working memory (Trick, Mutreja, & Hunt, 2012) . Moreover, the role of visuospatial memory in reading development has received little attention (Bacon, Parmentier, & Barr, 2013) . A study carried out by Gathercole and colleagues (2006) found that visuospatial short-term memory scores were low in dyslexic children and reported correlations with measures of complex memory, phonological processing and performance IQ.

For the pilot a paradigm similar to the Visual Patterns Test (VPT) (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999) was designed taking into consideration the need to limit the use of verbal coding (Brown, Forbes, & McConnell, 2006) . Briefly, children are shown a series of grids. In each grid, half the squares were filled in black while the remaining ones remained empty. The grids increased in size and thus complexity beginning with 4 squares (2 black), 6 squares (3 black) and so forth. The experiment was designed to tap into the visual aspects of non-verbal short-term memory whilst excluding the spatial-sequential component.

The Visual STM task was completed using the SuperLab 5 software package. Each pattern was presented to the participant for 3000 ms. The participant was then asked to reproduce the pattern by marking squares in an empty grid of the same size as the one bearing the pattern just presented. Each grid size represented a level, and each level had three differentially coded grids within it; in order to progress to the next level, children had to correctly code two of the three grids. The VSSTM score was calculated to be the level at which the child failed to progress.

<u>Working Memory (WM)</u> (Baddeley & Logie, 1999)
Working memory may be conceptualized as a processing resource of limited capacity, involved in the conservation of information whilst simultaneously processing the same or additional information (Swanson, Xinhua, & Jerman, 2009) . With respect to reading development, tasks of working memory are frequently used as a predictor variable. It was decided, similar to other research in the literature, that a backward digit span task would be utilized as an index of working memory capacity. Complex memory span tasks are postulated to elevate demands on both on the central executive and the phonological loop (Baddeley & Logie, 1999) . However, it must be noted that there is considerable debate as to whether the backward digit span test is representational of WM or STM processes (Gathercole, Pickering, Ambridge, & Wearing, 2004). Similar to the phonological STM task above, participants were verbally presented with a series of random digits (e.g., '8, 3, 4'). The lists of numbers range from 2 to 8 numbers long. In the Working memory task (backward digit-span task) the participant was required to reverse the order of the numbers. The length of the longest list a participant could recall was scored as their working memory span.

<u>Experimental Hypotheses</u>

H1: Beyond the first grade, word length and frequency effects will be evident for typically developing children in so far as longer words and less frequent words will take longer to name. This is thought to be reflective of both lexical and sublexical processes.

H2: Word reading accuracy will be close to ceiling level. Word reading speed will be the best predictor of reading ability, particularly with reference to older children (who have reached ceiling level performance on word reading accuracy (Wimmer & Schurz, 2010) .

H3: In typically developing Turkish children, it is expected that there will also be a lexicality effect positing that words will be named faster than nonwords. This being said, it is also reasonable to postulate that the transparent nature of Turkish will also allow for the processing of nonwords with relative ease. In addition to this, a lexicality by length interaction effect would be expected, stipulating the availability of two routes of reading aloud in children learning to read a highly transparent orthography.

H4: There will be a heterogeneous spread (large variance) in the manifestation of visual attention span and memory tasks that reflect individual differences in these underlying abilities. This should hold true for both typically developing children and dyslexics. This being said, it is expected that all cognitive measures would improve with age.

The following section will provide an account of the application of the newly developed battery of tests in a pilot study by planning to

i) measure the reliability of the constructs
ii) to provide initial insights into word decoding skills in Turkish-speaking children based on the novel methodological design of the pilot study
iii) to make further recommendations on how the battery can be improved for the main study.

4.5.2 Methods

Participants

Participants were 58 randomly selected Turkish-speaking children (Table 20) from Grade 2 and 5. Children were recruited from three mainstream primary schools in Famagusta, Cyprus after receiving backing from the Turkish Republic of Northern Cyprus Department of Primary Education. Informed assent and consent were obtained from all children and their parents. A further nine children who were originally recruited were removed from the study, seven due to technical issues[1], one due to non-compliance and one for requesting to be removed during the testing stage. This study

[1] The nature of technical difficulties and implications for future design of the methodology will be discussed in the final section of the pilot study.

was approved by the Brunel Psychology Research Ethics Committee (Appendix 4) and by the collaborating institution, namely the Eastern Mediterranean University's Psychology Department.

Children's demographic information was obtained together with language use and familial history from parents using a questionnaire recently constructed for the purpose of this study. Acquiring a wide range of information regarding a child's background is useful in developing individualized profiles that take into account factors such as language proficiency and familial environmental differences. An index of socioeconomic status (SES) for each child was derived from parental education level and was computed by the mean of the highest attained educational level of both parents rated on a 7-point scale. This approach is habitually utilized in the reading literature and is thought to be highly predictive of overall SES. The Edinburgh Handedness Inventory (Oldfield, 1971) was translated into Turkish in order to determine and control for handedness. All participants had normal or corrected-to-normal vision and are all monolingual Turkish-speakers. Table 16 illustrate the characteristics of the pilot cohort.

TABLE 20: CHARACTERISTICS OF THE PILOT COHORT BY GRADE.

	Grade 2		**Grade 5**	
N	27		31	
Sex [% girls]	56		45	
	Mean	SD	Mean	SD
Age (Months)	90.15	5.45	124.26	4.49
SES (2.5-6.5)	4.18	1.29	3.93	0.73
Handedness	21 RH	6 LH	28 RH	3 LH

Procedure

The battery of tests (described above) was carried out individually in a single session lasting approximately one hour in a quiet room within the children's school during school hours. The experimenter and a trained research assistant conducted all testing.

The battery was presented as a fully computerized task with the exception of the Visuo-spatial short-term memory task in which children were required to write answers in a booklet with blank grids. Each session followed a fixed order of presentation beginning with Visual Short-term Memory, followed by RAN, VA Span, Single Word/Pseudoword naming, Raven's CPM, Digit span (Forward and Backward) and ended with PA. Children were informed that they could take brief breaks between tasks but not during tasks. The exception to this was the single word/pseudoword reading task in which a break was built into the task design after children had reached the midpoint (40 words/ pseudowords) in order to avoid fatigue.

Statistical Analysis Plan

The data analysis approach for this pilot study was designed to address the previously stated research hypotheses. For each response of each participant for the single word/pseudoword reading task, reaction time (RT) was calculated from the onset of the stimulus until the onset of the response. Though this was a built-in measure of the SV-1 voice-activated key, each response was also checked off-line using the digital sound files recorded separately by the USB microphone. Furthermore, only RTs of correct responses were used to calculate the RT means. The digital sound files were also used to score accuracy and were reported as an overall percentage correct. With regards to error coding for the pseudowords, in line with previous research, a lenient error-coding criterion was adopted in which all phonologically plausible responses were considered correct (Landerl, Wimmer, & Frith, 1997; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003b) .

In order to gain valid results from the use of parametric tests, a common assumption is that the dependent and independent variables are approximately normally distributed. Transformations allow for data to be manipulated into a normal distribution. All raw scores with violations of normality were converted into z-scores within each grade level. It was anticipated that several variables would have highly skewed distributions, e.g., word reading accuracy and phoneme deletion as previously reported in transparent orthographies. When such distributions were detected, a ranking procedure was utilized prior to normalization (Landerl et al., 2012). Pearson's correlation was used to measure the strength of association between two continuous variables, namely Reaction Time and Accuracy measures and the predictor variable scores.

A series of multiple regressions were used to predict single word reaction time, a continuous dependent variable, given a number of independent variables. Multiple regression analysis was also utilised to offer an account of how much the independent variables can explain the variation of the dependent variable score. For the purpose of regression analysis, where several measures of a construct are measured, principal components analysis (PCA) was used to cluster cognitive skills (RAN, Memory and Phonological Awareness) into principal components. However, it is acknowledged that large sample sizes are required in order for a principal components analysis to produce a reliable result. Countless diverse guidelines have been suggested to calculate minimum numbers for reliable PCA that differ in terms of using either absolute sample size numbers or a multiple of the number of variables in the sample. Generally speaking, a minimum of 150 cases or 5 to 10 cases per variable has been recommended as a minimum sample size.

The mixed ANOVA will be used to compare the mean differences between groups (between-subjects factor) that have been divided on three independent variables (within-subjects factor). For the purposes of this study, the three independent variables in question will be lexical status, length and frequency. The central aim of a mixed ANOVA is to establish the presence or absence of an interaction between the two independent variables on the dependent variable. Statistical analysis for the pilot study was carried out using IBM SPSS Statistics 20.

4.5.3 Results

Reliability Analysis

Before proceeding with planned comparisons, reliability analysis was conducted on all measures used in the battery. The resulting alpha coefficients are reported in Table 21 (word/pseudoword tasks) and 22 (cognitive tasks) below. Cronbach's α is considered to be an index measure of internal consistency with a measurement criterion of 0.7 and above being indicative of a reliable measurement tool (Nunnaly, 1978). In line with this, the battery of tasks designed for this study (with the exception of the phonological awareness phoneme deletion task (0.66)) are considered reliable measures of their underlying constructs (0.72-0.96). Additionally, the word and nonword measures were also highly reliable (0.875-0.942).

TABLE 21: ALPHA COEFFICIENTS OF WORD/ PSEUDO WORD NAMING TASKS.

	Short			**Long**		
	Low Freq	High Freq	Pseudoword	Low Freq	High Freq	Pseudoword
Cronbach's α	0.918	0.923	0.942	0.875	0.912	0.922
N of items	10	10	20	10	10	20

TABLE 22: ALPHA COEFFICIENTS OF COGNITIVE BATTERY TASKS

Task	**RAN**	**VA Span**	**PA -Del**	**PA- Seg**	**PA – Spoon**
Cronbach's α	0.84	0.96	0.66	0.79	0.92
N of items	4	20	10	20	20
Task	**PSTM**	**VSTM**	**WM**	**RCPM**	
Cronbach's α	0.74	0.85	0.76	0.72	
N of items	8	27	7	36	

Descriptive

Regarding Word/Pseudoword naming speed, and in line with the current literature (see Baayen & Milin, 2010; Rønneberg & Torrance, 2017), all outliers with a residual outside -/+2.5 SD from the mean reaction time for each condition and for each grade were removed, and a new mean was calculated. From Grade 2, there was one extreme outlier in the SNW reaction times. For Grade 5, there were 2 extreme outliers:

1 in the Long Low Frequency and 1 in the Long High-Frequency group. Furthermore, there were 53 false triggers (1.5%) in naming (RT<350ms) and 177 (5%) incidences of a Non-Response, which were removed before accuracy for each condition was calculated. The descriptive statistics for non-verbal IQ, cognitive measures and reading skills by Grade are presented in Table 23. A preliminary examination of means and standard deviations of variables measured in this pilot suggested that, in general, there was a gradual development of these skills with age. Grade 5 students significantly outperformed second graders on all measures except for non-verbal IQ, phonological short-term memory and all but one (Long Low-Frequency Words) measures of single word/pseudoword reading accuracy. In general, Turkish-speaking children displayed a high degree of accuracy on the word/ pseudoword task regardless of grade and reached ceiling or near-ceiling levels of accuracy with the notable exception of long pseudowords. In contrast, all measures of reading speed were a particularly sensitive index of the variability in reading skill between the two different grades.

TABLE 23: MEANS, STANDARD DEVIATIONS, KURTOSIS AND SKEWNESS OF INDIVIDUAL TESTS AS WELL AS T-TESTS COMPARING MEANS ON TESTS BETWEEN GRADES. *DENOTES A SIGNIFICANT DIFFERENCE BETWEEN GRADES

	Grade 2				**Grade 5**					
	Mean	SD	Kurtosis	Skewness	Mean	SD	Kurtosis	Skewness	t(56)	P
Non-verbal IQ Max= 36	21.30	4.47	-0.75	-0.4	23.35	4.07	1.94	0.51	1.84	.72
Reading										
SLF Accuracy (%)	99.26	2.67	10.67	-3.45	99.35	2.50	12.72	-3.73	0.14	.889
Speed (ms)	1043	66.41	-0.87	0.37	868	35.46	-0.856	0.21	12.77	<.001*
SHF Accuracy	98.52	6.02	21.31	-4.53	99.68	1.80	31.00	-5.57	1.02	.311
Speed	892	51.07	2.84	0.59	716	50.57	0.708	0.83	13.16	<.001*
LLF Accuracy	95.93	9.31	8.28	-2.77	99.68	1.80	31.00	-5.57	2.2	.032*

Speed	1339	113.2	-0.34	0.21	1118	53.54	-0.103	-0.66	9.74	<.001*
LHF Accuracy	95.21	9.75	4.84	-2.27	96.56	12.22	25.91	-4.94	0.46	.647
Speed	1195	79.83	0.86	-1.01	952	69.18	0.512	0.79	12.45	<.001*
SNW Accuracy	92.76	6.27	-0.86	-0.4	91.82	9.96	5.37	-2.06	0.42	.675
Speed	1119	56.49	-0.48	0.32	893	70.00	-0.796	0.52	13.44	<.001*
LNW Accuracy	74.08	18.24	5.03	-1.9	73.59	19.14	-0.24	-0.72	0.1	.92
Speed	1582	121.01	0.42	-0.22	1331	115.50	0.498	1.69	8.07	<.001*
RAN										
Objects (secs)	67.04	14.86	-0.56	0.72	53.98	10.14	0.18	0.18	3.95	<.001*
Colors	62.76	14.02	-0.66	0.53	44.80	9.30	4.26	4.26	5.82	<.001*
Letters	37.23	8.06	1.29	1.16	26.70	5.11	1.64	1.64	6.02	<.001*
Numbers	41.19	7.15	1.22	0.78	27.76	4.53	2.95	2.95	8.66	<.001*
PA										
Deletion	8.74	1.29	-0.46	0.76	9.48	0.85	3.10	-1.85	2.62	.011*
Max = 10										
Segmentation	15.81	3.11	0.1	-0.76	18.68	1.80	5.47	-2.06	4.36	<.001*
Max=20										
Spoonerism	17.71	10.42	-0.6	0.02	24.34	9.40	0.39	-0.58	2.46	.018*
Max=40										
VA Span	55.74	11.5	0.12	-0.18	67.84	16.29	0.40	-0.59	3.22	.002*
Max= 100										

VSTM	9.59	2.55	-0.57	-0.01	13.84	3.84	0.21	0.59	4.88	<.001*
Max= 24										
PSTM	4.67	1.18	-0.65	0.41	4.97	1.02	-1.16	0.48	1.05	0.3
Max=8										
WM	2.15	0.77	-0.05	0.28	2.94	1.00	-0.11	0.136	3.33	.002*
Max =7										

H1: Do children learning to read Turkish manifest Length and Frequency effects in Word Naming?

A 3 x 2 ANOVA, using frequency (high-frequency words, low-frequency words and pseudowords) and length (short words and long words) as within-subject factors and Grade as a between-subjects factor, showed significant main effects of frequency, $F(2,55) = 579.16$, $p < .001$, and length, $F(1,56) = 2357.38$, $p < .001$ as well as a significant interaction between them, Length x Frequency, $F(2,55) = 129.81$. In addition, there was also a significant interaction between length and grade, $F(1,56) = 11.42$, $p = .001$. There was no significant interaction between frequency and grade, $F(2,55) = 3.03$, $p = .057$. The results of the above analysis are presented graphically in Figure 12. The findings here are consistent with Hypothesis 1, stating that length and frequency effects will be present in Turkish children that are learning to read as previously reported in Turkish adults and children of other transparent alphabetic languages.

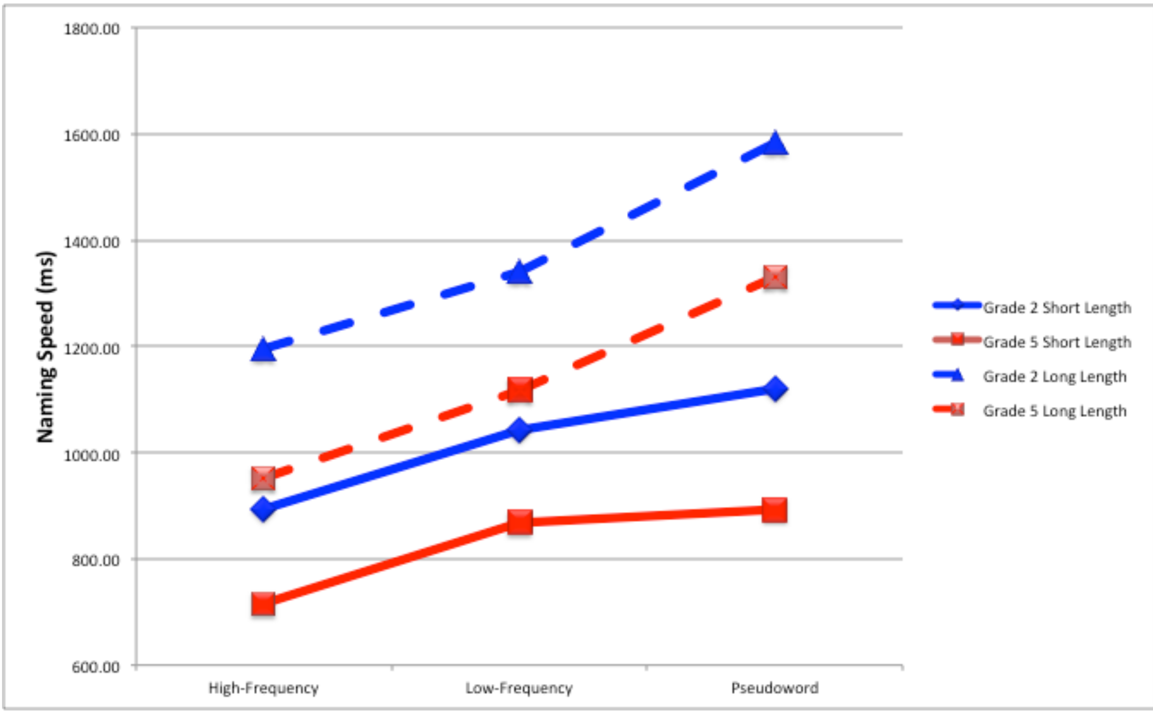

FIGURE 14: MEAN NAMING SPEED BY GRADE, FREQUENCY AND LENGTH

Following this, two 2 x 2 x 2 ANOVAs were conducted where length (short words and long words) and lexicality (words and pseudowords) were the within-subjects' factors and again grade (2 and 5) was the between-subjects factor for both low and high-frequency words. For low-frequency words, the results showed that the main effects of length, F (1,56) = 1770.65, p<.001, with short words (949ms) being named faster than long words (1221ms); and also found that lexicality F (1,56) = 292.75, p<.001 with words (1085ms) being named faster than nonwords(1223ms) was significant. The interaction between them was also significant, F (1,56) = 202.88, p<.001 in that the magnitude of the difference between short and long RTs for low-frequency words (271ms) was significantly increased for nonwords (450ms). There was also a significant interaction between Length and Grade F (1,56) = 4.22, p<.045 in that the magnitude of the difference between short and long RTs for Grade 2 stimuli (379ms) was significantly reduced for Grade 5 (344ms). In addition, there was a significant interaction between Lexicality and Grade F (1,56) = 6.10, p<.001 in that the magnitude of the difference between word and nonword RTs for Grade 2 stimuli (160ms) was significantly reduced for Grade 5 stimuli (119ms).

For high-frequency items, the results showed that the main effects of length, $F(1,56) = 2025.60$, $p<.001$ with short items (898ms) being named faster than long items (1257ms); and also found that lexicality $F(1,56) = 1160.52$, $p<.001$ was significant with words (938ms) being named faster than nonwords (1223ms). The interaction between them was also significant, $F(1,56) = 231.61$, $p<.001$ in that the magnitude of the difference between short and long RTs for high-frequency words (267ms) was significantly increased for nonwords (450ms). There was also a significant interaction between Length and Grade $F(1,56) = 8.24$, $p = .006$ in that the magnitude of the difference between short and long RTs for Grade 2 stimuli (383ms) was significantly reduced for Grade 5 (337ms). There was no significant interaction for Lexicality and Grade $F(1,56) = 2.86$, $p<.097$.

H2: How does the accuracy of Word/ Nonword naming across conditions reflect the development of reading skills in Turkish children?

Thus far reaction time, as an index of reading ability, have been shown to develop rapidly in Turkish children in that the time taken to process words is vastly reduced in older children. It is also imperative to explore the function of accuracy as a measure of reading ability. To this end, accuracy, as measured by percentage correct, was investigated between Grades. Similar to the investigation of naming speed, a 3 x 2 ANOVA, using frequency (high-frequency words, low-frequency words and pseudowords) and length (short words and long words) as within-subject factors and Grade as a between-subjects factor, showed significant main effects of frequency, $F(2,55) = 53.46$, $p<.001$ and length, $F(1,56) = 61.35$, $p<.001$ as well as a significant interaction between them, $F(2,55) = 24.80$, $p<.001$. As suspected from the near ceiling effects of accuracy scores, there were no significant interactions between length and grade, $F(1,56) = 0.53$, $p =.472$ as well as frequency and grade $F(2,55) = .37$, $p = .696$.

What is the relationship between the outcome and predictor variables?

Correlation analysis was conducted to explore the relationship between the outcome and predictor variables. Before doing so, cognitive constructs, that had more than one measure, were explored and, where appropriate, were clustered into a single component using Principal Component Analysis (varimax rotation with Kaiser normalization). There were three such constructs: RAN, PA and memory. Overall, a strong degree of inter-correlation was found amongst the four different stimulus types for RAN performance, which is illustrated in Table 24.

TABLE 24: CORRELATION OF MEASURES OF RAN

Correlations				
	RAN_Ob	RAN_Col	RAN_Lett	RAN_Numb
RAN_Ob	-			
RAN_Col	.613**	-		
RAN_Lett	.601**	.610**	-	
RAN_Numb	.520**	.642**	.760**	-
**** Correlation is significant at the 0.01 level (2-tailed).**				

A single component was extracted with an eigenvalue of 2.88 and accounting for 72% of the variance of RAN measures. With regards to PA, a strong degree of inter-correlation was also established amongst the three different stimulus types for PA performance which is illustrated in Table 25.

TABLE 25: CORRELATION OF MEASURES OF PA.

Correlations			
	Phon_Del	Phon_Seg	Spoonerism
Phon_Del	-		
Phon_Seg	.532**	-	
Spoonerism	.510**	.664**	-
**** Correlation is significant at the 0.01 level (2-tailed).**			

A single component was extracted with an eigenvalue of 2.14 and accounting for 71% of the variance of PA measures. Finally, with regards to memory, a strong correlation was found between the measures of visual short-term memory and working memory but not with phonological short-term memory which is illustrated in Table 26.

TABLE 26: CORRELATIONS OF MEASURES OF SHORT-TERM AND WORKING MEMORY.

Correlations				
	PSTM	WM	VSTM	
PSTM	-			
WM	0.176	-		
VSTM	0.048	.348**	-	
**** Correlation is significant at the 0.01 level (2-tailed).**				

Taking the above into consideration, it was decided that factor analysis would be conducted on WM and VSTM with PSTM being considered as a separate measure. The resulting component had an eigenvalue of 1.35 and accounted for 67% of the variance between the two measures.

Following this, two correlation analyses were conducted: one for naming accuracy and one for naming speed which were shown in Table 27. Correlation analysis of naming accuracy only revealed one significant correlation. Long low frequency words correlated moderately with PA, $r(58) = .397$, $p < .003$.

TABLE 27: CORRELATIONS OF NAMING SPEED AND COGNITIVE MEASURES. ** CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED). * CORRELATION IS SIGNIFICANT AT THE 0.05 LEVEL (2-TAILED).

	SLF	SHF	LLF	LHF	SNW	LNW	RCPM	PSTM	VA Span	RAN	PA	Memory
SLF	1											
SHF	.865**	1										
LLF	.885**	.836**	1									
LHF	.858**	.926**	.836**	1								
SNW	.897**	.906**	.858**	.883**	1							
LNW	.797**	.781**	.829**	.819**	.844**	1						
RCPM	-0.25	-0.119	-0.138	-0.123	-.269*	-0.151	1					
PSTM	0.051	-0.081	-0.071	-0.113	-0.068	-0.066	-0.022	1				
VA Span	-.304*	-0.24	-.291*	-.289*	-.336**	-.319*	.330*	0.145	1			
RAN	.576**	.638**	.573**	.645**	.636**	.702**	-0.198	-0.179	-.459**	1		
PA	-.418**	-.302*	-.467**	-.413**	-.388**	-.462**	.417**	0.147	.424**	-.466**	1	
Memory	-.495**	-.543**	-.435**	-.577**	-.600**	-.476**	.428**	0.136	.382**	-.543**	.454**	1

Briefly, correlation analysis revealed strong correlations between the six measures of naming speed. With the general exception of PSTM and nonverbal IQ, all of the outcome measures correlated with all of the predictor variables. The specific exception was that short high-frequency words did not show a correlation with VA Span, $r(58) = -.24$, $p = .07$.

Can reaction time data be used in conjunction with measures of cognitive skills to identify developmental dyslexia in Turkish-speaking children?

Using an arbitrary cut-off point of 1.25 SD (Landerl et al., 2013) , the reaction time data were explored for potential candidates who were either poor readers or developmental dyslexics. To this end, the cut off points for each condition were SLF: 1126, SHF: 956, LLF: 1481, LHF: 1295, SNW: 1190 and LNW: 1734 for Grade 2 and SLF: 912, SHF: 779, LLF: 1185, LHF: 1038, SNW: 980 and LNW: 1476 for Grade 5.

For Grade 2, 8 children were slower than the 1.25 cut-off in at least one naming condition, 3 children were slower in at least 2 conditions and 2 children were slower in at least 3 conditions. The three children in Grade 2 identified by being slower in word naming in at least 2 conditions were then further subjected to investigation of their cognitive skills in comparison to their age-matched peers. It was decided to approach the task of identifying predictors of developmental dyslexia by comparing the three "at-risk" children to the rest of the Grade 2 cohort. An ANCOVA, controlling for the covariation of nonverbal IQ and SES, revealed that working memory was the only significant predictor of "at-risk" status, $F(1,26) = 10.20$, $p = 0.005$. For Grade 5, 8 children were slower than the 1.25 cut-off in at least one naming condition, 3 children were slower in at least 2 conditions and 2 children were slower in at least 3 conditions. The three children in Grade 5 identified by being slower in word naming in at least 2 conditions were then further subjected to investigation of their cognitive skills in comparison to their age-matched peers. No quantitative difference was found between the "at-risk" children and their age-matched peers.

INTERIM DISCUSSION

The over-arching aim of the pilot study was to establish the adequacy of the newly created battery of tests and methodological design for measuring the development of reading skills in Turkish-speaking children. In addition, the data collected from 58 Turkish-speaking monolingual children aged 7 to 8 and 10 to 11 has yielded some interesting insights regarding the development of reading processes in a bi-directionally transparent language. The overarching aim of the researcher's doctoral thesis is to explore the Turkish language processing skills of children who are in the process of acquiring the ability to read with skill. The present pilot study is suggestive of the unique contribution of both specific factors involved in visual word recognition (Lexicality, Length and Frequency effects) as well as more general automation of cognitive functioning during the process of learning to read in Turkish. A brief account of the findings and their potential relation to the wider domain are discussed below.

Reliability

An initial goal of the pilot research was to identify possible sources of bias and inconsistent measurement at the item level prior to administering the final version of the newly created battery. While there has been a vast growth in the development of measurement tools available to assess foundation learning and literacy, the practice of adapting tests with demonstrated psychometric properties in Western contexts is prevalent amongst researchers investigating lesser-studied languages (Nag, 2017). One of the main criticisms of such an approach is that there is often a lack of rigour and care in the adaptation of such tests as well as a lack of widely available, open-source materials used by a wide range of researchers in any given language.

In the context of the current pilot study, all newly developed tests were (according to Cronbach's α coefficients) highly reliable, with the exception of the PA- Deletion measure. Furthermore, the measures developed were considered to be highly suitable for the purpose of this thesis (though see below for a discussion of limitations and future work).

Word Reading Accuracy and Speed

Word reading accuracy was high across conditions irrespective of grade. The exceptions to this finding were the significant difference between grades on the Long, Low-Frequency word condition and the overall poorer performance by both groups on the long nonword measure. These ceiling effect findings are largely in line with the literature concerning reading development in Turkish. Overall, evidence from the pilot study advocates that Turkish readers rapidly reach ceiling-level reading accuracy, which is expected given the shallow orthography of Turkish (Durgunoğlu & Öney, 1999; Öney & Durgunoğlu, 1997). This is also in agreement with previous outcomes in the adult literature (I.Raman et al., 2004) as well as the results from transparent orthographies such as Italian, Finnish, and Greek (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988; Holopainen, Ahonen, & Lyytinen, 2002; Porpodas, 1999). Additionally, the above findings appear to corroborate the position that reading speed is a superior index of reading than accuracy in transparent orthographies. From the above, it appears that hypothesis 2 can be accepted regarding ceiling effects for accuracy data, particularly with regards to single word reading.

Regarding word reading speed, all six of the word/nonword conditions were named significantly faster by children in Grade 5 than in Grade 2. The general direction (from fastest to slowest) of the word/nonword naming conditions, independent of Grade, was: SHF> SLF>SNW>LHF>LLF>LNW. In addition, single word RTs of the older children became faster and less sensitive to word length and frequency effects as a function of age as indicated by the significant interaction of both length by grade and Frequency by Grade. Similar to findings in other transparent languages (e.g., Acha, Laka, and Perea, (2010), the findings above are thought to signify that word naming is easier as evidenced by a decreasing word frequency effect in older readers. The robust word frequency effect found in beginning readers is thought to denote the development and storage of words in the mental lexicon and is thought to be reduced as older children are exposed to more words. However, this effect was small in older children, who were able to effortlessly recognize both frequent and infrequent words.

Lexicality, Length and Frequency Effects in Turkish children

The present pilot investigation propositions the availability of both lexical and sublexical reading strategies at early stages of reading acquisition in Turkish as word length, a sublexical variable, and frequency, a lexical variable, affected both Grades 2 and 5. Evidence from Turkish adults suggests that users of transparent orthographies make full use of the lexical route in naming words aloud and that under conditions which demand the use of the nonlexical route, they shift strategies (I.Raman et al., 2004). Further support for the flexible use of both the lexical and sublexical pathways in transparent orthographies comes from research in Italian (Pagliuca, Arduino, Barca, & Burani, 2008). Furthermore, as stated previously, reading strategies are thought to be partially dependent on the transparency/consistency of the orthographic system.

To the best of the researcher's knowledge, this is the first evidence for the presence of lexical variables from the early stages in Turkish reading development. Similar evidence has been found in Italian (Burani, Marcolini, & Stella, 2002) , Japanese (Sambai et al., 2012) and Spanish (Avdyli, Castejón, & Cuetos, 2014) . Furthermore, the supposition of lexical and sublexical strategies for reading forms the very basis of the Dual Route Cascaded model of Reading Aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). In contrast, connectionist frameworks offer an alternative account of observed frequency effects stipulating that the connection weights between the distributed representational units, i.e., orthography, phonology and semantics gradually strengthen to reduce model error (Seidenberg & McClelland, 1989) . Finally, the observed interaction between length and lexicality asserts that length effects were larger for orthographically unfamiliar word forms, i.e., pseudowords compared to words (Weekes, 1997). Thus, as the recognition of pseudowords is considered to reflect a serial, sublexical procedure and the recognition of words is presumed to reflect a parallel lexical procedure, these findings lend support to dual-routes of visual word recognition (Coltheart et al., 2001; Perry, Ziegler, & Zorzi, 2007; 2013) as well as the availability of both routes to Turkish primary school children learning to read. From the above, it appears that hypothesis 1 can be accepted regarding the presence of length and frequency effects as well as hypothesis 3 regarding the positive finding of a length by lexicality interaction.

Cognitive measures of visual word recognition

A separate goal of the pilot study was to assess the role of different cognitive skills in learning to read in Turkish, a highly transparent orthography. The first studies of

Turkish phonological awareness successfully established the rapid development of decoding skills in emergent readers (Durgunoğlu & Öney., 1999; Öney & Goldman, 1984) . In line with this, the results from this pilot investigation advocate that phonological awareness contributes to word recognition in the early stages of reading as expressed by ceiling effects for both word reading accuracy and phonological awareness even in children as young as 7 years old. Furthermore, the relationship between phonological awareness and reading speed measures in Turkish was weak to moderate (r ranging from -.302 to -.418). Likewise, the role of RAN, which has been suggested to be central to reading in transparent orthographies, was explored through a correlational analysis and found that the relationship between RAN and reading speed measures in Turkish were moderate to strong (r ranging from .573 to .702). In a similar manner of analysis, the relationship between reading speed measures and VA Span was weak (r ranging from -.240 to -.336), for VSSTM was moderate (r ranging from -.395 to -.538), for WM was weak to moderate (r ranging from -.251 to -.483). The findings above warrant further investigation of the role of these cognitive factors in reading in Turkish.

Learning to read in languages with transparent orthographies is thought to begin with alphabetic decoding though due to the high degree of consistency of these languages, this phase is transitory, and children rapidly start to develop a lexical procedure, as revealed in this pilot and previous investigations of reading development in transparent orthographies. In principle, this change is not necessary since nearly all words can be identified by the use of the sublexical route. Conversely, by making use of this strategy, reading becomes more fluent and rapid. Since the correspondence rules are straightforward in Turkish, early alphabetic decoding is acquired effectively by practically all children learning to read Turkish. Accordingly, Turkish second graders who have only formally been learning to read for a year can read slowly but accurately.

4.5.4 Limitations and future improvements to study design

While considering the aim to establish the utility of the newly created battery of tests and methodological design, the pilot study was successful in its objective though several limitations must be acknowledged. These limitations, along with future developments to the battery and study design, will be discussed below. Given the limited sample size, the preliminary findings of this study should be considered as a foundation for further research as the sample size restricted the types of analyses that could be carried out. In particular, whilst relationships were established between predictor and outcome variables, multiple regression analysis would allow a more refined evaluation of the relationships found in the pilot study. Additionally, some of the findings in the pilot study require replication in order for the results to be generalised to the wider population.

This issue of upscaling leads to the second limitation: Whilst SuperLab 5 proved to be a useful experiment presentation tool in general, the relatively large number of participants (10%) that were removed from the pilot due to technical issues was alarming. The root cause of the technical issues was the SV-1 voice-activated key which required voice-level adjustment for each child, a largely subjective occurrence, and in the researcher's opinion, nevertheless manifested with an unacceptably large number of false triggers (53) in naming (RT<350ms) as well as documenting 177 (5%) incidences of a Non Response. Furthermore, given the need to upscale in terms of participants and factoring in the cost of SuperLab licenses and SV-1 voice-keys, it was deemed unsuitable to carry on using SuperLab the experiment presentation software for this and similar studies in this thesis. In light of this, a decision was made to migrate to DMDX 5.1, an experiment presentation control software (Forster & Forster, 2003) .

Another limitation was that of recruitment strategy; whilst it was decided that for the pilot study children at the beginning and end of primary education should be recruited, this approach was less than optimal for several reasons. The primary criterion for inclusion into the study was that of grade, though within the Turkish educational system there is considerable variability regarding age in grades, as progress to the next year is dependent on adequate progress. Moreover, the random nature of recruitment for the pilot study proved problematic regarding the balancing of demographic information, particularly for SES and gender. It is evident that given a significantly larger number of participants will be recruited for the main study; a more systematic approach needs to be taken regarding participant recruitment. In order to overcome the potential limitations of the grade/ age issue above, an equal number of children from Grades 2, 3, 4 and 5 will be recruited for the main study and instead of using grade as a between-subjects factor, age in months will be calculated instead.

With regards to the battery itself, accuracy scores, as measured by percentage correct, proved to be overwhelmingly redundant due to ceiling effects. In order to overcome this, accuracy in the main study will be reported as an error rate. Also, PSTM as a construct and measure proved to be unhelpful regarding both the investigation of reading development across grades and as a predictor variable of reading outcomes. The null findings of PSTM are difficult to reconcile, and possible explanations of this finding will be considered in the discussion. The null finding, coupled with the demand to reduce the length of the experiment (see below), it was decided to remove the measure of PSTM from the battery. Though the majority of children completed the experiment with no difficulties, there was a significant majority (mainly younger children) that systematically complained about the length of the experimental procedure. In order to avoid discouraging participation as well as ensuring that performance was not compromised by fatigue, several changes to the experimental procedure were proposed:

i) Raven's CPM would be removed from the battery and instead would be carried out in small groups as has been done previously in the literature (Jerman, Reynolds, & Swanson, 2012; Uno, Wydell, Haruhara, Kaneko, & Shinya, 2009) .

ii) Several of the constructs with multiple measures would be reduced: For PA, the deletion task was removed as nearly all children hit ceiling and was deemed too easy of a task. Furthermore, the number of items in the segmentation and spoonerism tasks would be reduced; for RAN, the high correlation between the four measures allowed for its reduction. Upon review of the literature, it was decided to keep the two alphanumeric measures (Letters and Numbers) as they were named faster than the non-alphanumeric conditions and that the alphanumeric measures are more strongly related to reading skills (Bowey, 2005).

iii) As stated above, PSTM, as measured by the forward digit span task, would be removed from the battery.

It was also decided that due to the above modifications, there was room for the addition of a new outcome measure, Reading Fluency, which would not significantly increase the total time taken to complete the battery as it is frequently measured as a one-minute task. An overview of the reading fluency literature, as well as a solid justification of its usefulness in the current thesis, is given in the next section.

Reading Fluency

Reading fluency may be conceptualized as an index of effective word recognition skills in order to extract meaning from text. Indeed, on the surface, oral reading fluency is often cited to be a direct quantification of phonological segmentation and recoding skills (Fuchs, Fuchs, Hosp, & Jenkins, 2001) . Furthermore, the ability to read connected text fluently is considered to be an essential aspect of successful reading comprehension. Of particular interest to this present thesis is the notion that fluency is said to consist of properties regarding accuracy in word decoding, automaticity, and prosody (Hudson, Lane, & Pullen, 2005) . Though a comprehensive discussion regarding prosody remains beyond the limits of this thesis, accuracy and automaticity lie at the very centre of the current investigation. Thus, it seems logical that measuring reading fluency within the frameworks previously defined would ultimately be complementary.

For instance, Vaessen and colleagues (2010) conducted a cross-sectional comparison of French, Dutch, and Hungarian primary school children and found that reading fluency was predicted by phoneme awareness, letter knowledge, and RAN independent of which language the children use. The authors further state that the cognitive development of reading skill, as measured by reading fluency, appears to be universal in nature. Additionally, the manifestation of reading fluency difficulties (Eklund, Torppa, & Lyytinen, 2013) has been explored with regards to developmental dyslexia (Eklund, Torppa, & Lyytinen, 2013). Recently the relationship between reading fluency and proficiency in reading has been investigated among fifth-grade Turkish-speaking children (Yildirim & Rasinski, 2014) . Three hundred ninety-nine fifth-grade students were tested on measures of reading comprehension, word recognition and reading fluency. The authors reported that word recognition and fluency correlated significantly with reading comprehension. Evidently, whilst the relationship between reading fluency and comprehension is well defined, there is a need to further explore the relationship between measures of word recognition and reading fluency in Turkish-speaking children.

Development of measures of Oral Reading Fluency

It was decided that Oral reading fluency (ORF) was to be measured with a one-minute grade-appropriate passage reading task. Though this approach would not allow for the comparison of reading fluency across grades, it would act as an intra-grade measure of reading ability. Additionally, the methodological limitations of selecting a passage that would be equally difficult across age groups would be avoided. Short passages were selected from textbooks administered by the Turkish Ministry of Education, which are tabulated in Table 28.

TABLE 28: SUMMARY OF TEXTS SELECTED FOR AGE-APPROPRIATE MEASURES OF ORAL READING FLUENCY.

Grade	Age-appropriate text selected (English translation)	Length (Words)	Difference	Cumulative Difference
2	Üç Kelebek "Three Butterflies"	181	-	-
3	Uçan İlk Adam "The First Man to Fly"	190	9	9
4	Küçük Limon Ağacı "The Small Lemon Tree"	214	24	33
5	Şeker Dede "Sweet Granddad"	218	4	37

Children were asked to read the short passage aloud as quickly and accurately as possible for 1 minute. Oral reading fluency was calculated by computing the total number of words read during the minute (WPM), subtracting incorrectly read words, and consequently calculating the number of words read correctly per minute (WCPM) (Hasbrouck & Tindall, 2006).

4.6 THE MAIN STUDY

4.6.1 REVISED METHODS

Before carrying out the main study, sample size calculations in the form of a priori statistical power analyses were carried out using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2009). In order to investigate potential small-to-medium effect sizes, i.e. 0.3 (Cohen, 1977) for lexicality, length and frequency conditions as well as exploring linear regression models for the main study, power analysis revealed that a total sample size of 132 was needed to maintain power at 0.95 (with up to 10 predictor variables).

Participants

Twelve (12) schools were contacted regarding the research project of which four (4) expressed an interest in taking part. Following this, 320 students were invited to take part in the study (80 from each school; 20 from each year group per school). Of these, 153 (~48%) students and their families gave assent/consent to take part in the study. During data collection, two students were removed due to technical difficulties and finally, six students failed/refused to complete the individual session.

Therefore, participants were 145 second, third, fourth and fifth-grade children from four mainstream primary schools in Famagusta, Cyprus. At the time of testing, the mean ages of the children were 96 months (SD = 5.52 months) in Grade 2, 106 months (SD = 4.77 months) in Grade 3, 118 months (SD = 3.52 months) in Grade 4 and 131 months (SD = 4.25 months) in Grade 5. All participants reported having normal or corrected-to-normal vision and were all native Turkish-speakers. This study was approved by the Brunel Psychology Research Ethics Committee and by the collaborating institution, namely the Eastern Mediterranean University's Psychology Department in Cyprus. Permission to approach schools was approved by the Turkish Republic of Northern Cyprus Ministry of Education. All psycholinguistic testing was carried out in the children's schools during school hours.

Following data collection, a further 15 participants were removed from the dataset as suspected developmental dyslexics for further analyses in Chapter 5. The final participant number for the current study was, therefore, 130. Table 29: shows the characteristics of the main study demographics by Grade.

TABLE 29: CHARACTERISTICS OF MAIN STUDY DEMOGRAPHICS BY GRADE

	Grade 2		Grade 3		Grade 4		Grade 5	
N	28		32		31		39	
Sex [% girls]	61		44		48		64	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (Months)	99	8.3	116.5	6.01	127.1	6.91	137.4	6.96
SES (1-4)	2.3	1.26	2.2	1.1	2.68	1.16	2.2	0.9

Procedure

Following the suggested revisions from the pilot study, the data collection procedure is as follows. With the exception of data collection for Raven's Coloured Progressive Matrices, which was carried out in small groups, children were assessed individually in a quiet room in the children's school and lasted approximately 1 hour. Similar to the pilot study, the battery of tasks was presented as a fully computerized task with the exception of the Visuo-spatial short-term memory task in which children were required to write answers in a booklet with blank grids.

Each session followed a fixed order of presentation beginning with Visual Short-term Memory, followed by RAN, VA Span, Single Word/Psuedoword naming, Digit span Backward, PA and ended with the newly created ORF (Oral Reading Fluency) task. Children were informed that they could take brief breaks between tasks but not during tasks. The exception to this was the single word/pseudoword reading task in which a break was built into the task design after children had reached the midpoint (40 words/ pseudowords) in order to avoid fatigue.

Stimuli were presented on a laptop using DMDX v5.1 (Forster & Forster, 2003). In addition, all of the children received the same instructions, which were displayed on the screen and reinforced orally. In order to simulate the natural conditions of individual reading on self-teaching, participants did not receive feedback on their responses, nor were they corrected if they misread the pseudo-words (Álvarez-Cañizo, Suárez-Coalla, & Cuetos, 2018). Finally, children's responses were recorded in WAV format using DMDX and analysed with CheckVocal software (Protopapas, 2007) to calculate the number of correct responses and reaction times (RTs).

Statistical Analysis Plan

The recent move toward the use of linear mixed methods (LMM) provides an opportunity to analyse all the available word-level RT data without the reliance on averaging across participants or items. LMMs permit the separation of the effects of the predictor variables (fixed effects) from the differences on performance among participants and items (random effects) (Baayen, Davidson, & Bates, 2008). For the above reason, the single word reading data was analysed using linear mixed-effects models for continuous variables, i.e. naming speed and generalized mixed-effects models for binary variables, i.e. accuracy (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), with the lme4 package (version 1.1-12; Bates, Mächler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2016). In addition, in order to adjust the skew in its distribution when carrying out mixed-effect model analysis (Baayen, Davidson, & Bates, 2008), all RTs were transformed into inverse RTs (−1000/RT), in line with the recommendation by Brysbaert and Stevens (2018). The data were transformed back to raw RTs for ease of interpretation. Furthermore, the lmerTest package (version 2.0-30; Kuznetsova, Brockhoff, & Christensen, 2015) was used to calculate *p* values using Satterthwaite approximations to determine degrees of freedom. Following the recommendations of Barr, Levy, Scheepers, and Tily (2013), fixed effects in models will include both random effects terms between participants or items (random intercepts) and differences between participants or items in the slopes of the effects of the predictor variables (random slopes). The likelihood ratio test (LRT; Barr et al., 2013; Pinheiro & Bates, 2000) was used to evaluate whether the inclusion of fixed or random effects was warranted in order to select the model of best fit (Luke, 2017). As a final note of consideration, lme4 uses general-purpose nonlinear optimizers (e.g. Nelder-Mead or Powell's BOBYQA method) to examine the variance-covariance structures of the random effects though the calculation of convergence is challenging (Bates et al., 2015). In an attempt to address this issue (as well as mitigate against convergence issues) within the current study, during the RT analysis, the final models were all re-fitted with a range of optimizers using the allFit function in lme4. Where non-default optimizers provided a better model fit for the data, the optimizer used was also reported. Finally, it should be emphasised that the present study design, with concurrent measurement of all predictor variables, does not permit any causal interpretation of these effects.

Experimental Hypotheses

H1: As in the pilot study, word length and frequency effects will be evident for typically developing children in so far as longer words and less frequent words will take longer to name. This is thought to be reflective of both lexical and sublexical processes.

H2: As found in the pilot study, word reading accuracy will be close to ceiling level. Word reading speed will be the only word reading measure that differentiates between good and poor readers (Hasko, Groth, Bruder, Bartling, & Schulte-Körne, 2013) .

H3: In typically developing Turkish children, it is expected that there will also be a lexicality effect positing that words will be named faster than nonwords. This being said, it is also reasonable to postulate that the transparent nature of Turkish will also allow for the processing of nonwords with relative ease. In addition to this, a lexicality by length effect would be expected, stipulating the availability of two routes of reading aloud in children learning to read a highly transparent orthography.

H4: Regarding the cognitive predictors of both ORF and single word/pseudoword reading, it is anticipated that all of the measures used in the main study would have a significant effect on ORF and SWR measures.

H5: Based on findings from previous studies, it is expected that there will be a heterogeneous spread (large variance) in the manifestation of visual attention span and memory tasks that reflect individual differences in these underlying abilities.

4.6.2 RESULTS

Oral Reading Fluency

Descriptive

Table 30 below shows the characteristics of the oral reading fluency task as measured by an age (grade) appropriate one-minute text reading task. In order to explore the potential application of a measure across grades, the mean syllable length and mean syllables per word were calculated.

A non-parametric Friedman test of differences among the different grade-based ORF measures was conducted and rendered a Chi-square value of 6.12, which was nonsignificant ($p=.106$). Therefore, Correct Syllables per minute (SyllCPM) was considered to be a valid index of reading fluency across grades. The results of the Pearson correlation between WCPM and SyllCPM denoted that there was a very strong significant positive association between the two measures, ($r (128) = .99$, $p < .0001$)

TABLE 30: CHARACTERISTICS OF THE ORAL READING FLUENCY TASK

	Grade 2	Grade 3	Grade 4	Grade 5
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
mean Word Length	6.23 (2.83)	6.66 (2.99)	6.21 (2.89)	6.22 (2.32)
mean Syllable Length	2.64 (1.20)	2.82 (1.23)	2.61 (1.23)	2.67 (1.05)
mean Syll per word	2.36	2.36	2.38	2.33
ORF - WCPM	58.64 (15.74)	63.47 (16.66)	84.97 (20.84)	89.10 (26.27)
ORF - SyllCPM	158.96 (48.31)	167.31 (47.89)	215.90 (52.08)	245.87 (64.80)

Predicters of Oral Reading Fluency

In order to evaluate predictors of oral reading fluency across grades, it was decided to use the newly created SyllCPM measure in a linear regression analysis. Before statistical analysis took place, the two phonological awareness task scores (Segmentation and Spoonerism) were combined for each child and similarly a combined score was calculated for the two RAN tasks (RAN–Letter and RAN–Digit) by averaging raw scores in order to increase the predictive validity of phonological awareness and RAN. The correlations between the combined tasks were $r = .357$ ($p < .001$) between the phonological awareness tasks and $r = .637$ ($p < .001$) between the RAN tasks. Following this, correlation analysis revealed significant correlations between the dependent, i.e., SyllCPM and the cognitive predictors of interest in this study (See Table 31). With the exception of RAN (which was negatively correlated to SyllCPM), the predictor variables were all positively correlated to each other and the oral reading fluency measure. The simultaneous use of these measures in a regression analysis can be problematic (see Belsley et al., 2005). Therefore, multicollinearity was checked using bivariate correlations (Tabachnick & Fidell, 2007) and tolerance values in the regression output. In line with the recommendation by Tabachnick & Fidell, (2007), Table 31 indicates that none of the independent variables were too highly correlated to be run in the multiple regressions (< .70).

TABLE 31: CORRELATIONS BETWEEN **SYLLCPM AND THE COGNITIVE PREDICTORS**

	ORF_Syll	RCPM	Grade	PA	VSTM	WM	RAN
RCPM	.358**						
Grade	.545**	.341**					
PA	.625**	.319**	.427**				
VSTM	.458**	.487**	.522**	.500**			
WM	.329**	.229*	.259**	.372**	.301**		
RAN	-.670**	-.193*	-.556**	-.543**	-.461**	-.324**	
VA Span	.553**	.293**	.287**	.376**	.369**	.348**	-.453**
****. Correlation is significant at the 0.01 level (2-tailed).**							
***. Correlation is significant at the 0.05 level (2-tailed).**							

A multiple regression analysis was carried out in which all predictor variables (RCPM, Grade, PA, VSTM, WM, RAN and VA Span) were entered simultaneously. The recommendation to have at least 10 cases per independent variable (Hair, Babin, Anderson, & Tatham, 2005) was well satisfied. The results of the regression analysis are shown in Table 32 and indicated that the selected predictors explained 64.6% of the variance (R^2=.645, F (7, 110) = 31.46, p<.001). The individual predictors were examined further and indicated that Grade (β = 2.72, p < .01), PA (β = 4.42, p < .001), RAN (β = -4.02, p < .001) and VA Span (β = 3.29, p < .001) were all significant predictors in the model. The four significant predictors were then re-entered into a hierarchical regression model in the following order: Grade, PA, RAN and VA Span. The results are summarised in Table 33 below

TABLE 32: REGRESSION COEFFICIENTS OF ORAL READING FLUENCY

	Unstandardized Coefficients		Standardized Coefficients
	B	Std. Error	Beta
RCPM	0.679	0.104	1.607
Grade	4.266	0.198	2.722 **
PA	0.676	0.317	4.421 ***
VSTM	1.367	-0.090	-1.193
WM	5.157	0.016	0.259
RAN	0.926	-0.313	-4.018 ***
VA Span	0.253	0.219	3.293 ***
* p < .05. ** p < .01. *** p < .001.			

TABLE 33: HIERARCHICAL REGRESSION COEFFICIENTS OF ORAL READING FLUENCY

		B	Std. Error	Beta
1	Grade	31.397	4.270	0.545 ***
2	Grade	19.599	4.059	0.340 ***
	PA	4.328	0.636	0.479 ***
3	Grade	10.850	4.119	0.188 **
	PA	3.035	0.639	0.336 ***
	RAN	-4.612	0.927	-0.383 ***
4	Grade	10.583	3.883	0.184 **
	PA	2.614	0.611	0.290 ***
	RAN	-3.542	0.912	-0.294 ***
	VA Span	0.971	0.237	0.258 ***

The unique variance (Adjusted R^2) accounted for by the predictors in each subsequent model while controlling for variance explained by previous predictors was Grade (29.1%), PA (17.8%), RAN (8.2%) and VA Span (4.9%).

In a further investigation of the age-based influence of cognitive predictors on ORF, the Grade variable was recoded into Younger (Grade 2 and 3) and Older (Grade 4 and 5) groups. A new linear regression model with all cognitive variables (except VSSTM) and their relative interactions with the newly created age variable was created. Table 34 illustrates the outcomes of the new model.

TABLE 34: COGNITIVE PREDICTORS OF ORAL READING FLUENCY

Oral Reading Fluency			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	58.50	54.18 – 62.81	**<0.001**
Age	24.70	18.47 – 30.94	**<0.001**
PA	0.49	0.41 – 0.57	**<0.001**
RAN	-0.94	-1.03 – -0.85	**<0.001**
VAS	0.47	0.44 – 0.50	**<0.001**
WM	-0.75	-1.30 – -0.20	**0.007**
Age: PA	0.88	0.78 – 0.99	**<0.001**
Age: RAN	-1.11	-1.25 – -0.97	**<0.001**
Age: VAS	-0.14	-0.18 – -0.10	**<0.001**
Age: WM	0.72	-0.06 – 1.51	0.071
R^2 / adjusted R^2	0.653 / 0.653		

Post-hoc analyses of the significant interactions were conducted in order to further explore them. To perform the post-hoc analyses, the lsmeans package (Lenth, 2016) was used and examined the cognitive predictor effects across age groups using the Tukey adjustment. First, regarding the Age * PA significant interaction, the effect of PA on ORF increased between younger and older children (Younger: $Estimate_{PA}$ = 0.49, *SE* = 0.04; Older: $Estimate_{PA}$ = 1.38, *SE* = 0.03). Second, the analysis of the Age * RAN significant interaction showed that the effect of RAN was lower in younger ($Estimate_{RAN}$ = -0.94, SE = 0.05) than older children ($Estimate_{RAN}$ = -2.05, *SE* = 0.06). Third, the Age * VA span interaction meant that the effect of VA span, despite being significant in both groups, was lower for older children ($Estimate_{VAS}$ = 0.33, *SE* = 0.013) than younger children ($Estimate_{VAS}$ = 0.47, *SE* = 0.015).

Mediation Analysis

To explore if the effects of working memory and/or visuospatial STM on reading are mediated by VA span, two separate simple mediation analysis was carried out using the R package, mediation (Tingley et al., 2013). The motivation behind this analysis was that the total effects of WM and VSSTM on the ORF measure could be separated into direct and indirect effects. While the direct effect of WM and VSSTM on reading fluency were found to be non-significant, there is a need to further explore how visual attention can itself be limited and how in turn VA span may act as a mediator of the relationship between WM and VSSTM with ORF. VA span, WM and VSSTM scores all significantly varied between grades with the exception of no difference between Grades 2 and 3 for VA span. Considering the distribution of each of the measures: VA span had a range from 12 to 91; WM had a range from 2 to 5, and VSSTM had a range from 2 to 20.

Mediation analysis revealed that the relationship between WM and ORF scores was significantly mediated by children's VA span (total effect: $b = 0.37$, $p = .001$; direct effect: $b = -0.003$, $p = .99$; mediation indirect effect: $b = 0.37$, $p = .001$; 10,000 simulations). Additionally, the relationship between VSSTM and ORF scores was also significantly mediated by children's VA span (total effect: $b = 0.24$, $p = .001$; direct effect: $b = -0.02$, $p = .7$; mediation indirect effect: $b = 0.26$, $p = .001$; 10,000 simulations).

Word Reading

Data extraction and cleaning

For the single word/ pseudoword reading data, a total of 10,400 responses were recorded. Following data collection, the sound spectrograms of the recorded responses were analysed using CheckVocal (Protopapas, 2007) in order to extract corrected accuracy and RT measures. For the analysis of accuracy, all responses were considered. Transversely, for the analysis of RT, only correct responses were considered.

Overall, there were 709 (6.8%) errors in the single word naming data which corresponded to: 106 non-responses (no response, <250ms or >3,000 ms to respond); 88 were word naming errors; and 515 were pseudoword errors. The RTs of the remaining 9,691 correct responses remaining after the above eliminations were then explored for outlier using the same procedure outlined in Chapter 3 for the adult lexical decision data. Briefly, RTs that were determined to be outside of a range of whiskers of a boxplot adjusted for skewed distributions (Hubert & Vandervieran, 2004; 2006; 2008) were also removed from the dataset. This data cleaning approach was calculated independently for each participant, separately for words and nonwords and led to the removal of 513 RT data points from the data resulting in a final dataset of 9173 correct responses for further analysis.

Descriptive

The mean overall response time and accuracy across conditions and participants was 1077ms and 93%, respectively. Considering the lexicality factor, response time and accuracy for words was 971ms and 98%, respectively whereas response time and accuracy for nonwords was 1193ms and 88%, respectively. Considering the length factor, response time, and accuracy for short words/pseudowords was 1003ms and 96%, respectively whereas response time and accuracy for long words/ nonwords was 1156ms and 90%, respectively. Considering the frequency factor, response time and accuracy for high-frequency words was 957ms and 98% respectively whereas response time and accuracy for low-frequency words was 985ms and 98% respectively. Table 35 and 36 below provides a summary of the descriptive statistics for accuracy and RT data, respectively.

TABLE 35: MEAN ACCURACY FOR WORDS AND NONWORDS BY GRADE

	Grade 2	Grade 3	Grade 4	Grade 5
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Short Low Frequency	99 (11)	100 (6)	99 (10)	98 (13)
Short High Frequency	99 (11)	97(17)	99 (8)	99 (11)
Short Psuedoword	94 (23)	94 (23)	93 (25)	94 (24)
Long Low Frequency	93 (26)	99 (9)	97 (17)	97 (17)
Long High Frequency	97 (16)	98 (14)	96 (19)	99 (9)
Long Pseudoword	76 (43)	85 (36)	85 (36)	85 (36)

As can be seen from Table 35, reading accuracy across participants was at near celling level. As stated above, the accuracy of responses to words/ pseudowords was conducted by using Generalized-Mixed effects Modelling (GLMM). As an extension of multiple logistic regression, it allows for the evaluation of the log odds that a response would be accurate given a set of predictors. In addition, LMMs can be used to factor in random effects of both subjects and items. In line with similar research (e.g. Ricketts et al., 2016), the approach for the GLMM analysis involved conducting pairwise LRT comparisons (Pinheiro & Bates, 2000) of simpler models with more complex models, where each step of model comparison involves the former model building on the latter in order to determine the value of including various fixed and/ or random effects in the models of single word/ pseudoword reading accuracy.

Reading Speed

TABLE 36: MEAN RT FOR WORDS BY GRADE

	Grade 2	Grade 3	Grade 4	Grade 5
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Short Low Frequency	1034 (202)	949 (175)	908 (153)	817 (116)
Short High Frequency	1042 (215)	935 (155)	896 (160)	799 (114)
Short Psuedoword	1290 (415)	1116 (202)	1099 (332)	979 (188)
Long Low Frequency	1246 (386)	1090 (192)	1002 (307)	922 (157)
Long High Frequency	1217 (403)	1047 (197)	983 (269)	930 (190)
Long Pseudoword	1398 (215)	1430 (285)	1319 (376)	1219 (271)

Accuracy – Length, Lexicality and Grade

In the analysis of accuracy, first, a model (Empty model) that included participants and items as random factors and intercept as a fixed factor was tested:

Accuracy ~ 1 + (1|Subject) + (1|Item)

In the next step, Grade (2,3,4,5), Length (short vs long) and Lexicality (words vs pseudowords) were added as fixed factors.

Accuracy ~ Grade + Length + Lexicality + (1|Subject) + (1|Item)

It was found that the addition of the above, fixed effects significantly improved model fit (LRT, χ^2 = 125.56, 4 df, p < .001). In this main-effects model, there were significant effects of length and lexicality (p < .001) but not of grade (p = .102). The third step was carried out by adding interaction terms. In contrast with the main effects model, the new model also included the length by lexicality interaction and subsequently improved model fit (LRT, χ^2 = 7.17, 2 df, p = .028). The length by lexicality interaction was significant (p = .027). In order to explore further interactions, length by grade and lexicality by grade interaction terms were added one at a time to the model. The addition of the length by grade interaction term also improved model fit (LRT, χ^2 = 9.32, 1 df, p < .001) whereas the addition of the lexicality by grade interaction did not further improve model fit (p=1). Three-way interactions were not explored due to the inherent difficulty in interpreting the results. Therefore, the final model was:

Accuracy ~ Grade + Length + Lexicality + (Length * Lexicality) + (Length * Grade) + (1 | Item) + (1 | Subject)

As a penultimate step in selecting the model of best fit, a series of comparison models that varied on random effects were explored. First, a model with only the random effect of participants on intercepts was compared to the model above and found that the addition of participants on intercepts improved model fit (LRT, χ^2 = 577.61, 1 df, p < .001). Similarly, when a model with only the random effect of items on intercepts was compared to the full model above it was found that the inclusion of a random effect of items on intercepts improved model fit (LRT, χ^2 = 21.02, 1 df, p < .001).

Finally, in order to explore “maximal” models (Barr et al., 2013), the inclusion of random slopes was examined. Random slopes highlight random differences between participants or items regarding the fixed effects. Consequently, random slope terms analogous to random effects of participant differences on the slope of grade and random effects of item differences on the slopes of both length and lexicality were added to the model and did significantly improve the model (LRT, χ^2 = 42.52, 7 df, p < .001). Table 37 below offers a summary of the final model.

TABLE 37: SUMMARY TABLE OF THE FINAL GLMM MODEL OF WORD/PSEUDOWORD READING ACCURACY.

Fixed effects	**Estimated coefficient**	**SE**	**z**	**p**
(Intercept)	4.12	0.48	8.60	**<.001**
Grade	0.03	0.12	0.23	.82
Length	-1.76	0.45	-3.88	**<.001**
Lexicality	1.49	0.30	5.00	**<.001**
Length by Lexicality	0.79	0.31	2.51	**<.010**
Length by Grade	0.23	0.11	1.98	**<.050**
Random effects		**SD**		
Due to items				
Intercepts		0.49		
Grade		0.03		
Due to participants				
Intercepts		0.64		
Length		0.83		
Lexicality		0.81		

Note. Number of observations: 10400; 80 items; 130 participants. The marginal R^2 of the final model is 0.23, and the conditional R^2 of the final model is 0.38.

From Table 37, the estimated coefficients for the final model show that reading accuracy was higher for short words over long words as indicated by the length effect. There was also a significant effect of Lexicality, indicating that real words were named faster than pseudowords. Additionally, the model revealed a Length by Lexicality interaction. The interaction effect indicates that length effects were more evident for pseudowords rather than words. Of note, in the final model, the effect of grade was non-significant, however, there was a significant interaction of length by grade (Figure 13).

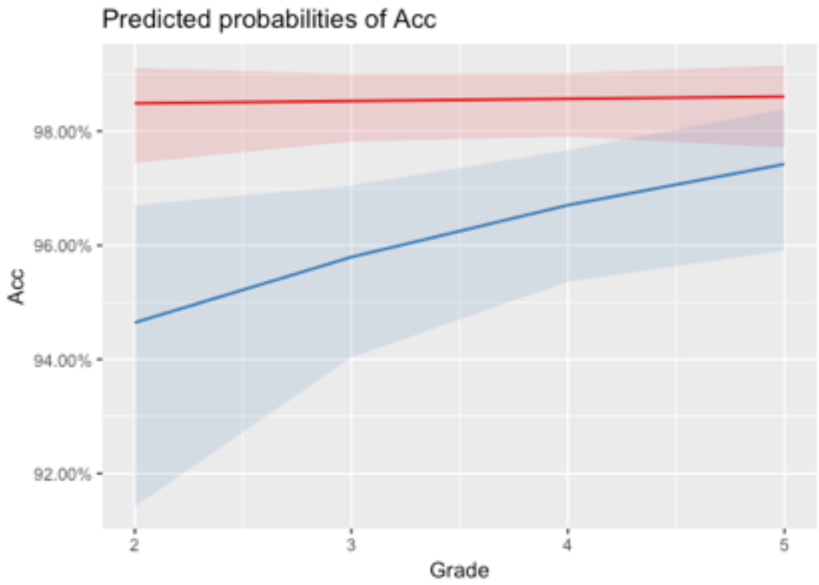

FIGURE 15: PREDICTOR PROBABILITIES OF ACCURACY HIGHLIGHTING THE LENGTH BY GRADE INTERACTION. RED DENOTES SHORT WORDS/ PSEUDOWORDS AND BLUE DENOTES LONG WORDS/ PSUEDOWORDS.

Accuracy – Length, Frequency and Grade

In order to explore the effects of word frequency, a separate (yet identical) analysis of participants' word reading accuracy (n=5200) was carried out. In this analysis, lexical frequency was added as a fixed effect predictor, and the lexicality term used in the above analysis was removed. The final model (including random intercepts and slopes) for the accuracy of responses to words found no significant fixed effects or interactions perhaps reflective of the ceiling effects of word naming in Turkish-speaking children.

Naming Speed - Length, Lexicality and Grade

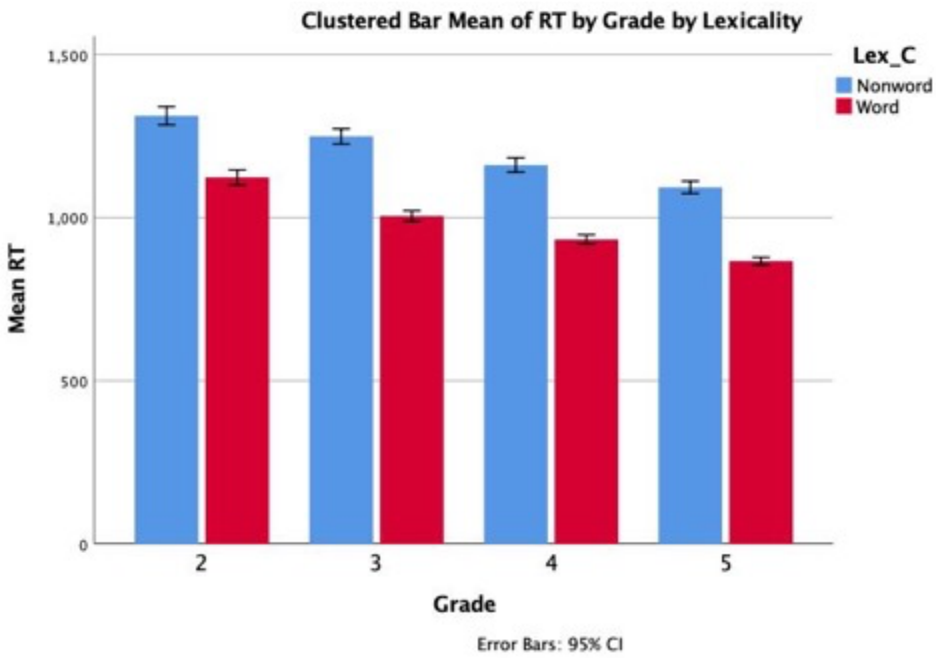

FIGURE 16: CLUSTERED BAR CHART OF MEAN REACTION TIMES BY GRADE AND LEXICALITY

Following on from the analysis of the accuracy data, the same model selection procedure was applied to the RT data considering only correct responses (n=9178). Figure 14 suggests that RT to both words and pseudowords decrease as a function of grade (and by extension age) suggesting an overall development of word reading automaticity as the cohort get older. The effects of Grade, Length and Lexicality were explored using a Linear Mixed-effects model. The final formula for the best model fit was:

Inverse RT ~ Grade + Length + Lexicality + (Length * Lexicality) + (Length*Grade) + (Lexicality*Grade) + (Lexicality | Subject) + (Length | Subject) + (1| Item)

Table 38 below offers a summary of the final model. From Table 38, the estimated coefficients for the final model show that single word/pseudoword RTs were greater for short words over long words as well as words being named faster than pseudowords. There was also a significant effect of Grade. Additionally, the model discovered a significant Length by Lexicality interaction. The interaction effect indicates that length effects were more evident for pseudowords rather than words. There was also a Lexicality by Grade effect indicating that lexicality effects were more evident in younger rather than older children. The marginal R^2 of the final model is 0.311 (variance explained by the main effects), and the conditional R^2 of the final model is 0.529 (variance explained by the entire model, including both fixed and random effects).

TABLE 38: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD/PSEUDOWORD READING RT.

Fixed effects	**Estimated coefficient**	**SE**	**T**	**p**
(Intercept)	-0.79	0.07	-11.73	**<.001**
Grade	-0.06	0.02	-4.17	**<.001**
Length	-0.09	0.03	10.33	**<.001**
Lexicality	0.07	0.03	-2.43	**.037**
Length by Lexicality	-0.06	0.02	-2.49	**.015**
Lexicality by Grade	-0.02	0.01	-3.79	**<.001**
Length by Grade	0.01	0.01	1.71	0.09
Random effects	**Variance**	**SD**		
Due to items				
Intercepts	0.002	0.05		
Due to participants				
Intercepts	.02	0.10		
Length	.003	.05		
Lexicality	.005	.07		

Note: Convergence issue resolved with the use of bobyqa (optimx) optimizer with 300000 iterations.

Naming Speed - Length, Frequency and Grade

In addition, to explore the effects of word frequency, a separate (yet identical) analysis of participants' word reading RTs was carried out. In this analysis, lexical frequency was added as a fixed effect predictor, and the lexicality term used in the above analysis was removed. The final model had the equation:

InverseRT ~ Frequency + Grade + Length + (Length | Subject) + (Grade| Item)

TABLE 39: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RT.

	Dependent variable:
	Inverse RT
Frequency	-0.004
	(0.005)
Grade	-0.072^{***}
	(0.016)
Length	0.095^{***}
	(0.014)
Constant	-0.843^{***}
	(0.068)
Observations	4,819
Log-Likelihood	1,065.581
Akaike Inf. Crit.	-2,113.163
Bayesian Inf. Crit.	-2,054.344
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Convergence issue resolved with the use of bobyqa (optimx) optimizer with 200000 iterations.

From table 39 ,as above, the estimated coefficients for the final model show that single word RTs were faster for short words over long words as well as words being named faster as a function of age. However, the frequency effect was found to be non-significant after length was added to the model. The marginal R^2 of the final model is 0.111 (variance explained by the main effects), and the conditional R^2 of the final model is 0.604 (variance explained by the entire model, including both fixed and random effects).

As a post-hoc follow up to explore the non-significant finding of a frequency effect, the above model was rerun with several theoretical and statistical considerations. It is currently understood that transformations of RT data such as the one used in this study lead to distorted interval differences in the DV (Lo and Andrews, 2015). In order to explore this issue further, the distributional properties of the RT data were further explored using a number of non-Gaussian (non-normal) distributional assumptions for the RT data; namely Gamma and Inverse Gaussian distributions with both variants including an identity and inverse link function. Based on the comparison above, the model was refitted using an Inverse Gaussian distribution applied to the raw RT data. The new model has the formula:

RT~ Length + Frequency + Grade + (Length*Frequency) + (Length*Grade) + (Length|Subject) + (Frequency|Subject)

From Table 40 below, it is evident that the pattern of main effect findings is similar between the two LMM models of word reading in that the estimated coefficients for the new model show that single word RTs were faster for short words over long words as well as words being named faster as a function of age. The frequency effect remained non-significant after length was added to the model. However, the new model also found a significant interaction effect of Frequency by Length and of Length by Grade. Further investigation into the interaction effects established that frequency effects were more evident for long rather than short words. Similarly, Length by Grade interaction indicated that length effects were more evident in younger rather than older children. The marginal R^2 of the final model is 0.705 (variance explained by the main effects), and the conditional R^2 of the final model is 1.00 (variance explained by the entire model, including both fixed and random effects).

TABLE 40: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RT WITH MODIFICATIONS TO RT

Fixed effects	**Estimated coefficient**	**SE**	**T**	**p**
(Intercept)	1236.14	17.81	69.39	**<.001**
Grade	-56.47	6.59	-8.57	**<.001**
Length	199.65	15.26	13.08	**<.001**
Frequency	-4.90	3.01	-1.63	.103
Length by Frequency	-12.23	3.87	-3.17	**.002**
Length by Grade	-15.21	4.82	-3.16	**.002**
Random effects	**Variance**	**SD**		
Due to participants				
Intercepts	181.9	13.49		
Length	5870	76.61		
Frequency	249.7	15.80		

Cognitive Predictors of Word/Pseudoword Reading

In a final analysis, the cognitive predictors of single word/pseudoword reading were considered. First, both words and pseudowords were considered together for both accuracy and RT measures. Following this, a separate analysis was carried out for words and pseudowords.

Word/Pseudoword Reading Accuracy

The same model selection procedure used in the previous analyses was applied to both the accuracy and RT data. The effects of Grade, PA, RAN, VA Span, RCPM, WM and VSSTM on accuracy scores were explored using a Generalised Linear Mixed-effects model. The final formula for the best model fit was:

Accuracy ~ PA + VA Span + (PA | Subject) + (1 | Item)

TABLE 41: SUMMARY TABLE OF THE FINAL GLMM MODEL OF WORD READING ACCURACY USING COGNITIVE PREDICTORS

	Accuracy		
Predictors	**Odds Ratios**	**CI**	**p**
(Intercept)	4.37	1.70 – 11.29	0.002*
PA	1.06	1.02 – 1.10	0.001*
VA Span	1.02	1.00 – 1.03	0.017*
Random Effects			
σ^2	3.29		
Subject	1.87		
Item	1.65		
Subject.PA	0.01		
Subject	-0.74		
ICC _Subject	0.27		
ICC _Item	0.24		
Observations	10400		
Marginal R^2 / Conditional R^2	0.055 / 0.506		

From Table 41, the estimated coefficients for the final model show that reading accuracy was more likely to produce correct responses by students with higher scores in PA and VA Span. No other cognitive predictors or interactions reached significance for inclusion into the final model. Separate analysis for words and pseudowords yielded the same final model and as such will not be reported further.

Word/Pseudoword Reading Speed

The same model selection procedure used in the previous analyses was applied to the RT data. The effects of Grade, PA, RAN, VA Span, RCPM, WM and VSSTM on RT scores were explored using a Linear Mixed-effects model. The final formula for the best model fit was:

InverseRT ~ PA + RAN + VA Span + (Grade | Item) + (PA + RAN + VA Span | Subject)

From Table 42 below, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with higher scores in VA Span and PA and lower scores in RAN.

TABLE 42: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD/ PSUEDOWORD READING RT USING COGNITIVE PREDICTORS

Inverse RT			
Predictors	**Estimates**	**CI**	**p**
(Intercept)	-1.04	-1.30 – -0.77	**<0.001** ***
PA	-0.005	-0.01 – -0.00	**0.049** *
RAN	0.01	0.00 – 0.02	**0.006*****
VA Span	-0.002	-0.00 – -0.00	**0.029** *
Random Effects			
σ^2	0.03		
Subject	0.03		
Item	0.01		
ICC _Subject	0.50		
ICC _Item	0.08		
Observations	9176		
Marginal R^2 / Conditional R^2	0.131 / 0.563		

Note: Convergence issue resolved with the use of bobyqa (optimx) optimizer with 200000 iterations.

Word Reading Speed

Considering the effects of Grade, PA, RAN, VA Span, RCPM, WM and VSSTM on word reading speed scores using a Linear Mixed-effects model. The final formula for the best model fit was:

InverseRT ~ Grade + RAN + VA Span + (1 | Subject) + (1 | Item)

From Table 43 below, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with higher scores in RAN and VA Span but not PA. Additionally, there was a near significant finding that speed decreased as a function of age in that older children were faster than younger children in word naming. No significant interaction effects were found.

TABLE 43: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RT

Inverse RT			
Predictors	**Estimates**	**CI**	**p**
(Intercept)	-1.16	-1.48 – -0.84	**<0.001** ***
Grade	-0.03	-0.06 – 0.00	0.057
RAN	0.01	0.00 – 0.02	**0.001*****
VA Span	-0.00	-0.00 – -0.00	**0.038 ***
Random Effects			
σ^2	0.03		
Subject	0.03		
Item	0.00		
ICC _Subject	0.45		
ICC _Item	0.06		
Observations	4819		
Marginal R^2 / Conditional R^2	0.150 / 0.587		

Nonword Reading Speed

Considering the effects of Grade, PA, RAN, VA Span, RCPM, WM and VSSTM on pseudoword reading speed scores using a Linear Mixed-effects model. The final formula for the best model fit was:

InverseRT ~ RAN + VA Span + PA + (Grade | Item) + (RAN | Subject)

From Table 44 below, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with higher scores in RAN and VA Span and PA. No significant interaction effects were found.

TABLE 44: SUMMARY TABLE OF THE FINAL LMM MODEL OF NONWORD READING RT

	Inverse RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.91	-1.18 – -0.64	**<0.001**
RAN	0.01	0.00 – 0.02	**0.010**
VA Span	-0.00	-0.00 – -0.00	**0.013**
PA	-0.00	-0.01 – -0.00	**0.030**
Random Effects			
σ^2	0.03		
$\tau_{00\ Subject}$	0.21		
$\tau_{00\ Item}$	0.00		
$\tau_{11\ Subject.RAN}$	0.00		
$\tau_{11\ Item.GradeNew}$	0.00		
$\rho_{01\ Subject}$	-0.97		
$\rho_{01\ Item}$	1.00		
ICC $_{Subject}$	0.87		
ICC $_{Item}$	0.01		
Observations	4357		

4.7 Discussion

The current chapter explored the factors that influence reading aloud in typically developing readers in Turkish, a highly transparent orthography. Reading was considered at the word and text levels. As predictors, the effect of length, lexicality and frequency (and their interactions) were explored as well as considering the influence of a set of cognitive predictors that have been indicated to influence word and text reading such as phonological awareness, rapid naming, and visual attention span.

With reference to the experimental hypotheses of this study, the current investigation found supporting evidence for length effects as well as some evidence for frequency effects resulting in a general acceptance of H1. Additionally, word reading accuracy was found to be high across participants (H2) as well as reporting a lexicality main effect and a lexicality by length interaction effect (H3). Regarding cognitive predictors of reading in Turkish-speaking children (H4), LME model analysis reported direct effects of Grade, PA, RAN and VA span while also finding an indirect influence of WM and VSSTM measures on ORF via a mediation analysis. Finally, regarding the hypothesis that there would be a large variance in VA span and memory tasks (H5), there was partial evidence to support this position. The below sections will offer a further critical interpretation of the results in light of the existent literature.

Oral Reading Fluency

The decision to include measures of Oral Reading Fluency within the current study appears to be justified in that the measures developed were able to differentiate good from poor readers using a single index. In addition, the growth of oral reading fluency as measured by correct words and syllables per minute across grades is reflective of the general reading proficiency of the current cohort and the ability to read connected text fluently is advocated to be an essential requirement for adequate reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001).

When considering the cognitive predictors of ORF, the current study found a significant effect of PA, RAN and VA Span. Concerning PA's influence on ORF, the findings of this study indicate that increased phonological awareness skills predict reading fluency irrespective of age and as such offer support for the view that phonological processes continue to contribute to effective word recognition even in fluent readers (e.g. Rayner et al. 2012). Furthermore, this finding lends support to the view that PA may be a significant universal contributor to ORF (Katzir, Schiff, & Kim, 2012). Also, the contribution of phonological processing to word recognition may have an indirect influence on oral reading fluency (Torgesen, Rashotte, & Alexander, 2001). Considering the findings of the younger vs older children comparison, PA's role in ORF is sustained over and above its role in single-word recognition.

When considering the influence of RAN on ORF, the findings of the current study are broadly in line with the literature in that RAN is considered to be a robust predictor of ORF (Christo & Davis, 2008; Papadopoulos, Spanoudis, & Georgiou, 2016). Furthermore, the parallel development of reading automaticity, as measured by RAN and ORF may be indicative of shared mental resources, i.e. domain-general factors such as serial processing and articulation (see Georgiou, Aro, Liao, & Parrila, 2016 for an in-depth discussion). In line with this study, both Babayiğit and Stainthorp (2011) and Karadağ, Keskin and Arı (2019) report that RAN was a significant predictor of fluent reading in Turkish children. Beyond Turkish, there is growing evidence that RAN predicts reading fluency equivalently well across languages (e.g. Georgiou, Aro, Liao, & Parrila, 2016). In addition, considering the findings of the younger vs older children comparison, RAN's role in ORF is sustained over and above its role in single-word recognition.

When considering the influence of Visual Attention Span on ORF, the findings of a unique and significant influence of VA Span on ORF are primarily in line with previous studies that have reported that VA Span correlates with oral text reading speed (Lobier, Dubois, & Valdois, 2013). Besides, there are several accounts of the unique contribution of VA Span to oral reading (Bosse & Valdois, 2009; Valdois et al., 2003). However, a recent study carried out by van den Boer, van Bergen & de Jong (2014) found that although VA span correlated equally strongly with measures of oral and silent reading, VA Span made a significant unique contribution to silent reading exclusively. In light of the current findings, it may be reasonable to postulate that the agglutinative nature of Turkish requires the rapid development of VA Span in order to facilitate the correct reading of increasingly long words. The current findings signified that ORF is a reliable and valid measure of reading skill in Turkish grades 2–5 and is straight forward and quick to administer. As such, ORF measures may contribute to the early identification of Turkish students at risk for reading difficulties. This position will be further investigated in Chapter 5. Furthermore, considering the findings of the younger vs older children comparison, VA span's role in ORF appears to diminish for older students. That is, for Turkish, it appears that the contribution of VA span to reading beyond the single-word level may be time-limited. It is feasible to postulate that this may be due to a language-specific feature of Turkish such as clearly defined syllable boundaries or agglutination or even a combination of both. In transitioning from sublexical to lexical reading strategies, the need to develop a VA span beyond three would serve little advantage for Turkish children learning to read. Further research is needed to elucidate the exact nature of these findings.

With regard to the mediation analysis, both WM and VSSTM capacity predicted VA span and as such, could be central limiting factors in VA span performance. Consequently, children with smaller reported WM and VSSTM capacities would have fewer available memory slots to store encoded letters than those with a larger WM and VSSTM capacities. Accordingly, they report fewer letters and their total VA Span score is lower. This finding is largely in line with the Theory of Visual Attention (TVA) (Bundesen, 1990).

In addition, the current study contributes to the growing literature that VA span can operate within and beyond the single-word level (Chen, Schneps, Masyn, & Thomson, 2016). Overall, these findings provide support for H4 and H5 regarding the influence of several cognitive predictors on ORF as well as the large variance observed in VA span and memory measures.

Word/ Pseudoword reading

The current study also considered single word and pseudoword recognition in addition to the manipulation of length and frequency. In general, the study determined the presence of lexicality, length, and frequency effects indicating the use of both a sub-lexical reading route (length effect) and a lexical route (lexicality and frequency effect). These results are mostly coherent with previous reports on Turkish adults (I.Raman, 2006). Furthermore, the collective evidence for a lexicality effect with specific evidence for a frequency effect establishes the influence of lexical knowledge on reading development in Turkish.

When assessing word/ pseudoword reading accuracy, the current study found that irrespective of age, reading accuracy was near ceiling. As reported in the pilot investigation, the ceiling effect findings are in line with the literature concerning reading development in Turkish (Durgunoğlu & Öney, 1999; Öney & Durgunoğlu, 1997) and other transparent orthographies such as Italian (Burani et al., 2002), Finnish (Holopainen, Ahonen, & Lyytinen, 2002), Greek (Porpodas, 1999) and Spanish (Davies et al., 2013). Taken together, the results of this study extend the predominantly European alphabetic findings of the influence of orthographic transparency on reading development. From the above, it appears that hypothesis 2 can be accepted regarding ceiling effects for accuracy data, particularly with regards to single word reading.

Further to this, there was evidence of a length and lexicality effects as well as a length by lexicality interaction on reading accuracy data. The smaller effect of length on words than on pseudowords implies that both lexical and sublexical procedures were available to Turkish children in the naming task. While the main effect of grade was found to be non-significant, there was a significant interaction of length by grade which indicated that regardless of lexical status, older children were better than younger children regarding reading long letter strings accurately while there was no difference reading short letter strings accurately.

When considering the influence of grade, length and frequency on word reading accuracy, the ceiling effects of word naming accuracy in Turkish-speaking children found no significant fixed effects or interactions. Additionally, the above finding appears to corroborate the position that reading speed is a superior index of reading than accuracy in transparent orthographies.

Reading speed, as indexed by RT in this study, was then further considered. Regarding the influence of lexical and sublexical factors on RTs, a significant word length and lexicality effect, as well as a significant effect of grade, were revealed. Persistent length effects in transparent orthographies are thought to be reflective of reliance on the sublexical route as this strategy is said to be both fast and efficient (Wydell, Vuorinen, Helenius, & Salmelin, 2003). Similar findings have been reported in other transparent orthographies such as Finnish (Leinonen et al., 2001) and Italian (Paulesu et al., 2000). Furthermore, the current study reported a significant Length by Lexicality interaction that is indicative of length effects being more evident for pseudowords rather than words as well as a Lexicality by Grade effect indicating that lexicality effects were more evident in younger rather than older children. While the Length by Grade interaction failed to reach significance ($p=0.09$), there was a trend which suggested that there was a diminishing effect of word length on RTs as children grew older conceivably insightful of a gradual shift from sublexical to lexical route use.

The analyses of the influence of grade, length and frequency on word RTs discovered several remarkable results. The first iteration of the model found that word RTs were faster for short words over long words as well as words being named faster as a function of age. However, the frequency effect was found to be non-significant. The evidence regarding the influence of lexical frequency on RTs in reading in Turkish, as well as other transparent orthographies, is variable. For example, small but significant frequency effects have been reported in Turkish adults (Raman, Baluch, & Sneddon, 1996). In Spanish children, Valle Arroyo (1989) reported no frequency effect and a significant length effect, similar to the findings of the current study.

Conversely, a recent series of studies by Davies and colleagues (2007; 2013) did report frequency effects in Spanish children. Comparably, in Italian, frequency effects have been reported to be more sizeable in younger children (Barca et al. 2007). However, Burani et al. (2002) stated finding no significant interaction between the effects of frequency and age. Taken together, the results of a null word frequency effect were perplexing, not due to the Turkish adult or Spanish and Italian developmental data but more so since frequency effects were found in the pilot investigation using the same stimuli.

In order to address the seemingly confounding results, a further analysis was carried out using raw rather than transformed RTs in order to address the potential compression of the long end of the RT distribution relative to the short end (Liceralde & Gordon, 2018). The resulting LMM model reported the same main effects as the previous model, i.e. significant length and grade effects and non-significant frequency effects. However, the new model also included significant interactions of Length by Frequency as well as Length by Grade, which substantially increased the variance explained by the fixed effects of the model. Taken together, the interactions stipulate that length effects were greater for infrequent than frequent words. Similarly, the Length by Grade interaction indicated that length effects were more evident in younger rather than older children. The length by frequency interaction has previously been reported in both oral and silent reading tasks in children (Rau, Moll, Snowling, and Landerl, 2015; Tiffin-Richards & Schroeder, 2015). Within the dual-route framework and per the self-teaching hypothesis (Share, 1995), word representations in the mental lexicon arise as a consequence of repeated phonological decoding with a gradual increase in the number of words stored in the lexicon. As Turkish children that are learning to read are anticipated to possess fewer word representations in their lexicons, words are highly likely decoded using the sublexical route, which is serial and sensitive to word length, resulting in more substantial length effects for infrequent words.

Furthermore, the length by grade interaction is indicative that the differences between short and long word stimuli are greater in the lower grades and diminish in the higher grades. The diminishing length effect with increasing experience has previously been reported in Spanish children (Suárez-Coalla, Álvarez-Cañizo, & Cuetos, 2014) and is considered to be indicative of a gradual shift away from sublexical toward lexical processing of words. Taken together the findings reported in the above section provide strong support for length and lexicality effects (H1 and H3) as well as partial support for frequency effects being present in Turkish-speaking children who are learning to read.

Cognitive predictors of words/pseudowords

A separate avenue of research was to explore the influence of several cognitive skills in word and pseudoword reading accuracy and speed in Turkish children. To this end, several LMMs were produced. While considering the influence of cognitive predictors on reading accuracy, PA and VA Span were found to be exclusively influential on reading accuracy.

First, considering the role of PA, there is a large body of evidence within both the cross-linguistic (Caravolas et al., 2012; Moll et al., 2014; Vaessen et al., 2010; Ziegler et al., 2010) and within-language literature of the importance of PA (e.g. de Jong & van der Leij, 2003; Landerl & Wimmer, 2008; Nag & Snowling, 2012; Park & Uno, 2015; Torppa et al., 2010). In reading development, the vast number of studies carried out to date are suggestive that PA may be more significant for reading development in opaque than in transparent orthographies. This being said, there is growing support for the view that the effect of PA in transparent/ consistent orthographies appears to be temporary (Verhagen, Aarnoutse, & Van Leeuwe, 2008; Wimmer & Mayringer, 2002). The first studies of PA in Turkish (Durgunoğlu & Öney., 1999; Öney & Goldman, 1984) recognized the rapid development of decoding skills in emergent readers. In line with this, the results from the current study support the view that phonological awareness contributes to word recognition in the early stages of reading as expressed by ceiling effects for both word reading accuracy and phonological awareness.

Considering the role of VA Span on reading accuracy, the hypothetical role of visual processes on reading acquisition is still under examination. This study lends support to the role of VA Span in reading development in line with previous research carried out in Dutch (van den Boer, van Bergen & de Jong, 2015) and French (Bosse and Valdois, 2009) monolingual children. However, it remains unclear whether the relationship between the VA span and reading changes across languages (Lallier et al., 2018). Taking this into consideration, weaker length effects are associated with larger VA spans in a study involving Dutch children (van den Boer, de Jong, & Haentjens-van Meeteren, 2013) suggesting that increasing VA span may act to facilitate the fast-lexical procedure of reading (Lobier, Dubois, & Valdois, 2013) by permitting more letters at each fixation to be processed. Within the context of Turkish, VA span may similarly accelerate the processing of very long words known to be associated with agglutinative orthographies.

In a final analysis, the influence of several cognitive skills in word and pseudoword reading speed in Turkish children were considered. To this end, several LMMs were produced. While considering the influence of cognitive predictors on reading speed, RAN, PA and VA Span were found to be exclusively influential on reading speed. However, when considering word reading speed only, the effect of PA disappeared. It appears that phonological skills are related to literacy skills that involve decoding as indexed by a significant finding in nonword reading speed and lack of a significant finding in word reading speed. In this study, RAN and VA span has been found to be related to word reading speed.

Considering the role of RAN in word reading speed, RAN was a better predictor of word than of nonword reading. Theoretically, if RAN is held to be a reflection of orthographic processing (Bowers & Newby-Clark, 2002; Manis, Seidenberg, & Doi, 1999), the integration of visual information concerning letter orders in words would make RAN redundant in nonword reading processes as found in this study. Still, the mechanisms underlying the RAN-reading relationship are still not fully understood (Kirby, Georgiou, Martinussen, & Parrila, 2010).

Considering the role of VA Span in word reading speed, the current study found that VA span directly influences single word reading speed. In addition, there was a relationship between VA span abilities and reading skills at all grades. Similar findings have been reported in French children (Bosse and Valdois, 2009).

4.7.1 Limitations

As with any experimental study, there are some limitations of the current study that need to be highlighted. First, the current study was undertaken with children who were attending several different schools from one district, and as participation was voluntary, and the participation rate was moderately low, a selection bias cannot be excluded. The findings of the current study need to be replicated with a more representative sample that moves beyond the singular geographic region covered in this thesis. Second, all of the participants, as a function of the inclusion criteria for this study, had already mastered the alphabetic principle; in order to fully explore earlier relationships between predictors and the Turkish outcome variables, future studies that include younger children are needed. Related to this, as the present study was concurrent, the causal role of how lexical and cognitive factors develop over time need to be addressed using a longitudinal study design that follows students over several years.

The lack of standardized measures (as previously reported in Chapter 3) led to the development of measures for each cognitive skill that was hypothesized to be involved in Turkish reading. While this approach was appropriate for this study, future experimental research could address this limitation following the standardization of the measures developed in this work. Also, the findings of the current study are limited to the specific set of measures used, i.e., it is entirely plausible that a different set of measures (e.g. syllabic manipulation in PA tasks or n-back tasks for working memory) for the constructs measured would yield different results. Future research would benefit from including well-designed measures of morphological awareness in studies exploring reading development in Turkish. Additionally, frequency measures beyond surface frequency such as root frequency may have a differential influence in agglutinative orthographies and warrant further research. The bottom-up approach used for the analysis was one of personal preference if the alternative top-down approach had been adopted, it is feasible that results would be different.

The null findings of PSTM are challenging to reconcile, given that, they could be due to a poorly conceived measure design through digit span tasks are widely used in the literature. Future studies may consider an alternative, purer, measure of PSTM such as a modified version of Gathercole and Baddeley's (1989) Nonword repetition test. Alternatively, the findings could be reflective of ceiling level phonological processing development in Turkish children (see Babayiğit & Strainthrop, 2007). Within the literature, there are mixed findings regarding the role of PSTM. For example, Parrila et al. (2004) and Torgesen et al. (1997) found that when considered along with PA and RAN, PSTM was only weakly associated with reading measures. Conversely, Swanson and colleagues (Swanson & Alexander,1997; Swanson & Howell, 2001) report that the contribution of PSTM to reading was significant. Perhaps it is feasible to postulate that, like the other measures of memory used in the current investigation, PSTM may have a mediating role in Turkish reading. Alternatively, if phonological memory reaches capacity early on in reading development, then the emphasis would switch to the contribution of other cognitive skills. This finding presents an opportunity for further investigation in future studies to shed light on the findings regarding PSTM in this study.

Finally, it is important to consider that while outlier removal is a common approach within the literature, it is not without contention when carried out regarding children's RTs. The approach adopted within this thesis is the use of means and standard deviations which has been identified as being problematic for a number of reasons (Jones, 2019). The first issue is that means and standard deviations are particularly sensitive to extreme values and as such, extreme outliers can often mask less extreme cases. The second issue is that the use of means and standard deviations also assumes symmetry of distribution, which is not the case in reaction time data. With particular relevance to this thesis, children's RT data, is sensitive to outlier removal with the use of absolute cut-offs risking real data being eliminated (Ratcliff, 1993). Therefore the removal of outliers may bias the results (Bakker & Wicherts, 2014). Furthermore, slow RTs may simply be reflective of the developmental nature of word reading. Future studies in reading development that utilise RTs as a measure of interest need to consider alternative approaches such as the median absolute deviation (MAD; Leys et al., 2013) to detect outliers or the use of robust LMM models (Koller, 2016) which require no outlier removal.

4.8 Summary

The findings of this study contribute to a number of topics concerning the underlying cognitive and linguistic mechanisms of reading development in Turkish children. Taken together, for children reading in Turkish, performance appears to be dependent on a mixture of both lexical and sub-lexical knowledge. Furthermore, the results obtained in the current study reveal that phonological awareness, rapid automatized naming and visual attention span differentially influence reading ability.

As reading accuracy reaches ceiling quickly, the focus shifts toward developing reading speed by automating a superior method of decoding and progressively developing the lexical reading route. To this end, VA span appears to play a vital role in decoding speed, conceivably by processing multiple letter-clusters as single units. In addition, RAN is related both to decoding and to sight word reading and may be involved in the essential function of fluently converting visual stimuli into their corresponding phonological representations.

The following chapter will introduce the literature regarding atypical reading development in Turkish in addition to exploring the contribution of several linguistic and cognitive factors to reading development in atypically developing monolingual Turkish children.

CHAPTER 5: DEVELOPMENTAL DYSLEXIA IN TURKISH-SPEAKING CHILDREN

5.1 OVERVIEW

While typical reading development in Turkish has been sufficiently explored in Chapter 4, there is also a need to examine the manifestation of Developmental Dyslexia (DD) in Turkish-speaking children. This chapter will introduce the topic area covering current developments in the domain, explicitly concerning differing theoretical accounts of DD taking into consideration the influence of orthographic transparency on reading development. Subsequently, the chapter will critically review both cognitive predictors and the diverse conceptualisations of DD subtypes. Based on the literature review below, comparisons between DD children and chronological age-matched and younger controls will be reported using both a group and multiple case study approach. The chapter will conclude with a discussion of the findings and potential future developments.

5.2 INTRODUCTION

Developmental dyslexia is characterised as a neurodevelopmental disorder that manifests as a disorder that principally affects the acquisition of literacy skills, in particular, learning to read (Habib, 2000; Peterson & Pennington, 2012; Scerri & Schulte-Körne, 2010). According to the World Health Organization (1993), DD can additionally be described as *"a disorder manifested by difficulty learning to read, despite conventional instruction, adequate intelligence and sociocultural opportunity"* (World Health Organization, ICD-10). However, concerning intelligence, this definition is considered outdated as average intelligence is no longer considered a pre-requisite for DD identification (Siegel & Hurford, 2019). Poor readers with a lower than average IQ have not been found to differ from poor readers with average or above IQ in several literacy and cognitive measures (See Ellis, McDougall & Monk 1996; Vellutino et al., 1996).
According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V, 2013), DD should be identified if reading is behind that expected for the person's age and the

difficulties observed cannot better be accounted for by any other neurological, physical, social or personal condition. This age-based definition will be adopted for this thesis. Globally, DD is the most common learning disorder accounting for approximately 80% of all reported specific learning disorders (Handler et al., 2011).

5.2.1 Theories of Developmental Dyslexia

DD is both particularly difficult to define and currently has an inadequately understood aetiology. There are a large number of conflicting hypotheses for the causes of DD. Many of these competing theories refer to diverse theoretical approaches developed to explain DD from a causal perspective (Reid, 2001). For example, in a seminal study, Bradley and Bryant (1978) reported that poor readers attained lower scores on tasks of rhyme oddity and production than good readers, resulting in the proposal of the Phonological Deficit Hypothesis (PDH) (Bradley & Bryant, 1983). The core ability to read, at least in alphabetic orthographies, is considered to be phonological processing. As such, is reported to be the core deficit in DD (Frith, 1997; Goswami, 2002; Marshall et al., 2001; Snowling, 1998; Snowling and Hulme, 2011; Stanovich, 1988). According to this phonological deficit model, DD results from an inability to break down words into their phonological parts and map each letter to its corresponding sound. In individuals with DD, phonemes are less well defined and are thought to affect processing via both the lexical and the sublexical route. In support of this, a longitudinal study examining the cognitive skills of children with a familial risk of DD (Pennington and Lefly, 2001) found that this cohort performed significantly poorer than low-risk control children on both implicit VSTM (visual short term memory) and RAN (rapid automatized naming), and explicit (PA) phonological processing tasks. However, an outstanding issue within the literature concerns whether PA precedes reading acquisition or is a consequence of learning to read (Morais, Cary, Alegria, & Bertelson, 1979). There is some evidence from two recent extensive longitudinal studies (Landerl et al., 2019: Peterson et al., 2018) that the predictive power of PA for reading may have been overemphasised in previous studies. Additionally, the authors of these studies propose that the contribution of PA to reading development may be less causal and conjecture that PA may function as a corequisite skill for typical reading development.

An extension of the phonological account of reading disorder is the *double-deficit hypothesis* (Wolf & Bowers, 1999; 2000). This hypothesis postulates that in addition to the observed phonological difficulties experienced by dyslexic children, there is a second equally important deficit in naming speed as measured by RAN tasks (Allor, 2002; Denckla & Rudel, 1976b). RAN is defined as the ability to rapidly name highly familiar visual stimuli, such as digits, letter, objects and colours (Wolf & Bowers, 2000). As such, RAN tasks are intended to measure reading speed. Although RAN is frequently reported to be moderately correlated with PA, most studies also report that RAN makes a unique contribution to explaining the variance of reading skill (Georgiou, Parrila, & Liao, 2008). Considering transparent orthographies, reading speed deficits are at the core of DD, and RAN is reported to be a better predictor of reading differences than PA (Guzmán et al., 2004; Meisinger, Bloom, & Hynd, 2010; Serrano & Defior, 2008). The double-deficit hypothesis further stipulates that individuals who have both PA and RAN deficits show more significant reading impairments compared to those with a single deficit. However, there is a conflict in the literature as to the exact nature of RAN; one view posits that RAN tasks index the retrieval speed of phonological information from memory and as such is an aspect of phonological processing, which is coherent with the phonological deficit theory (Snowling, 2000). However, the alternative view is that RAN tasks index processes that are as a minimum partially independent of phonological processing. Evidence in support of this view comes from correlational studies that report a relationship between RAN and reading skills that are independent of measures of PA (Bowers & Wolf, 1993; Wolf & Bowers, 1999). Taken together, it appears that RAN may be the best predictor of reading speed while phonological processing may be the best predictor of reading accuracy and spelling (Moll et al., 2014).

Beyond this broad consensus, the underlying biological and cognitive causes of DD are still contested (Ramus 2003; Démonet et al., 2004; Ramus et al., 2006) and a growing number of studies identify deficits other than phonological. Several alternative hypotheses regard developmental dyslexia as a primarily visual deficit. Firstly, the magnocellular theory of DD (Stein, 1989; 2001) proposes that the dysfunctional frequency and amplitude sensitive Magno-cells negatively impact motion sensitivity (binocular instability) and rapid auditory processing (Ray, Fowler, & Stein, 2005; Sperling, Lu, Manis, & Seidenberg, 2003) resulting in the observed deficits manifested by people with dyslexia. In a similar but distinct line of enquiry, the *visual stress theory* (Wilkins, 2003) stipulates that visual stress leads to distortions of text and headaches when reading and advocates for the use of coloured lenses to reduce visual stress. However, it is not considered to be a specific theoretical theory of dyslexia though some dyslexic children do show marked patterns of visual stress (Ramus et al., 2003). The *visual processing deficit hypothesis* attempts perhaps inadvertently to merge several theories into a coherent framework, but the theory remains contentious at least regarding alphabetic languages (Wang, Bi, Gao, & Wydell, 2010). Finally, and of most interest to this thesis, the *visual attention span deficit hypothesis* (Bosse et al., 2007) has received increased attention as it has been established that phonological awareness and visual attention span measures each made a unique contribution to the reading performance of dyslexic children. They also report that phoneme awareness was responsible for a large amount of variance in pseudoword reading, by extension, adding weight for the argument of the robust impact of phonological processing on reading skills (Ziegler et al., 2008). VA Span impairments are often manifested as a deficit in the ability to recall strings of consonants. However, they can identify consonants in isolation and often have preserved phonological processing abilities (Bosse & Valdois, 2009b). A recent interventional case study on a French-Spanish bilingual dyslexic girl by Valdois and colleagues (2014) found that after specific training in a VA Span task, the dyslexic child reported higher scores on the VA Span tasks. Also, the study reported increased activation in her superior parietal lobes bilaterally; an area thought to be associated with the neural underpinnings of VA Span (Peyrin, Démonet, N'Guyen-Morel, Le Bas, & Valdois, 2011). Based on this finding, the authors concluded in favour of a causal relationship between VA Span and DD, though stated that more extensive studies need to be conducted before a true conclusion could be drawn.

It is also important to highlight that processes involved in WM and STM have also received much attention in children with reading disorders over the last 30 years (Swanson, Cooney, & McNamara, 2004). STM is considered to be essential to phoneme recall, and WM is required for the simultaneous processing and storage of letter sequences. STM deficits are thought to contribute to the manifestation of DD by decreasing the orthographic and phonological information that is needed to be co-activated during reading. Additionally, with reference to reading development, STM is particularly important during the development of phonological recoding, when GPC are not yet fully specified (Gathercole and Baddeley, 1993). According to WM models, WM is controlled by the central executive system in order to manage attentional processing demands and is indirectly involved in phonological processes and naming speed. Poor WM is hypothesised to contribute to poor performance on complex PA tasks leading to a negative impact on word reading ability (see de Jong & van der Leij, 2003). Accordingly, a child with DD may show deficits in both STM and WM processes (Swanson et al., 2009).

The notion that DD could have diverse cognitive characteristics is increasingly acknowledged in the domain (Menghini et al., 2010). It is now becoming evident that DD may best be defined as a multi-faceted disorder and that a combination of factors contributes to its heterogeneous manifestation. It is conjectured that phonological factors (Siegel, 1990; Snowling, 1995; Stanovich, 1996), working memory (Baddeley, 1993; Rack, 1994), visual processing (Stein, 1989; Wilkins, 2003) and processing speed of information (Rack, 1994) all play vital roles in explaining DD (Pneuman, 2009). The following section will consider several approaches that have been created to explore the subtypes of DD. A critical review of these different methodological approaches will be offered before considering some of the theoretical frameworks that have been proposed regarding reading development and disorder in orthographies beyond English.

5.2.2 Subtypes of Developmental Dyslexia

The heterogeneous manifestations of dyslexia are thought to characterise diverse patterns of performance on reading and reading-related cognitive tasks, and as such lend support to the idea that there may be homogenous subtypes within DD populations. There are, at present, numerous theoretical frameworks which are used to characterise and classify the diversity of children with DD (e.g. Coltheart, Curtis, Atkins, & Haller, 1993; Pennington, 2006; Ramus et al., 2003). Moving beyond single-deficit accounts of DD, one of the most influential theoretical models of reading and reading disorder come from the dual-route cascaded (DRC) model of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). While the model was initially developed to explain skilled reading and acquired dyslexia, it has also had a significant influence on the field of DD especially concerning the question of whether there are subtypes of DD (Castles & Coltheart, 1993; Manis, Seidenberg, Doi, McBride-Chang & Peterson, 1996). As mentioned previously, according to the dual-route model, written words are processed by use of either the lexical or sublexical route. According to the DRC model, reading acquisition involves establishing these two partially independent routes. There are two widely used approaches to subtyping DD within the framework of the dual-route framework. Using the so-called classical method, the assumption is that phonological DD is indexed by impaired pseudoword reading but with preserved irregular word reading; Conversely, surface DD manifests with impaired irregular word reading with spared pseudoword reading. A mixed profile of DD is reported when both pseudoword and irregular word reading are impaired. This dissociation was first reported in several adult cases of acquired dyslexia (Newcombe & Marshall, 1973; Holmes, 1973; Shallice, 1981) and forms the cornerstone of evidence for support of the dual-route models of reading (Coltheart, Masterson, Byng, Prior, & Riddoch, 1983; Coltheart, 1987) but has also been replicated in the connectionist triangle model of Seidenberg and McClelland (1989) as well as the multi-trace memory model of Ans et al. (1998). Alternatively, the regression method considers reading disorder as a relative deficit with either orthographic skills relative to phonological skills, or vice versa. Using this approach, DD subtypes may be classified as "soft" as contrasting to the "hard" subtypes defined using the previously mentioned classical method (Stanovich et al., 1997). Soft DD subtypes are established by plotting pseudoword reading performance against irregular-word performance (and vice versa) and then inspecting the 90% or 95% confidence intervals around the regression lines as defined by the control group. A participant is considered to be a phonological dyslexic when they are an outlier once pseudowords are plotted against irregular words but are in the normal range when irregular words are plotted against pseudowords. Equally, a participant is considered to be a surface dyslexic in the opposite direction, i.e., outlier status when irregular words are plotted against pseudowords but normal range when pseudowords are plotted against irregular words. Participants who are outside of the normal range for both regression lines are considered to be mixed profiles. Both of these methods of classifying DD have been widely utilised in the domain across several orthographies that differ in their transparency including the opaque orthographies of English (Castles & Coltheart, 1993; Manis, Seidenberg, Doi, McBride-Chang & Peterson, 1996; Stanovich, Siegel & Gottardo, 1997) and French (Génard, Mousty, Content, Alegria, Leybaert & Morais, 1998; Sprenger-Charolles et al., 2000; Ziegler et al., 2008) as well as more transparent alphabetic orthographies such as Greek (Douklias, Masterson, and Hanley, 2009; Niolaki, Terzopoulos & Masterson, 2014) and Spanish (Jiménez et al., 2009). Interestingly, the above studies that explored DD subtypes in transparent orthographies (Greek and Spanish) used word reading latency rather than irregular word accuracy as a measure of lexical processes. The lack of irregular words in Turkish would also warrant such an approach.

In a converse conceptualisation, the interaction between DD and the DRC model have been proposed so that the DRC model informs DD research (Friedmann & Coltheart, 2016). Within this framework, Developmental Dyslexias are characterised as selective deficits in the different modules or connections of the DRC model. A deficit in each component or connection is proposed to manifest in diverse patterns of reading difficulty, characterised primarily by different error types. Using this framework, Dyslexias are broadly distributed into peripheral Dyslexias which are considered to be reading deficits that stem from the orthographic-visual analysis stage of the model, and alternatively, the central Dyslexias, which stem from reading impairments in the later stages of the lexical and sublexical routes. Using this approach, it has been proposed that there are 19 subtypes of DD (Friedmann & Coltheart, 2016)

Another framework for DD subtyping uses cluster analysis to classify DD subtypes based on the children's differential deficits in distinctive cognitive domains. In an early attempt at using this approach, Lyon and Watson (1981) reported six independent DD subgroups in a cohort of 100 children with DD: Cluster 1 was defined by deficits in language comprehension, auditory memory, sound blending, visual-motor integration, visual-spatial and visual memory skills; Cluster 2 was defined by deficits in language comprehension, auditory memory, and visual-motor integration skills; Cluster 3 manifested with deficiencies in language comprehension and sound blending; Cluster 4 was defined by deficits in visuospatial capacity; Cluster 5 was defined by deficits in verbal, visual memory, and Cluster 6 manifested with no cognitive deficits. This approach has also found some utility in orthographies beyond English such as Chinese (Ho et al., 2004), Dutch (Willems, Jansma, Blomert, Vaessen, 2015), German (Heim et al., 2008), Portuguese (Pacheco et al., 2014), and Spanish (Soriano and Miranda, 2010). For example, the data-driven study by Pacheco and colleagues (2014) profiled children with DD on a wide-ranging array of cognitive abilities, including measures of PA, RAN, Verbal WM and vocabulary. The study found evidence for a cluster with phoneme deletion and RAN deficiencies as well as a cluster with phonological processing difficulties (phoneme deletion and digit span) without a RAN deficit. The authors suggest that the results are best explained by a hybrid perspective, initially proposed by Pennington and colleagues (2012), to explore the heterogeneity of dyslexia further. Adopting a hybrid perspective incorporates all contending models of DD and therefore permits the occurrence of both single and multiple-deficit cognitive profiles. In summary, the significance of characterising the cognitive profiles of Turkish children with DD would contribute to both a better understanding of the aetiology of DD as well as offering a valuable resource in order to develop enhanced tools for the identification and remediation of DD in Turkish-speaking populations.

However, the evidence for discrete subtypes is controversial (e.g. Bryant & Impey, 1986; Stanovich et al., 1997; Sprenger-Charolles et al., 2011). In a recent study, Giofrè et al. (2019) identified seven DD subtypes with cluster analyses in a large Italian cohort. Taking a distinct approach, the authors propose a continuum of phonological–visual impairment in which the position of an individual along the continuum might determine the distinctive features of their cognitive profile. The study found the presence of two distinct clusters of children with DD: Cluster 1 reported a more prominent phonological deficit, while both clusters were found to be impaired in visual processing. Also, a continuum approach has been implemented into the previously discussed surface, and phonological types proposed by the DRC model (see Peterson et al., 2014) as support for distinct aetiologies is limited (Gustafson, Ferreira, & Ronnberg, 2007), and prevailing evidence implies that the two subtypes symbolise two ends of a continuum rather than discrete groups (Castles et al., 1999; Griffiths & Snowling, 2002; Olson et al., 1985).

Another pertinent issue within the DD subtyping literature concerns the use of control groups. The performance of children with DD are typically compared to chronological age (CA) matched controls. Using a CA control group approach, it is hypothesised that the overall portion of mixed profiles is high, and there should be little evidence for dissociated profiles. However, there is a strong case for the inclusion of average readers of the same reading level (RL) controls for several reasons. Firstly, RL controls are necessary to establish if there are differences in vocabulary size and phonemic awareness between people with dyslexia and CA controls (Bryant & Impey, 1986) are a consequence of the lower reading level of people with DD. Second, there is a need to consider the shift in the use of the sublexical and the lexical routes that depend on the overall level of visual word recognition that has developed (Waters, Seidenberg, & Bruck, 1984; Sprenger-Charolles et al., 2000).

Consequently, comparisons of DD participants with both CA or RL controls are based on skills that differ both quantitatively and qualitatively. Additionally, when compared to RL controls, it can be determined if people with DD have a developmental trajectory that is either deviant or delayed. For instance, when people with DD are at the same reading skill level as RL controls, then their developmental trajectory is considered to be delayed. When their phonological or orthographic reading skills are below the level of RL controls, then their developmental trajectory is considered to be deviant.

It is, however, essential to consider that the majority of the theoretical frameworks concerning DD come from studies conducted with English-speakers. The outlier status of the English orthography is well established (Share, 2008) and accumulating evidence from cross-linguistic studies (e.g. Joshi & Aaron, 2006; Seymour, Aro, Erskine, 2003) raises uncertainties regarding if studies of DD conducted in the English orthography can be readily generalised to other orthographies (Ziegler et al., 2003). It is currently understood that the manifestation of DD differs according to the orthographic transparency of a language. In a seminal study, Wydell and Butterworth (1999) reported a case of a well-educated English-Japanese bilingual boy, AS who manifested with monolingual dyslexia in English. The study found that AS's ability to read in Japanese was at an equal or superior level to that of his peers, notwithstanding his severely impaired reading ability in English. He was especially poor at English tasks involving phonological manipulation. To explain this dissociation, Wydell and Butterworth (1999) inaugurated the Hypothesis of Granularity and Transparency (HGT). As mentioned in Chapter-3, the hypothesis suggests that any orthography where the print-to-sound translation, i.e. the transparency dimension is transparent will not produce a high incidence of phonological DD, regardless of the grain size, i.e., phoneme, syllable, and character. In addition to this, they stipulate that any orthography whose smallest orthographic unit representing speech sound is coarse (granular level), i.e., a whole character or whole word, should also not produce a high incidence of phonological DD. Support for the HGT comes from a range of studies that report the estimated prevalence rates of DD across orthographies. For example, use of the different script variants of Japanese, namely syllabic Hiragana, syllabic Katakana and logographic Kanji, report distinct prevalence rates of DD, i.e. 0.2%, 1.6% and 6.9% respectively (Uno et al., 2009). In addition, a higher incidence rate of phonological DD has been reported in English (25–55%) when compared to Spanish (22%) or French (4%) (Castles and Coltheart, 1993, Genard et al., 1998, Jimenez and Ramirez, 2002, Stanovich et al., 1997). Finally, the prevalence of DD in Italian has also been reported to be considerably lower than opaquer and less consistent languages (De Luca, Burani, Paizi, Spinelli, & Zoccolotti, 2010). However, prevalence rates vary widely across studies depending on the exact definition and measures used for identification (Elliott & Grigorenko, 2014). The dimensions of the HGT are illustrated in Figure 15 below.

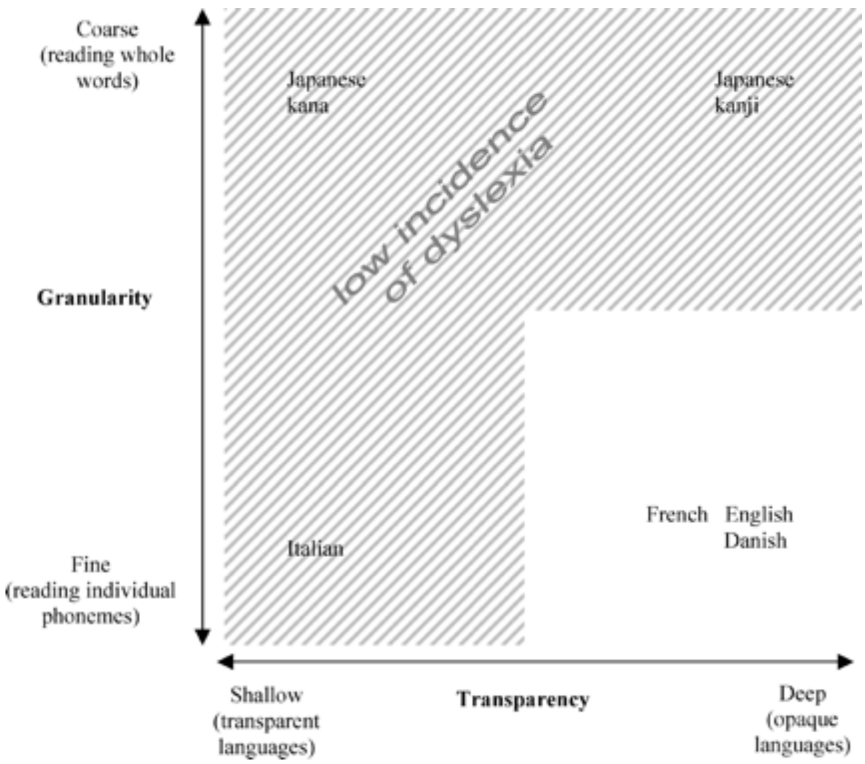

FIGURE 17: THE INCIDENCE OF PHONOLOGICAL DYSLEXIA IN LANGUAGES THAT DIFFER IN TRANSPARENCY AND GRANULARITY. (ADAPTED FROM WYDELL & BUTTERWORTH, 1999)

In contrast, the Psycholinguistic grain size theory (PGST) also mentioned in Chapter-3, proposed by Ziegler and Goswami (2005) does not predict that orthographic transparency and granularity reduces the incidence of DD. They, instead, argue that if reading is established by phonological skills, then children with DD will experience similar difficulties even in consistent orthographies. The PGST further suggests that although the incidence of DD will be similar regardless of the transparency of the orthography, the manifestation of the disorder may differ with orthographic consistency. Additionally, they posit that the incidence of Phonological DD should not be reduced by larger grain sizes, as PA of "*subsyllabic units may still be necessary for the acquisition of the characters or symbols used in coarse grain-size orthographies".*

In summary, the variance among readers with dyslexia is significant; for each proposed causal mechanism, there are readers with dyslexia that show the deficit predicted, but many others that do not. To date, no single-deficit approach has offered a fully comprehensive account for such heterogeneity. In line with this, Menghini and colleagues

(2010) provide support to the view of DD as a multifactorial deficit. Consequently, the multiple deficit perspective of DD suggests that cognitive profiling is critical for a true understanding of DD subtypes. Furthermore, one of the main limitations of causal interpretations of DD is that research looking into the reading disorder has primarily been carried out with English speaking participants. It is, therefore, imperative to consider how the characteristics of writing systems influence the manifestation of DD.

5.2.3 Developmental Dyslexia in Transparent Orthographies

The orthography that children learn to read in has been acknowledged as a critical factor that influences the manifestation of DD. Before exploring the literature regarding Turkish DD, it would be useful to critically evaluate the research literature concerning DD in other transparent alphabetic orthographies such as Greek, Italian, Finnish and Spanish, amongst others. Finnish represents a particularly remarkable orthography to explore further given the orthographic, morphological and phonological similarities with the Turkish writing system. First, the accuracy and latency of visual word recognition will be considered, followed by an overview of specific cognitive deficits associated with DD in transparent orthographies.

The established view within the literature proposes that participants with DD perform both less accurately and slower than age-matched controls on reading tasks (King, Lombardino & Ahmed, 2005; Tressoldi, Lorusso, Brenbati & Donini, 2007; Wimmer, 1993; Wimmer & Goswami, 1994,). As previously mentioned, there is also accumulating evidence that children with DD in transparent orthographies manifest with slow word reading rates (Campton & Carlisle, 1994) and is supported by results in Finnish (Holopainen, Ahonen, & Lyytinen, 2001; Müller & Brady, 2001), Italian (Brizzolara et al., 2006; Tressoldi, Stella, & Faggella, 2001), and Spanish (Jiménez González & Hernández-Valle, 2000). In Finnish, parallel to Turkish, reading accuracy and efficient decoding skills are established early on in reading development, and reading difficulties in Finnish are primarily revealed in tasks of reading fluency and slow single word reading (Aro, Huemer, Heikkilä & Mönkkönen, 2011; Holopainen, Ahonen, & Lyytinen, 2001). Considering this position further, Tressoldi, Stella and Faggella (2001) examined Italian children's text passage reading accuracy and speed and reported that Italian children with DD manifested with difficulties involving speed or automatization of the reading process. In addition, in a seminal study, Porpodas (1999) established that Greek children with DD read 93% of nonwords accurately representing a statistically significant difference when compared to 97.1% nonword accuracy for the typically developing controls. As mentioned in Chapter 4, the transparent mapping of graphemes to phonemes appears to facilitate the development of word decoding skills to the degree that suggests reading accuracy is relatively unaffected in most cases of DD; although see Güven & Friedmann (2019) for a counter-argument related to Letter Position Dyslexia in Turkish.

Further to this, DD affects and is affected by the cognitive skills that underpin reading ability though the exact nature and mechanisms of these cognitive skills remain a matter of considerable debate: It has also been proposed that depending on orthographic transparency, cognitive processes may be differently involved in producing symptoms of DD, (Landerl et al., 2013). Studies carried out in several transparent orthographies suggest that deficits in PA are a universal feature of reading disorder and that PA is central to reading (Goulandris, 2003). Evidence for this comes from studies carried out in Finnish (Puolakanaho et al., 2004), Greek (Porpodas, 1999; Nikolopoulos et al., 2003) and Spanish (Jimenez Gonzalez, Alvarez Gonzalez, Estevez Monzo, & Hernadez-Valle, 2000). Furthermore, a study on the phoneme awareness skills of Czech and English children with DD found that children with DD in grades 3 to 7 manifested with significant and persistent phoneme awareness difficulties irrespective of orthographic consistency (Caravolas, Volín, & Hulme, 2005). Adding weight to this argument, Nikolopoulos et al. (2003) compared the performance of children with DD and typically developing readers at Grades 2 and 4, on a series of PA and cognitive tasks and found that Greek children with DD struggled on complex PA tasks, such as spoonerism and phoneme deletion. However, in a counter position, there is also some evidence that specific tasks that measure PA such as are less demanding for children with DD that are learning to read a transparent orthography (Ziegler & Goswami, 2003). Overall, the discriminatory power of PA measures of reading deficit can be evaluated in terms of task demand in that complex

PA tasks maintain high discriminatory power in studies including children with DD who are learning to read in a transparent orthography (de Jong & van der Leij, 2003).

Beyond phonology, there is growing evidence that, at least in transparent orthographies, RAN appears to be one of the strongest predictors of reading among children with DD (See Landerl & Wimmer, 2008; Torppa et al., 2010). For instance, Brizzolara and colleagues (2006) evaluated naming speed (RAN) and phonological skills in Italian children with DD. The study reported that the majority of children manifested with problems with RAN. In contrast, children who exhibited reduced scores in phonological measures had a history of language delay. The authors concluded that in transparent orthographies, phonological deficits might not accurately represent the primary cognitive marker of DD. In support of this, Holopainen and colleagues (2001) explored the role of phonological awareness, letter knowledge, and naming speed measures in predicting the deficits in reading acquisition observed in Finnish children with DD. The study found that rapid naming speed was the most important predictor of reading differences in Finnish and argued that RAN was a superior measure to PA in predicting at-risk status in reading development. Overall, it appears that PA is an essential universal skill early on in reading acquisition then as children reach ceiling levels in their ability to accurately decode words, there is a shift towards a stronger relationship between RAN and reading. Taken together, orthographic depth is believed to prescribe when this shift occurs with children reading in transparent languages shifting earlier in schooling (Vaessen et al. 2010).

In addition to PA and RAN, there are a number of other cognitive skills that are thought to be involved in the manifestation of DD. Firstly, there is an increased focus within the DD literature concerning the role of visual attention span (e.g., Bosse & Valdois, 2009; Valdois, Bosse, & Tainturier, 2004; van den Boer, de Jong, & Haentjens van Meeteren, 2013). Mentioned previously, Visual attention span is generally conceptualised as the ability to report back briefly presented letter strings and is thought to reflect the number of orthographic units that can be processed in a glance (Valdois et al., 2004). Considering the role of VA Span in DD, evidence from French and English-speaking children with DD (Bosse et al.,2007), suggest that, independent of their phonological skills, children with

DD frequently report with difficulty to simultaneously process multiple elements. Beyond opaque orthographies, there is growing evidence of the role of VA Span in DD concerning transparent orthographies (e.g. Awadh et al., 2016; Valdois et al., 2014). For instance, Germano, Reilhac, Capellini, and Valdois (2014) investigated the manifestation of DD in Brazilian Portuguese (intermediate transparency). Results suggest that both PA and VA span skills contributed independently to reading fluency and that deficits in each of the cognitive abilities define specific subtypes of DD. Furthermore, evidence from the relatively transparent orthography of Dutch (Van Den Boer, Van Bergen, & de Jong, 2015) confirms the contribution of VA span to reading fluency. Interestingly the study controlled for both rapid naming and verbal short-term memory, adding further support for the independent role of VA span to reading development and reading disorder.

Concerning working memory deficits in DD, there is abundant evidence for the role of verbal working memory in processes related to word decoding. Several studies of transparent orthographies have reported differences between typical and poor readers (Dutch: Tilanus, Segers, & Verhoeven, 2013; Greek: Constantinidou, & Evripidou, 2012). Italian: Menghini, Finzi, Carlesimo, & Vicari 2011; Spanish: Jiménez, Rodríguez, & Ramírez, 2009) suggesting that poor readers have difficulty in keeping phonological information in working memory. However, in contrast to verbal working memory, visual-spatial working memory has received little attention in the domain (Provazza, Adams, Giofrè, & Roberts, 2019). Interpreted in terms of the working memory model of Baddeley and Hitch (1974), visual-spatial working memory refers to the visuospatial sketchpad which has limited capacity to represent information in terms of its visual and spatial characteristics (Baddeley, 2000). Visual-spatial memory performance is typically measured by tasks relating to the recall/ recognition of visual patterns (Gathercole & Baddeley, 2014). Concerning the role of visual-spatial memory in DD within transparent orthographies, the evidence is variable. For instance, in support of the role of visual-spatial working memory deficits, Giovagnoli, Vicari, Tomassetti, and Menghini (2016) reported that in a cohort of Italian children with DD, there were significant differences between typically developing and DD subgroups in tasks of mental rotation, visual-spatial memory, global visual-perception and visual-motor integration. However, a longitudinal study carried out with German children with DD (Fischbach, Könen, Rietz & Hasselhorn, 2014) reported no deficits in static visual-spatial working memory though there were significant differences in dynamic measures of visual-spatial working memory. Furthermore, the authors propose that their results may be reflective of compensational effects in that phonological deficits are equalised by strengths in the visual-spatial domain (See von Kàrolyi et al., 2003; Winner, French, Seliger, Ross, & Weber, 2001). Also, the role of WM, generally, and visual-spatial WM, specifically, in agglutinative orthographies such as Turkish has received little attention. The current study will aim to bridge the gap in the literature concerning the role if any, of visual-spatial memory on reading disorder in Turkish.

Considering reading disorder within the framework of dual-route theories, evidence from Italian children with DD suggests a dependence on the sublexical route as indexed by the presence of length effects for both words and pseudowords (De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; Zoccolotti et al., 2005). Similar reports of persistent word length effects have been reported in Dutch (Martens & de Jong 2008; Verhoeven, & Keuning, 2018), Finnish (Hautala et al., 2013) and Spanish (Davies, Rodríguez-Ferreiro, Suárez & Cuetos, 2013). In support of this assertion, a series of studies of Greek children with DD have found distinct surface dyslexia subgroups who read highly familiar words slowly (Douklias, Masterson, Hanley, 2009; Niolaki, Terzopoulos, Masterson, 2014; Sotiropoulos, & Hanley, 2017). However, in another Italian study, word frequency effects were reported in both children with DD and typically developing controls. This has also been reported in Dutch (van der Leij, & van Daal, 1999), Finnish (Hautala et al., 2013) and Spanish (Davies, Rodríguez-Ferreiro, Suárez & Cuetos, 2013). Overall, the above studies conclude that the lexical route is still available to developmental dyslexics of shallow orthographies thus rejecting the idea of an over-reliance on sublexical processing (Barca, Burani, Di Filippo, & Zoccolotti, 2006). It has also been posited that a persistent word length effect may be the prominent behavioural manifestation of dyslexia in shallow orthographies (Hautala, Hyönä, Aro, & Lyytinen, 2011).

Regarding dual-route approaches to DD subtyping, there is growing evidence of the presence of subtype dissociations in transparent alphabetic orthographies. For instance, consistent with the manifestation of surface and phonological subtypes of DD, there have been reported cases in Filipino (Surface: Dulay & Hanley, 2014), Greek (Douklias, Masterson, & Hanley (2009), Italian (Lorusso, Cantiani and Molteni, 2014; Surface: Zoccolotti et al., 1999) and Spanish (Jiménez, Rodríguez, & Ramírez, 2009). These studies appear to corroborate the position that "*individuals with pure surface and phonological dyslexia can be observed in transparent alphabetic orthographies*" (Hanley, 2017, p.14). Additionally, in order to explore the supposition that the underlying impairment in DD that produces slow reading of familiar, regular words in transparent orthographies is parallel to inaccurate reading of irregular words in English, Sotiropoulos, and Hanley (2017) carried out a study of seven Greek-English bilinguals with DD. The study found that the cohort had slow reading of Greek familiar words (surface DD). In addition, the group also reported inaccurate reading of English irregular words. Sotiropoulos and Hanley (2017) argue that the observed co-occurrences provide strong evidence that the underlying impairment seen in the slow reading of real words in Greek is equivalent the impairment that produces inaccurate reading of irregular words in English.

In conclusion, it seems that both surface and phonological subtypes of DD can be readily detected in transparent and opaque alphabetic orthographies. The underlying impairment observed in surface DD appears to be equivalent though with a somewhat different manifestation in transparent and opaque alphabetic writing systems. From a theoretical perspective, research on developmental dyslexia is informative regarding the effect of the properties of the various orthographies on reading. Therefore, the simple GPC mappings in transparent writing systems are likely to cover difficulties observed in DD in opaque orthographies resulting in the manifestation of less severe and distinct deficit patterns by comparison.

5.2.4 Developmental Dyslexia in Turkish

The topic of DD was purportedly first introduced to the Turkish literature in a series of books (Razon, 1976, 1980, 1982). While benefitting from being written in the Turkish language, the books were merely reflective of the theoretical understanding of DD at the time from an Anglocentric perspective and contained no experimental study carried out in Turkish populations. Perhaps the first experimental report of DD in Turkish speakers comes from Vanlı (1988) who investigated literacy errors in children with DD by comparing 18 dyslexic boys aged 7-9 years with 18 typically reading boys matched in terms of IQ, age and SES. The study found that the number of reading and writing errors made by children with dyslexia was significantly higher than those in the control group. Furthermore, the most common mistakes were found to be missing/incorrect letter and syllable production, letter/syllable rotation, and nonword writing. However, the lack of an RL control group and the use of group means for statistical analysis makes both the reliability and interpretability of these results difficult. Similarly, Çapan (1989) reported two case studies of two poor readers with average IQ who made misread and misspelt many words and showed a tendency to omit and substitute suffixes. However, this study suffers from small sample sizes and lack of generalisability that are inherent in single case study designs.

Following these early investigations, there have been several studies examining reading performances of students with DD (Baydık, 2002; Çaycı & Demir, 2006; Gökçe-Sarıpınar & Erden, 2010; Güzel-Özmen, 2005; Karaman, Türkbay, & Gökçe, 2006). For example, in her doctoral thesis, Baydık (2002) conducted, a study examining word reading strategies of Grade 1 students with and without reading difficulties in Turkish. Her thesis reported that despite the orthographic transparency of Turkish, word and pseudoword reading accuracy was lower than expected. Baydık (2002) attributed this finding to the use of holistic approaches in the teaching of reading of the children participating in the research. In similar results, Akyol and Yildiz (2010) report that case study that found that the most frequent error type in a grade 5 student with DD was substitution with insertion leading to a high error rate. In another study examining the reading fluency errors of grade

4 students with DD (Sidekli, 2010) the most frequently reported errors were found to be related to word repetition, syllable skipping and adding letters. Similarly, Akyol and Temur (2006), examined the reading errors of the grade 3 students and reported that the students made the most errors concerning self-correction, syllable/word repetition, adding letters, letter skipping and syllable reading errors. Taken together, these case studies are suggestive that it is critical to also consider children with DD's reading accuracy scores despite the transparent nature of the Turkish orthography.

More recently, Gökçe-Sarıpınar and Erden (2010) carried out a large norm study in which 909 children in grades 1-5 took part. Additionally, the DD group was made up of 64 children, selected based on a formal diagnosis of DD by a state psychiatrist and psychologist. The study reported that children with DD performed worse in reading fluency (words correct per minute) and reading comprehension compared to their typical reading peers. While the study reflects upon an important and understudied area, the exploration of DD in Turkish within this study is mostly descriptive and does not consider the cognitive and linguistic skills involved in reading. In addition, nor the diagnostic criteria used to identify students with DD, nor potential subgroup membership has been stated in this study. Furthermore, the large variability seen between the various texts used for the study (80-269) calls into question the comparability of scores across age groups.

To date, there has been minimal empirical investigation of the existence and extent to which DD is manifested in Turkish. One of the main barriers within this domain is the lack of knowledge and understanding of specific learning disorders in Turkey and Northern Cyprus (Bingöl, 2003; Erden, Kurdoğlu, & Uslu, 2002; Esen & Çiftçi, 1998). Consequently, the number of students that are identified with learning disorders is limited. One of the reported reasons for this is the lack of tools available to identify learning difficulties in Turkish (Arslan & Dirik, 2008; Bingöl, 2003; Erden et al., 2002; Gökçe-Sarıpınar & Erden, 2010). The distinct lack of standardised measurement tools developed to measure reading skills presents a major challenge for gauging the manifestation for DD objectively. Additionally, the lack of measurement tools creates difficulty in identifying which areas children with DD experience difficulties and subsequently in developing suitable intervention programs specific to those areas of difficulty. In addition, educators are poorly informed as to the nature of learning disorders (Esen & Çiftçi, 1998). Additionally, the limited papers that have examined DD in Turkish to date have not made the distinction between different subtypes of DD. The principal exception to this is a case study carried out by I. Raman and Weekes (2005) reporting BRB a Turkish-English bilingual stroke patient who as a consequence of his stroke had an acquired lexical-phonological retrieval deficit. According to dual-route theory, BRB was reported to be unable to utilise the lexical route which led to a manifestation of surface dyslexia in English, and imageability effects in reading Turkish, with good reading of nonwords.

The current study aims to start filling the apparent research gap by exploring, in detail, the cognitive manifestation of DD in Turkish. The rationale for the proposed research is partially driven by the notion that a comprehensive account of dyslexia must accommodate phonological impairments as a potential determinant of reading and writing problems across different languages. It is therefore essential to examine the supposition that phonological impairments are associated with DD in languages that contain completely transparent orthography such as Turkish.

5.2.5 A note on DD policy in Turkey and Northern Cyprus

In Turkey, special education law was established when the Ö*zel Eğitime Muhtaç Çocuklar Kanunu* (Children with Special Education Need Law) (No. 2916) was legislated in 1983. The Turkish Ministry of Education (MEB) has formally recognised Developmental Dyslexia since 1997 (Special Education Regulation (No. 573)), although primarily uses the term “Specific Learning Difficulties” (SLD) rather than “Dyslexia”. In 2005, the Disability Act, Ozurluler Kanunu, no. 5378 (2005) was approved to protect the rights of people with disabilities, though not explicitly mentioning dyslexia. However, there are other legal regulations (*kararname*) which state that students with dyslexia have the right to special measures in class.

The identification of DD is exclusively carried out by state, and university hospitals using the Turkish version of tests such as the WISC-R (Savasir & Sahin, 1988) or Stanford-Binet (Ugurel-Semin, 1987) which are ill-suited for the identification of DD and the need to develop better reading achievement tests specific to Turkish are highlighted (Bingöl, 2003) in order to determine the prevalence of reading difficulties more accurately. In recent years, there have been several attempts at developing appropriate assessments for DD in Turkish-speaking populations. For example, the *Kelime Okuma Bilgisi Testi* (KOBİT) (Babür, Haznedar, Erdat-Çekerek, Erçetin, & Özerman, 2009) was designed to measure Turkish reading and was administered to 283 primary school students in Istanbul. While the authors assert that the reliability and validity of these specific tests have been established, they have not yet been shared with other researchers. In addition, the geographical restriction of the testing to Istanbul makes the generalisability to the broader Turkish population limited.

In sum, while there has been progress made towards both developing a national policy regarding DD and developing standardised/normalised measures for the identification of DD, there is still much room for further improvement. This chapter will aim to contribute to the evidence base regarding the selection of suitable materials for the development of standardised assessments for DD in Turkish-speaking children.

5.3 Method

The aims of the current study were four-fold. First, there was a need to investigate whether the classification of DD into subtypes is the same in Turkish as in other orthographies. The theoretical framework adopted for this purpose was the DRC model, and the same methodological procedures (classic and regression) (Castles & Coltheart, 1993) were explored. Second, there was a necessity to examine the incidence rates of surface and phonological DD subtypes in Turkish. Third, the current study aimed to explore individual-specific cognitive and linguistic deficits including phonological awareness, rapid automatized naming, visual attention span, working memory and visuo-spatial short-term memory using both a group and multiple-case study approach. Multiple-case studies, in contrast to single-case studies, are designed to study individual cases that are assumed to be representative of the larger population of individuals with DD within a given population. Therefore, multiple-case studies are considered to be more relevant than single-case studies in assessing the prevalence and the reliability of distinct profiles of DD (Sprenger-Charolles, Siegel, Jiménez & Ziegler, 2011). Besides, the detection of distinct subtypes of DD in Turkish children would be of pronounced significance for the identification of DD in Turkish. Finally, the applicability of the HGT (Wydell & Butterworth, 1999) would be evaluated for Turkish. Based on the above aims, the following hypotheses would be considered:

H1: Developmental Dyslexia, conceptualised primarily as a word-level literacy learning difficulty will be less evident in Turkish (Van Orden & Kloos, 2005).

H2: It is anticipated that the highly transparent nature of Turkish should not produce a high incidence of phonological developmental dyslexia (Wydell & Butterworth, 1999).

H3: In addition to this, Ziegler & Goswami, (2005) postulated, RAN may play a more significant role than phonological processing in the development of skilled reading for dyslexic learners in transparent orthographies.

H4: Considering cognitive skills, RAN would be the best predictor of reading at the word and text level in Turkish.

H5: There should be distinct word length effects observed in children with DD. It is hypothesised that these effects should be greater in the DD cohort.

5.3.1 Participants

Due to the lack of objective screening tools for developmental dyslexia in Turkish, the DD cohort was identified primarily by the distribution of Oral Reading Fluency scores and then by Reaction Times (RT) and error rates across all the tasks used in the study, employing a similar method used in Turkish adults (I. Raman, 2011). RTs are considered to provide a window into lexical processing and have been effectively utilised as a diagnostic tool especially in highly transparent orthographies that typically yield very low error rates, e.g. Italian (Tressoldi, Stella, & Faggella, 2001) and Spanish (Serrano & Defior, 2008). For the current experiment, the SD cut off for Turkish children was set at 1.25 (Landerl et al., 2013). This methodology was used in a large European study investigating Finnish, Hungarian, German, Dutch, French and English. Landerl and colleagues justify the use of a 1.25 SD cut-off as a pragmatic compromise between the standard criteria of -1 and -1.5 SDs. Finally, children with DD were included if they have been identified as having a reading difficulty in the absence of any overwhelming sensory, neurological and intellectual disorders or sociocultural factors.

Using the above approach, seven (7) children met the criteria for inclusion in the DD group for the current investigation. A further eight (8) participants were referred to the study by the Ministry of Education as suspected DD cases. In order to examine the equivalence of the two groups, independent samples t-tests were carried out for age, non-verbal IQ, SES and the ORF task. Examination revealed no significant differences for age, $t(13) = -.106$, $p > 0.5$, non-verbal IQ, $t(13) = -1.27$, $p > 0.5$, SES, $t(8.26) = -1.32$, $p > 0.5$, or ORF scores, $t(13) = -1.35$, $p > 0.5$. As such, the two groups were combined to form the DD group for the current investigation. Except for investigating incidence and prevalence rates, the newly formed cohort would be considered as one group for the remainder of this study.

Overall, fifteen (15) children (6 female) met the criteria for inclusion in the DD group for the current investigation. Children in the DD group had a mean age of 115.4 months (*SD*=8.48, range 103-129). In order to explore group differences, each child in the DD group was matched, when possible, to a child in an older typically developing (TD) control group based on age and non-verbal IQ and a child on the basis of ORF scores and non-verbal IQ, where possible, in order to create a younger TD control group. The typically developing control groups were drawn from the cohort previously described in Chapter 4. In a similar vein to Davies, Rodríguez-Ferreiro, Suárez & Cuetos (2013), the younger TD group used in the current study were not matched on reading ability to the children in the

DD group as ability matches on the basis of ORF were not possible with the current sample of children. Table 45 below provides an overview of group comparisons.

TABLE 45: SUMMARY OF GROUP COMPARISONS BETWEEN DD AND YOUNGER AND OLDER TD CHILDREN

	DD	Older TD	Younger TD	
	n=15	n=15	n=15	
Age (Months)	115.40 (8.48)	116.40 (7.80)	106.13 (12.65)	DD = Older TD DD > Younger TD*
Nonverbal IQ (36)	22.27 (7.42)	24.13 (4.66)	23.80 (6.41)	DD = Older TD DD = Younger TD
Oral Reading Fluency (WCPM)	29.00 (10.95)	70.47 (19.69)	42.20 (5.29)	Older TD > DD* Younger TD > DD*
Socioeconomic Status (1-4)	1.93 (0.59)	1.73 (0.80)	2.20 (1.08)	DD = Younger TD DD = Older TD

From table 45, as above, the DD and younger/ older TD groups showed no significant differences in terms of Nonverbal IQ and socioeconomic status. In terms of chronological age, the DD group was a similar age to the Older TD group but were significantly older (~9 months) then the Younger TD group. Oral reading fluency was significantly different between DD and younger TD children as well as DD and older TD children.

Following this, each child in the DD group was individually matched with 14-15 Older TD controls based on age, SES and non-verbal IQ and with 13-14 Younger TD controls based on ORF scores, SES and non-verbal IQ.

5.3.2 MATERIALS

The same tests that were created and used in Chapter 4 were used in the current study. The eight cognitive constructs explored by the battery were:

· Reading accuracy and speed in single-word naming

· Oral reading fluency (ORF)
· phonological awareness (PA)
· Rapid Automatized Naming (RAN)
· Visual Attention (VA) Span
· Non-verbal IQ
· Visuo-spatial short-term memory (VSSTM)
· Working memory (WM)

5.3.3 Procedure

The data collection procedure followed the exact protocol described in Chapter 4. Children were tested individually in a quiet room within the school. Stimuli were presented on a laptop using DMDX v5.1 (Forster & Forster, 2003). In addition, all of the children received the same instructions, which were displayed on the screen and reinforced orally. In order to simulate the natural conditions of individual reading on self-teaching, participants did not receive feedback on their responses, nor were they corrected if they misread the pseudowords (Álvarez-Cañizo, Suárez-Coalla, & Cuetos, 2018). Finally, children's responses were recorded in WAV format using DMDX and analysed with CheckVocal software (Protopapas, 2007) to calculate the number of correct responses and reaction times (RTs).

5.3.4 Statistical Analysis Plan

Statistical analyses were performed using RStudio. Planned comparisons between the DD, younger and older TD groups were carried out using statistical analyses based on Linear Mixed models previously described in Chapter 4.

In order to establish subtypes, the generally used procedure used in studies in opaque orthographies for identifying DD subtypes is based on pseudoword and irregular word reading accuracy. However, irregular words in transparent orthographies such as Turkish do not exist. Taking this into consideration, it has been suggested that RT data can be used as an alternative approach of detecting surface dyslexia in Spanish (Jiménez, Rodríguez, & Ramírez, 2009) and Greek (Douklias, Masterson & Hanley, 2009). Both the classic and the regression-based approaches to DD subtyping (Castles and Coltheart, 1993) were used in the current study. For the soft subtype approach, a regression analysis was conducted using two separate approaches. First by plotting nonword reading accuracy against word reading latencies and secondly by plotting nonword reading latency against high-frequency word naming latencies. Both approaches used the data for 30 of the Older TD control children in order to establish 90% confidence intervals (CIs), and children whose scores were outside of the established CIs were identified. Following this, modified t-tests (Crawford & Howell, 1998) were used to look for differences in scores in the background and experimental tasks between these cases and Older TD- and Younger TD-matched children. The resulting DD subgroups, i.e. phonological, surface and mixed were also compared to Older TD and Younger TD controls using univariate analysis of variance (ANOVA) tests. Post-hoc comparisons were conducted with either the Bonferroni or Games-Howell correction for multiple comparisons. Partial eta-squared was additionally calculated to determine the effect size of the differences between the groups.

5.4 RESULTS

5.4.1 GROUP COMPARISONS: WORD READING

Data extraction and cleaning

For the single word/pseudoword reading data, a total of 3,600 responses were recorded. Following data collection, the sound spectrograms of the recorded responses were analysed using CheckVocal (Protopapas, 2007) in order to extract corrected accuracy and RT measures. For the analysis of accuracy, all responses were considered. Transversely, for the analysis of RT, only correct responses were considered. Overall, there were 442 (12.8%) errors in the single word naming data which corresponded to 16 were non-responses (no response, <250ms or >3,000 ms to respond); 83 were word naming errors; and 343 were pseudoword errors. The removal of these RT data points from the data resulted in a final dataset of 3158 correct responses for further analysis.

Descriptive

The mean overall response time and accuracy across conditions and participants were 1328ms and 88.06%, respectively. Considering the group variable, overall accuracy was 82.67% for the DD group, 87.33% for the Younger TD group and 94.17% for the Older TD group. Overall response time was 1462ms for the DD group, 1324ms for the Younger TD group and 1197ms for the Older TD group. Considering the lexicality factor, overall response time and accuracy for words were 1180ms and 95.55%, respectively whereas response time and accuracy for nonwords was 1476ms and 80.55%, respectively. Considering the length factor, response time, and accuracy for short words/pseudowords were 1081ms and 97.56%, respectively whereas response time and accuracy for long words/ nonwords was 1285ms and 93.56%, respectively. Considering the frequency factor, response time and accuracy for high-frequency words was 1173ms and 96.22% respectively whereas response time and accuracy for low-frequency words was 1193ms and 94.89% respectively. Table 46 and 47 below provides a summary of the descriptive statistics for accuracy and RT data, respectively.

TABLE 46: MEAN ACCURACY FOR WORDS AND NONWORDS BY GROUP

	DD	**Younger TD**	**Older TD**
	Mean (SD)	Mean (SD)	Mean (SD)
Short Low Frequency	94.00 (9.10)	98.67 (3.52)	100.00 (0) *
Short High Frequency	96.00 (8.28)	97.33(5.94)	99.33 (2.58)
Short Psuedoword	83.33 (13.45)	90.00 (20.00)	95.67 (7.04) *
Long Low Frequency	86.67 (13.97)	92.00 (10.14)	98.00 (7.75) *
Long High Frequency	90.00 (15.12)	96.67 (6.17)	98.00 (5.61)
Long Pseudoword	64.00 (22.22)	67.00 (24.11)	83.33 (21.93) *

TABLE 47: MEAN RT FOR WORDS AND NONWORDS BY GROUP

	DD	**Younger TD**	**Older TD**
	Mean (SD)	Mean (SD)	Mean (SD)
Short Low Frequency	1195 (288)	1112 (219)	964 (162) *
Short High Frequency	1199 (291)	1057 (227)	959 (187) *
Short Psuedoword	1531 (305)	1360 (382)	1248 (406) *
Long Low Frequency	1434 (283)	1323 (381)	1132 (362) *
Long High Frequency	1408 (263)	1295 (369)	1119 (416) *
Long Pseudoword	1709 (359)	1500 (286)	1444 (271) *

As can be seen from Table 46, except long pseudowords, reading accuracy across participants was at near celling level. The general trend for accuracy scores was Older TD > Younger TD > DD with no exceptions. While the differences with between the DD and older TD groups was significant, the differences with the DD and younger TD groups were not. From table 47, as above, the general trend for RT scores was Older TD > Younger TD > DD with no exceptions. While the differences with between the DD and older TD groups were significant concerning low-frequency words and pseudowords (but not high-frequency words), the differences with the DD and younger TD groups were not. As stated above, the accuracy of responses to words/ pseudowords manipulated by length and lexicality was conducted GLMM analysis.

Accuracy – Length, Lexicality and Group

In the analysis of accuracy, the same approach as adopted in Chapter 4 was used to build LMM models. Briefly, the accuracy of responses to words/ pseudowords was conducted by using Generalized-Mixed effects Modelling (GLMM). The approach for the GLMM analysis involved conducting pairwise LRT comparisons (Pinheiro & Bates, 2000) of simpler models with more complex models, where each step of model comparison involves the former model building on the latter in order to determine the value of including various fixed and random effects in the models of single word/pseudoword reading accuracy. Table 48 below offers a summary of the final model. The final model for the analysis of Length, Lexicality and Group, was:

Accuracy ~ Length + Lexicality*Group + (Length| Subject) + (1|Item)

TABLE 48: SUMMARY TABLE OF THE FINAL GLMM MODEL OF WORD/PSEUDOWORD READING ACCURACY.

	Accuracy		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	17.78	9.24 – 34.20	**<0.001**
Length	0.24	0.18 – 0.40	**<0.001**
Lexicality	5.23	3.39 – 8.05	**<0.001**
DD vs Older TD	8.64	3.44 – 21.70	**<0.001**
DD vs Younger TD	2.04	0.90 – 4.63	0.089
Lexicality: DD vs Older TD	3.32	1.27 – 8.67	**0.014**
Lexicality: DD vs Younger TD	1.87	1.01 – 3.48	**0.046**
Random Effects			
σ^2	3.29		
τ_{00} **Item**	0.22		
τ_{00} **Subject**	1.09		
τ_{11} **Subject. Length**	0.46		
ρ_{01} **Subject**	-0.27		
ICC Item	0.05		
ICC Subject	0.24		
Observations	3600		
Marginal R^2 / Conditional R^2	0.365 / 0.549		

From Table 48, the estimated coefficients for the final model show that reading accuracy was higher for short words over long words as indicated by the length effect. There was also a significant effect of Lexicality, indicating that real words were named faster than pseudowords. Additionally, the model revealed Group effects indicating that the DD group was significantly less accurate than the Older TD group but not the Younger TD group (although this neared significance). The observed Lexicality by Group interaction effects indicates that lexicality effects were more evident for the DD group than both the younger and older TD groups.

Accuracy – Length, Frequency and Group

In order to explore the effects of word frequency, a separate (yet identical) analysis of participants' word reading accuracy (n=1800) was carried out. In this analysis, lexical frequency was added as a fixed effect predictor, and the lexicality term used in the above analysis was removed. The final model (including random intercepts and slopes) for the accuracy of responses to words is shown below and had the formula:

Accuracy ~ Length + Group +(1|Group:Subject) +(1|Item)

TABLE 49: SUMMARY TABLE OF THE FINAL GLMM MODEL OF WORD READING ACCURACY.

	Accuracy		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	37.56	16.13 – 87.45	**<0.001**
Length	0.33	0.17 – 0.61	**<0.001**
DD vs Older TD	10.61	2.97 – 37.97	**<0.001**
DD vs Younger TD	2.12	0.79 – 5.65	0.134
Random Effects			
σ^2	3.29		
$\tau_{00\ \text{Group: Subject}}$	1.16		
$\tau_{00\ \text{Item}}$	0.27		
$ICC_{\text{Group: Subject}}$	0.25		
ICC_{Item}	0.06		
Observations	1800		
Marginal R^2 / Conditional R^2	0.214 / 0.451		

From Table 49, the estimated coefficients for the final model show that word reading accuracy was higher for short words over long words as indicated by the length effect. Additionally, the model revealed Group effects indicating that the DD group was significantly less accurate than the Older TD group but not the Younger TD group.

Reading Speed - Length, Lexicality and Group

Following on from the analysis of the accuracy data, the same model selection procedure was applied to the RT data considering only correct responses (n=3170). The effects of Group, Length and Lexicality were explored using a Linear Mixed-effects model. The final formula for the best model fit was:

InverseRT ~ Length * Lexicality * Group + (Length | Subject) + (1 | Item)

Table 50 below offers a summary of the final model. From Table 50, the estimated coefficients for the final model show that single word/pseudoword RTs were shorter for short words over long words as well as words being named faster than pseudowords. There was also a significant effect of Group indicating that the DD group was significantly slower in word reading than the Older TD group but not the Younger TD group. Additionally, the model discovered a significant Length by Lexicality interaction. The interaction effect indicates that length effects were more evident for pseudowords rather than words. There was also a significant Length by Lexicality by Group effect between the DD group and the Older TD group (but not the Younger TD group) indicating that Length by Lexicality effect is statistically more evident in Older TD children than in children with DD. However, the Length by Lexicality effect was still significant in DD children (Figure 16).

TABLE 50: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD/PSEUDOWORD READING RTS.

	Inverse RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.83	-0.91 – -0.74	**<0.001**
Length	0.12	0.07 – 0.17	**<0.001**
Lexicality	-0.20	-0.25 – -0.16	**<0.001**
DD vs Older TD	-0.17	-0.28 – -0.05	**0.007**
DD vs Younger TD	-0.07	-0.18 – 0.05	0.274
Length: Lexicality	0.08	0.01 – 0.14	**0.018**
Length: DD vs Older TD	0.01	-0.05 – 0.07	0.678
Length: DD vs Younger TD	0.01	-0.04 – 0.07	0.752
Lexicality: DD vs Older TD	0.00	-0.04 – 0.04	0.995
Lexicality: DD vs Younger TD	0.01	-0.03 – 0.05	0.715
Length: Lexicality: DD vs Older TD	-0.14	-0.20 – -0.08	**<0.001**
Length: Lexicality: DD vs Younger TD	-0.02	-0.08 – 0.05	0.630
Random Effects			
σ^2	0.03		
$\tau_{00\ \mathrm{Item}}$	0.00		
$\tau_{00\ \mathrm{Subject}}$	0.03		
$\tau_{11\ \mathrm{Subject.Len1}}$	0.01		
$\rho_{01\ \mathrm{Subject}}$	-0.21		
$\mathrm{ICC}_{\mathrm{Item}}$	0.05		
$\mathrm{ICC}_{\mathrm{Subject}}$	0.43		
Observations	3170		
Marginal R^2 / Conditional R^2	0.227 / 0.600		

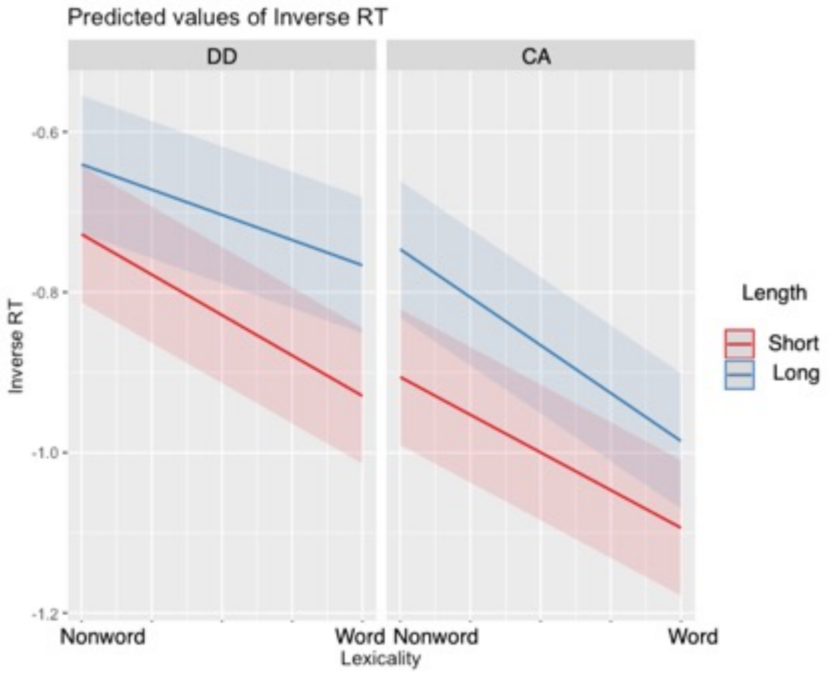

FIGURE 18: LENGTH BY LEXICALITY BY GROUP (DD VS OLDER TD) INTERACTION EFFECTS

Reading Speed - Length, Frequency and Group

In addition, to explore the effects of word frequency, a separate (yet identical) analysis of participants' word reading RTs was carried out. In this analysis, lexical frequency was added as a fixed effect predictor, and the lexicality term used in the above analysis was removed. The final model had the equation:

InverseRT ~ Length +Frequency+ Group + (Length | Subject) + (1 | Item)

TABLE 51: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RTS.

Inverse RT			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.92	-1.01 – -0.83	**<0.001**
Length	0.14	0.09 – 0.18	**<0.001**
Frequency	-0.03	-0.06 – 0.00	0.093
DD vs Older TD	-0.18	-0.30 – -0.06	**0.005**
DD vs Younger TD	-0.05	-0.17 – 0.07	0.440
Random Effects			
σ^2	0.03		
$\tau_{00\ Subject}$	0.03		
$\tau_{00\ Item}$	0.00		
$\tau_{11\ Subject.Len1}$	0.01		
$\rho_{01\ Subject}$	-0.23		
ICC $_{Subject}$	0.45		
ICC $_{Item}$	0.05		
Observations	1720		
Marginal R^2 / Conditional R^2	0.154 / 0.579		

From table 51, as above, the estimated coefficients for the final model show that single word RTs were faster for short words over long words as well as words being named faster as a function of Group with the DD group naming single words significantly slower than the Older TD group but not the Younger TD group. Of note, the frequency effect was found to be non-significant after length was added to the model.

In order to explore specific length by group interactions, the random slope for Len| Subject was removed with the final model having the formula:

InverseRT~ Len*Group + (1|Subject) + (1|Item)

TABLE 52: SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RTS WITH THE RANDOM SLOPE FOR LEN|
SUBJECT REMOVED

	Inverse RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.93	-1.02 – -0.84	**<0.001**
Length	0.16	0.12 – 0.21	**<0.001**
DD vs. Older TD	-0.17	-0.29 – -0.05	**0.009**
DD vs. Younger TD	-0.06	-0.18 – 0.06	0.326
Length: DD vs. Older TD	-0.06	-0.11 – -0.02	**0.004**
Length: Younger TD	-0.00	-0.05 – 0.04	0.893
Random Effects			
σ^2	0.03		
τ_{00} Subject	0.03		
τ_{00} Item	0.00		
ICC Subject	0.42		
ICC Item	0.05		
Observations	1720		
Marginal R^2 / Conditional R^2	0.160 / 0.555		

From table 52, as above, the estimated coefficients for the final model show that single word RTs were faster for short words over long words as well as words being named faster as a function of Group with the DD group naming single words significantly slower than the Older TD group but not the Younger TD group. Additionally, there was a significant interaction between Length and Group between DD and Older TD groups. The interaction effect indicates that length effects were more evident for DD children than Older TD controls.

Furthermore, in a similar vein to Chapter 4, to explore the non-significant finding of a frequency effect, the above model was rerun using an Inverse Gaussian distribution applied to the raw RT data. The new model has the formula:

RT~ Length + Frequency +Group + (Len*Group) +(Frequency*Group) + (Len*Freq*Group) + (1|Subject) + (1|Item)

TABLE 53 SUMMARY TABLE OF THE FINAL LMM MODEL OF WORD READING RTS WITH MODIFICATION TO RT

	RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1294.40	1227.64 – 1361.16	**<0.001**
Length	248.93	193.82 – 304.04	**<0.001**
Frequency	12.46	-35.64 – 60.57	0.612
DD vs Older TD	-97.61	-154.43 – -40.78	**0.001**
DD vs Younger TD	-11.90	-76.31 – 52.51	0.717
Length: DD vs Older TD	-176.54	-216.82 – -136.26	**<0.001**
Length: DD vs Younger TD	-105.08	-149.84 – -60.32	**<0.001**
Frequency: DD vs Older TD	-40.76	-79.02 – -2.50	**0.037**
Frequency: DD vs Younger TD	-86.98	-126.35 – -47.61	**<0.001**
Length: Frequency	-26.64	-88.97 – 35.69	0.402
Length: Frequency: DD vs Older TD	26.92	-23.41 – 77.25	0.295
Length: Frequency: DD vs Younger TD	70.74	19.78 – 121.71	**0.007**
Random Effects			
σ^2	0.00		
τ_{00} Subject	17309.67		
τ_{00} Item	3202.33		
ICC Subject	0.84		
ICC Item	0.16		
Observations	1720		
Marginal R^2 / Conditional R^2	0.407 / 1.000		

From table 53, as above, the estimated coefficients for the final model show that single word RTs were faster for short words over long words as well as words being named faster as a function of Group with the DD group naming single words significantly slower than the Older TD group but not the Younger TD group in support of the previous analysis. Additionally, there was a significant interaction between Length and Group between DD and both TD groups. The interaction effect indicates that length effects were more evident for DD children than TD controls. Furthermore, there was a significant interaction between Frequency and Group between DD and both TD groups. The interaction effect indicates that frequency effects were less evident for DD children than TD controls. Finally, a 3-way interaction between Length, Frequency and Group for DD vs Younger TD children revealed that the greater frequency effect for long words was less marked in the DD group compared to younger TD readers

5.4.2 Cognitive Predictors of Word/Pseudoword Reading

In a final analysis, the cognitive predictors of single word/pseudoword reading were considered. First, both words and pseudowords were considered together for both accuracy and RT measures. Following this, a separate analysis was carried out for words and pseudowords.

Word/Pseudoword Reading Accuracy

The same model selection procedure used in the previous analyses was applied to both the accuracy and RT data. The effects of Group, PA, RAN, VA Span, WM and VSSTM on accuracy scores were explored using a Generalised Linear Mixed-effects model. The final formula for the best model fit was:

Accuracy ~ PA + Group + (1 | Subject) + (1 | Item)

TABLE 54: SUMMARY TABLE OF THE FINAL GLMM MODEL OF COGNITIVE PREDICTORS OF WORD READING ACCURACY.

	Accuracy		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	4.86	1.56 – 15.10	**0.006**
PA	1.13	1.04 – 1.24	**0.005**
DD vs Older TD	3.71	1.00 – 13.78	**0.050**
DD vs Younger TD	1.34	0.52 – 3.44	0.542
Random Effects			
σ^2	3.29		
$\tau_{00\ \text{Subject}}$	0.80		
$\tau_{00\ \text{Item}}$	0.51		
ICC $_{\text{Subject}}$	0.17		
ICC $_{\text{Item}}$	0.11		
Observations	1800		
Marginal R^2 / Conditional R^2	0.236 / 0.454		

From Table 54, the estimated coefficients for the final model show that reading accuracy was more likely to produce correct responses by students with higher scores in PA. In addition, there was a significant difference between children in the DD group and Older TD children in PA scores. Similar to the parallel approach in Chapter 4, separate analysis for words and pseudowords yielded the same final model and as such will not be reported further.

Word/Pseudoword Reading Speed

The same model selection procedure used in the previous analyses was applied to the RT data. The effects of Grade, PA, RAN, VA Span, WM and VSSTM on RT scores were explored using a Linear Mixed-effects model. The final formula for the best model fit was:

InverseRT ~ PA + Group + (1|Subject) + (1 | Item)

From Table 55 below, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with higher scores in PA. No other cognitive predictors or interactions reached significance for inclusion into the final model.

TABLE 55: SUMMARY TABLE OF THE FINAL LMM MODEL OF COGNITIVE PREDICTORS OF WORD/PSUEDOWORD READING RT.

	Inverse RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.65	-0.78 – -0.52	**<0.001**
PA	-0.01	-0.02 – -0.00	**0.028**
DD vs Older TD	-0.09	-0.21 – 0.04	0.195
DD vs Younger TD	-0.03	-0.14 – 0.08	0.575
Random Effects			
σ^2	0.03		
$\tau_{00\ \text{Item}}$	0.02		
$\tau_{00\ \text{Subject}}$	0.02		
ICC_{Item}	0.23		
ICC_{Subject}	0.31		
Observations	3170		
Marginal R^2 / Conditional R^2	0.095 / 0.582		

Furthermore, in a similar vein to Chapter 4, the above model was rerun using an Inverse Gaussian distribution applied to the raw RT data. The new model has the formula:

RT ~ PA + RAN + Group + (RAN * Group) + (1 | Subject) + (1 | Item)

TABLE 56: SUMMARY TABLE OF THE FINAL LMM MODEL OF COGNITIVE PREDICTORS OF WORD/PSUEDOWORD READING RT WITH MODIFICATIONS TO RT

	RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1269.85	1127.36 – 1412.34	**<0.001**
PA	-3.10	-19.42 – 13.23	0.710
RAN	10.95	5.99 – 15.91	**<0.001**
DD vs. Older TD	166.43	114.52 – 218.33	**<0.001**
DD vs Younger TD	-359.05	-472.31 – -245.80	**<0.001**
RAN: Older TD	-3.93	-13.99 – 6.14	0.445
RAN: Younger TD	11.19	3.90 – 18.49	**0.003**
Random Effects			
σ^2	0.00		
$\tau_{00\ \text{Item}}$	12611.73		
$\tau_{00\ \text{Subject}}$	14934.04		
ICC_{Item}	0.46		
ICC_{Subject}	0.54		
Observations	3170		
Marginal R^2 / Conditional R^2	0.290 / 1.000		

From Table 56, as above, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with lower scores in RAN. In addition, there was a significant difference between children in the DD group and younger TD children in RAN scores. The interaction effect indicates that the effect of RAN on RT was less evident for DD children than younger TD controls.

Separate analysis for words revealed the final formula of the best model fit to be:

InverseRT ~ PA + RAN + VA Span + Group + (PA | Subject) + (Group | Item)

From Table 57 below, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with lower scores in RAN.

No other cognitive predictors or interactions reached significance for inclusion into the final model.

TABLE 57: SUMMARY TABLE OF THE FINAL LMM MODEL OF COGNITIVE PREDICTORS OF WORD READING RT.

	Inverse RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-1.09	-1.41 – -0.77	**<0.001**
PA	-0.00	-0.01 – 0.01	0.779
RAN	0.01	0.00 – 0.01	**0.038**
VA Span	-0.00	-0.00 – 0.00	0.632
DD vs Older TD	-0.09	-0.23 – 0.05	0.231
DD vs Younger TD	-0.03	-0.13 – 0.08	0.653
Random Effects			
σ^2	0.03		
τ_{00} Subject	0.00		
τ_{00} Item	0.01		
τ_{11} Subject.PA	0.00		
τ_{11} Item.DD vs older TD	0.00		
τ_{11} Item.DD vs younger TD	0.00		
ρ_{01} Subject	1.00		
ρ_{01} Item.DD vs older TD	-1.00		
ρ_{01} Item.DD vs younger TD	-1.00		
ICC Subject	0.04		
ICC Item	0.25		
Observations	1720		

Furthermore, in a similar vein to Chapter 4, the above model was rerun using an Inverse Gaussian distribution applied to the raw RT data. The new model has the formula:

RT ~ RAN + Group + (RAN * Group) + (1 | Subject) + (1 | Item)

TABLE 58: SUMMARY TABLE OF THE FINAL LMM MODEL OF COGNITIVE PREDICTORS OF WORD READING RT WITH MODIFICATION TO RT

	RT		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	772.85	698.08 – 847.61	**<0.001**
RAN	15.88	12.07 – 19.68	**<0.001**
DD vs Older TD	248.01	165.71 – 330.30	**<0.001**
DD vs Younger TD	145.60	74.34 – 216.87	**<0.001**
RAN: DD vs Older TD	-8.07	-15.57 – -0.58	**0.035**
RAN: DD vs Younger TD	-4.06	-9.82 – 1.71	0.168
Random Effects			
σ^2	0.00		
τ_{00} Subject	14440.14		
τ_{00} Item	5335.64		
ICC Subject	0.73		
ICC Item	0.27		
Observations	1720		
Marginal R^2 / Conditional R^2	0.412 / 1.000		

From Table 58, as above, the estimated coefficients for the final model show that reading speed was more likely to produce faster responses by students with lower scores in RAN. In addition, there was a significant difference between children in the DD group and older TD children in RAN scores. The interaction effect indicates that the effect of RAN on RT was more evident for DD children than older TD controls.

5.4.3 Subtyping

Moving beyond group studies, two methods were used to analyse the presence of subtypes within the DD cohort, namely, the classical method and the regression method.

Chronological Age Comparison: Hard subtypes (Classical method)

In order to identify hard Phonological and Surface DD cases, the classical method with the typically used criterion of one SD below the mean accuracy score of Older TD controls (or above the mean for RT). For accuracy, grade-based cut-offs for nonword accuracy were 71.97% for Grade 2, 74.73% for Grade 3, 72.18% for Grade 4 and 78.59% for Grade 5. Out of the 15 dyslexics, eight were reduced in pseudoword reading accuracy. Grade-based cut-offs for word accuracy were 92.91% for Grade 2, 95.92% for Grade 3, 91.35% for Grade 4 and 96.24% for Grade 5. Out of the 15 dyslexics, eight were deficient in word reading accuracy. Taken together, 5 had both deficits and had four neither. Accordingly, only three children were selectively impaired in pseudoword reading accuracy (Phonological DD), and three were selectively impaired in word reading accuracy (Surface DD). Hence, 6 of the 15 dyslexics had a selective deficit (40%).

Considering the classical method on word/pseudoword RTs, 11 of the 15 children in the DD subgroup were slow in pseudoword reading, 11 were slow in word reading, 10 had both deficits, and 3 had neither deficit. Only one child was selectively slow in pseudoword reading (Phonological DD), and one was selectively slow in word reading (Surface DD). Thus, a selective deficit was found in only 2 of the 15 dyslexics (13%). As is commonly used in subtyping studies in transparent orthographies, when both nonword accuracy and word reading RT were considered together, 3 of the 15 children in the DD subgroup were inaccurate in pseudoword reading (Phonological DD), six were slow in word reading (Surface DD), 5 had both deficits (Mixed DD), and 1 had neither deficit.

Chronological Age Comparison: Soft subtypes (Regression-based method)

In order to further explore the profiles of phonological and surface dyslexia, the regression-based technique employed by Castles and Coltheart (1993) and adapted for use in transparent orthographies by Niolaki, Terzopoulos and Masterson (2014) was used. A regression analysis was carried out using scores of 30 Older TD control children for nonword reading accuracy and word reading RT.

The regression discovered a significant relationship (r = -.626, P <.001) between the two variables (F (1,29) = 18.04, p<.001). In order to investigate if age accounted for this significant relationship, an additional analysis with chronological age as a covariate was carried out. The resulting partial correlation indicated that the relationship between the two variables remained significant (partial r=-.498, p<.001). Predicted values based on the relationship were used to identify those children in the DD group with scores markedly below expectation on one task based on their performance on the supplementary task. The resulting mean predicted values and 90% CI for a subgroup of 30 Older TD matched typically developing children as well as the DD children were plotted and are presented in Figure 17 below.

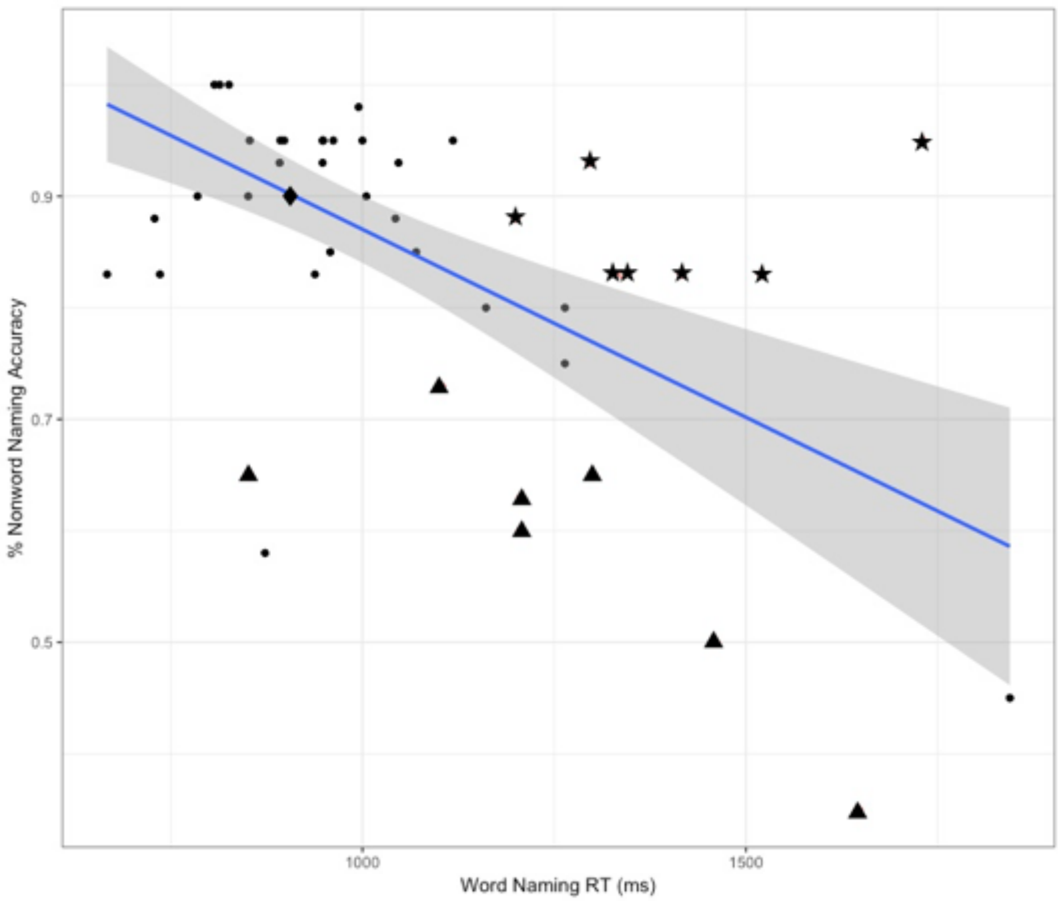

FIGURE 19: NONWORD READING ACCURACY BY WORD READING RT FOR 30 OLDER TD CHILDREN, AND 90% CONFIDENCE INTERVALS.

Note: Stars represent dyslexic children with the profile of surface dyslexia (n=7), triangles represent dyslexic children with the profile of phonological dyslexia (n=7) and diamond symbols represent dyslexic children with a mixed profile (n=1).

Before analysing the difference between groups, there was a need to consider the potential group dissimilarities in non-verbal IQ. Due to the small group sizes, a one-way ANOVA was used instead of the LMM approach. The analysis revealed that there was no significant difference between groups on non-verbal IQ $F(3,34) = 1.06$, $p = 0.38$. Analysing Oral Reading Fluency at the word and syllable level, there were significant differences between groups at the word, Welch $F(3,14.71) = 16.17$, $p<0.001$, $\eta^2 = .60$ and syllable, Welch $F(3,14.53) = 13.87$, $p<0.001$, $\eta^2 = .58$, levels. Planned post-hoc comparisons using the Games-Howell correction revealed significant differences between the Phonological DD subgroup and the Older TD (46 words, $p<0.001$) and the Younger TD (19 words, $p = 0.037$) controls but not with the Surface DD subgroup ($p> 0.05$) for the ORF word level. The surface DD group only differed significantly from Older TD (36 words, $p< 0.001$). Post-hoc comparisons using the Games-Howell correction for ORF syllable scores revealed significant differences between the Phonological DD subgroup and the Older TD (120 syllables, $p<0.001$) and the Younger TD (47 syllables, $p = 0.009$) controls but not with the Surface DD subgroup ($p> 0.05$) for the ORF word level. The surface DD group only differed significantly from Older TD controls (87 syllables, $p< 0.001$). Taken together, the Phonological DD group manifested with an overall deficit in ORF, whereas the Surface DD group appeared to reflect a delay in their ORF.

Following this, single word/ nonword naming accuracy and speed were analysed for group differences. The resulting ANOVAs discovered significant group differences for nonword reading accuracy, $F(3,43) = 5.72$, $p = 0.002$, $\eta^2 = .30$ and word reading RT, $F(3,43) = 3.16$, $p = 0.035$, $\eta^2 = .19$ but not for word reading accuracy Welch $F(3,14.54) = 3.18$, $p=0.056$ or nonword reading RT, $F(3,43) = 2.52$, $p=0.072$. Post-hoc comparisons using the Bonferroni correction for nonword reading accuracy scores revealed significant differences between the Phonological DD subgroup and the Older TD (29%, $p= 0.001$)

and the Younger TD (20%, $p = 0.042$) controls as well as the Surface DD subgroup (25%, $p> 0.001$). The surface DD group only differed significantly from the Phonological DD group previously reported above. Post-hoc comparisons using the Bonferroni correction for word reading RT scores revealed no significant differences between the groups. However, group differences between the surface DD group and the Older TD control group approached significance ($p= 0.056$).

Considering the cognitive profiles of the subgroups, there were significant differences between groups for PA, Welch $F(3,13.86) = 5.47$, $p = 0.011$, $\eta^2= .34$, and RAN, Welch $F(3, 14.37) = 4.65$, $p = 0.018$, $\eta^2= .27$ but not VA Span $F(3,43) = 2.66$, $p = 0.061$, WM, Welch $F(3,15.70) = 0.80$, $p = 0.52$,or VSSTM, $F(3,43) = 0.48$, $p = 0.70$. Post-hoc comparisons using the Games-Howell correction for PA scores revealed significant differences between the Phonological DD subgroup and the Older TD (12, $p= 0.003$). The differences between the phonological DD group and the Younger TD controls approached significance ($p= 0.051$). The surface DD group did not significantly differ from controls. Post-hoc comparisons using the Games-Howell correction for RAN scores only revealed significant differences between the Older TD and Younger TD control groups (5.64 seconds, $p= 0.031$).

In a further consideration, and in line with Jiménez, Rodríguez, Ramírez (2009), the regression analysis approach for subtyping was further explored using the RT for high-frequency words and pseudowords. See Figure 18 below.

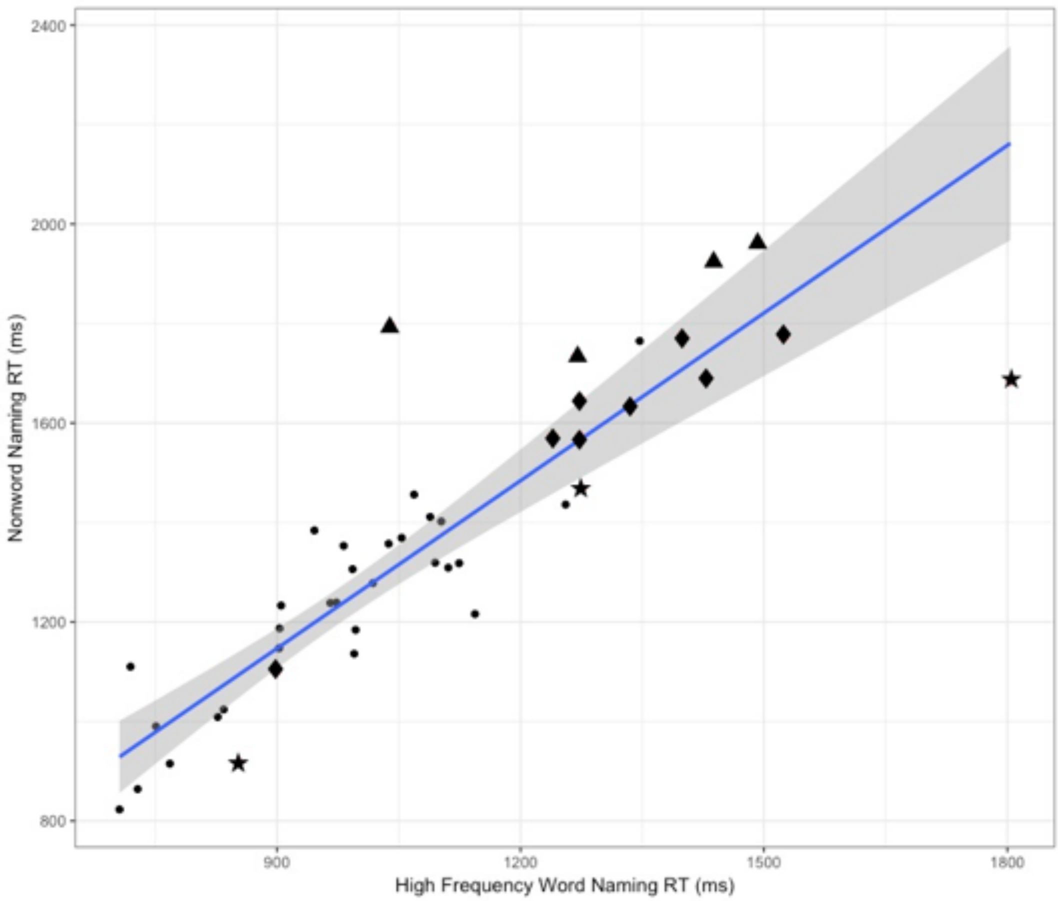

FIGURE 20: NONWORD READING RT BY WORD READING RT FOR 30 OLDER TD CHILDREN, AND 90% CONFIDENCE INTERVALS.

Note: Stars represent dyslexic children with the profile of surface dyslexia (n=3), triangles represent dyslexic children with the profile of phonological dyslexia (n=4) and diamond symbols represent dyslexic children with a mixed profile (n=8).

Using this approach, three children belonging to the phonological subgroup and four belonging to the surface subtype appear as mixed profiles. The resulting analysis also suggests that the approach originally adopted by Douklias, Masterson, and Hanley (2009) may be better suited for differentiating subgroups of DD in highly transparent orthographies.

Comparing individual DD children and typically developing controls

Using modified t-tests, the scores of the battery of tests of each individual child of the DD group was compared to those of 6 - 10 older TD children on the basis of chronological age, SES and nonverbal IQ as well as 5-8 younger TD controls on the basis of scores in the oral reading fluency task, SES and nonverbal IQ. Three children (PD4, SD4, SD5) did not have suitable younger TD children comparisons available. The resulting comparison is reported in Table 59 and 60.

TABLE 59: AGE, GENDER AND WORD/ PSEUDOWORD READING SCORES FOR CHILDREN WITH DD IN THE SURFACE, PHONOLOGICAL AND MIXED SUBTYPE GROUPS WITH OLDER TD AND YOUNGER TD CONTROLS (STANDARD DEVIATIONS ARE REPORTED IN PARENTHESES). * *P* < .05 (SIGNIFICANCE TEST: CRAWFORD & HOWELL, 1998).

	Phonological DD							Surface DD							Mixed DD
	PD1	PD2	PD3	PD4	PD5	PD6	PD7	SD1	SD2	SD3	SD4	SD5	SD6	SD7	MD1
Age (Months)	107	129	112	116	105	103	107	123	121	122	109	110	117	128	122
Gender	F	M	M	M	F	F	M	M	F	F	M	M	F	M	M
Older / Younger TD (N)	10 / 8	9 / 7	9 / 7	7 / 8	8 / NA	7 / 5	6 / 8	7 / 8	8 / 6	7 / 5	8 / NA	7 / NA	6 / 6	6 / 6	8 / 8
Older TD mean age (Months)	106.4 (2.8)	129.4 (1.0)	113.2 (0.83)	115.6 (1.62)	100.4 (8.94)	103.3 (1.50)	105.5 (1.87)	123.7 (1.25)	120.8 (3.15)	121.4 (1.62)	111.1 (2.47)	111.6 (2.99)	117.0 (2.53)	125.5 (4.97)	121.5 (5.0)
Older TD mean RCPM	25.2 (3.8)	27.6 (3.4)	25.1 (4.9)	24.3 (3.09)	16.5 (1.6)	15.03 (6.8)	24.33 (3.3)	25.0 (2.52)	24.6 (2.20)	19.14 (7.01)	24.67 (3.78)	27.20 (3.63)	25.17 (1.60)	15.50 (6.35)	30.9 (1.0)*
Older TD mean SES	1.7 (.08)	2.9 (1.05)	2.0 (1.07)	2.0 (1.0)	2.0 (1.15)	0.87 (.68)	1.83 (.75)	2.67 (1.03)	2.14 (1.21)	1.83 (1.33)	1.50 (.55)	1.71 (.50)	1.83 (.98)	2.50 (1.38)	2.4 (0.7)
Younger TD mean age (Months)	93.5 (6.1)*	101.9 (12.9)*	95.1 (6.6)*	93.6 (6.32)*	-	92.2 (4.55)*	92.8 (2.49)*	101.0 (10.65)*	98.8 (11.14)	97.00 (8.80)*	-	-	98.7 (10.88)	98.00 (8.25)*	109.3 (15.4)
Younger TD mean RCPM	21.25 (6.76)	26.17 (4.07)	26.00 (5.13)	24.4 (6.57)	-	18.8 (3.70)	23.6 (2.20)	24.6 (5.62)	26.6 (4.39)	15.6 (1.14)	-	-	26.17 (1.47)	15.83 (1.17)	26.7 (3.44)*
Younger TD mean SES	1.9 (.08)	2.3 (1.38)	2.6 (1.13)	2.5 (1.19)	-	1.0 (1.41)	2.6 (.92)	2.2 (1.36)	2.5 (1.38)	1.80 (.84)	-	-	2.17 (1.47)	1.67 (.82)	1.9 (0.9)
Older TD mean ORF Syllables	171.4 (52.7)*	235.0 (54.3)*	165.3 (59.0)*	184.0 (38.4)*	156.0 (41.1)	168.4 (44.3)*	183.7 (26.9)*	195.6 (39.4)*	206.4 (39.5)*	186.1 (56.67)	178.9 (66.42)	186.4 (55.11)*	172.2 (49.87)	200.33 (57.72)	218.3 (29.4)*
Younger TD mean ORF Syllables	125.0 (19.9)*	98.6 (6.8)*	108.3 (13.7)*	109.6 (13.4)	-	157.8 (36.93)*	143.6 (40.5)*	104.1 (9.6)	99.8 (6.46)	132.2 (12.58)	-	-	102.3 (9.61)	150.33 (45.82)	129.7 (9.55)
Word Accuracy	70%	88%	85%	80%	98%	98%	90%	88%	100%	100%	100%	93%	100%	90%	98%
DD vs. Older TD	96.94 (4.10)*	98.61 (2.2)*	100.00 (0.00)*	100.00 (0.00)*	96.25 (4.23)	97.50 (3.82)	98.75 (2.09)*	98.93 (1.34)*	99.06 (2.65)	97.14 (4.66)	99.38 (1.77)	99.64 (.95)*	99.17 (2.04)	96.00 (5.00)	99.7 (.009)*
DD vs. Younger TD	95.00 (3.78)*	96.43 (4.05)*	95.71 (3.45)*	95.63 (3.20)*	-	97.00 (4.11)	96.88 (3.72)	96.88 (3.95)*	96.67 (4.38)	94.58 (3.68)	-	-	96.67 (4.38)	95.00 (4.00)	98.2 (.02) *
Word RT	1647	1460	1286	852	1103	1300	1520	1340	1416	1337	1207	1200	1297	1731	907
DD vs. Older TD	1204 (459)	858 (102)*	1053 (231)	952 (151)	1088 (164)	1056 (185)	962 (164)*	871 (62)*	910 (128)*	945 (139)*	1002 (166)	1005 (223)	955 (91)*	942 (153)*	917 (75)
DD vs. Younger TD	1126 (113)*	1184 (166)	1276 (388)	1255 (366)	-	1063 (130)	1094 (112)*	1217 (128)	1226 (135)	1097 (136)	-	-	1215 (124)	1085 (147)*	1085 (163)
Nonword Accuracy	35%	50%	60%	63%	65%	65%	73%	83%	83%	83%	83%	88%	93%	95%	90%
DD vs. Older TD	78.88 (19.16)*	86.94 (12.73)*	90.56 (6.22)*	95.00 (6.29)*	86.56 (8.44)*	90.36 (6.20)*	89.58 (4.85)*	90.71 (10.97)	93.75 (5.82)	89.64 (6.03)	91.25 (6.55)	88.93 (7.20)	96.67 (3.03)	88.00 (5.00)	93.4 (11.70)
DD vs. Younger TD	81.25 (5.51)*	87.14 (6.36)*	81.79 (16.25)	80.00 (14.20)	-	86.50 (11.54)	88.75 (9.06)	84.69 (4.52)	85.42 (4.85)	81.67 (7.36)	-	-	83.75 (4.68)	84.00 (8.00)	77.9 (26.75)
Nonword RT	1960	1689	1574	919	1799	1735	1771	1776	1927	1634	1574	1651	1469	1684	1099
DD vs. Older TD	1496 (627)	1070 (174)*	1347 (322)	1205 (192)	1290 (207)*	1305 (190)*	1168 (166)*	1186 (166)*	1136 (145)*	1177 (198)*	1309 (168)	1331 (331)	1215 (103)*	1155 (217)*	1085(112)
DD vs. Younger TD	1333 (136)*	1414 (119)*	1560 (477)	1483 (480)	-	1241 (175)*	1297 (146)*	1443 (188)	1436 (113)*	1307 (138)*	-	-	1484 (177)	1307 (138)*	1404 (254)

TABLE 60: COGNITIVE PROFILES FOR CHILDREN WITH DD IN THE SURFACE, PHONOLOGICAL AND MIXED SUBTYPE GROUPS WITH OLDER TD AND YOUNGER TD CONTROLS (STANDARD DEVIATIONS ARE REPORTED IN PARENTHESES). * *P* < .05 (SIGNIFICANCE TEST: CRAWFORD & HOWELL, 1998).

	Phonological DD							Surface DD							Mixed DD
	PD1	PD2	PD3	PD4	PD5	PD6	PD7	SD1	SD2	SD3	SD4	SD5	SD6	SD7	MD1
Age (Months)	107	129	112	116	105	103	107	123	121	122	109	110	117	128	122
Gender	F	M	M	M	F	F	M	M	F	F	M	M	F	M	M (9)
PA	4	2	10	13	6	17	11	15	27	15	11	13	14	5	27
DD vs. Older TD	18.00 (7.25)*	27.44 (6.42)*	21.67 (6.93)	20.00 (2.38)*	15.63 (3.93)*	17.00 (5.51)	20.33 (7.66)	21.71 (5.47)	22.00 (5.66)	23.29 (6.50)	20.75 (9.54)	22.14 (8.73)	20.33 (3.78)	23.00 (4.56)*	24.38 (7.98)
DD vs. Younger TD	14.38 (2.77)*	15.43 (2.57)*	15.14 (2.19)*	15.00 (2.20)	-	14.00 (6.44)	15.13 (5.51)	14.75 (2.43)	15.17 (2.71) ↑	14.5 (2.74)	-	-	14.17 (1.33)	13.83 (2.14)*	16.43 (2.37) ↑
RAN	58.69	53.78	33.36	27.18	38.39	32.68	38.04	37	43.31	34.58	38.23	30.65	28.04	64.64	29.27
DD vs. Older TD	30.72 (4.95)*	22.77 (2.86)*	27.50 (3.37)	27.88 (3.25)	33.83 (5.72)	32.48 (4.22)	28.53 (3.18)*	27.32 (3.14)*	27.80 (2.85)*	27.60 (4.74)	28.31 (4.24)*	26.90 (2.72)	30.47 (7.22)	24.92 (3.93)*	26.44 (3.05)
DD vs. Younger TD	37.29 (3.49)*	35.81 (6.08)*	36.71 (4.75)	36.72 (4.52)	-	31.34 (7.17)	31.93 (6.40)	33.95 (5.22)	34.62 (5.69)	36.01 (2.89)	-	-	34.84 (5.38)	35.03 (4.83)*	31.05 (3.06)
VA Span	35	20	10	54	49	37	34	39	31	19	45	26	51	29	63
DD vs. Older TD	61.00 (11.73)*	65.00 (18.30)*	55.00 (12.14)*	55.43 (12.18)	45.13 (17.63)	50.71 (13.65)	61.67 (13.85)	65.86 (11.77)*	66.50 (13.70)*	58.71 (19.41) º	53.88 (13.21)	58.14 (8.28)*	62.00 (14.38)	54.00 (26.43)	77.88 (7.10)*
DD vs. Younger TD	45.13 (15.40)	39.14 (14.37)	38.00 (14.61)	37.38 (14.29)	-	44.60 (24.83)	48.75 (19.48)	39.75 (13.45)	38.67 (15.68)	41.00 (16.70)	-	-	36.67 (13.49)	42.00 (17.44)	47.86 (12.85)
WM	2	2	2	5	2	2	4	3	2	2	3	3	4	2	4
DD vs. Older TD	3.44 (.88)	4 (.50)*	3.78 (.44)*	3.43 (.79)	3.00 (.76)	3.29 (.76)	4 (0)	3.43 (.79)	3.50 (.76)	3.14 (.90)	3.5 (.76)	3.86 (.38)*	3.50 (0.55)	3.33 (.82)	3.88 (.41)
DD vs. Younger TD	3.00 (.76)	3.43 (.53)*	3.00 (.58)	3.00 (.53) ↑	-	2.80 (.45)	2.75 (.46) ↑	3.38 (.52)	3.50 (.55)*	3.00 (.63)	-	-	3.33 (.52)	2.83 (.75)	3.29 (.76)
VSSTM	9	13	9	13	4	10	11	7	16	13	19	8	11	6	11
DD vs. Older TD	8.89 (2.89)	14.67 (2.35)	12.33 (2.92)	12.29 (2.81)	9.88 (2.64)*	11.14 (2.27)	9.00 (2.76)	11.71 (3.15)	11.50 (3.89)	10.43 (4.31)	12.13 (3.87)	11.00 (3.65)	12.33 (3.39)	10.33 (4.63)	14.75 (2.60)
DD vs. Younger TD	8.89 (2.90)	10.14 (1.77)	9.14 (2.19)	9.13 (2.17)	-	7.60 (.89)	7.88 (1.96)	10.50 (1.69)	10.33 (1.86) ↑	9.17 (2.32)	-	-	10.33 (	9.33 (2.45)	9.00 (5.23)

For word reading accuracy, 5 out of 7 children in the Phonological DD subgroup (PD1, PD2, PD3, PD4 and PD7), 2 out of the seven children in the Surface DD subgroup (SD1 and SD5) and MD1 had scores that were significantly lower than those of Older TD controls. In addition to this, PD1, PD2, PD3, PD4, SD1 and MD1 reported word reading accuracy scores that were significantly below that of Younger TD controls. Regarding word reading RT, 2 out of 7 children in the Phonological DD subgroup (PD2, PD7) and 5 out of 7 children in the Surface DD subgroup (SD1, SD2, SD3, SD4 and SD7) were significantly slower than those of Older TD controls. Additionally, PD1, PD7 and SD7 were significantly slower in single-word naming than Younger TD controls.

For nonword reading accuracy, all of the children that fit the phonological DD profile had scores that were significantly lower than those of Older TD controls. None of the children that fit the surface DD profile had scores that were significantly lower than those of Older TD controls. In addition to this, PD1 and PD2 reported nonword reading accuracy scores that were significantly below that of Younger TD controls—regarding nonword reading RT, four (4) of the seven children in the phonological DD group (PD2, PD5, PD6 and PD7) and 5 of the children in the surface DD group (SD1, SD2, SD3, SD6 and SD7 were significantly slower than those of Older TD controls. Additionally, PD1, PD2, PD6, PD7, SD2, SD3 and SD7 was significantly slower in single nonword naming than Younger TD controls.

Considering the cognitive profile of the DD subgroups, scores on PA were significantly lower than Older TD controls for 4 of the seven children that fit the phonological DD profile (PD1, PD2, PD4 and PD5) as well as 1 of the seven children that fit the surface DD profile (SD7). In addition, PD1, PD2, PD3, and SD7 scored significantly lower than Younger TD controls for PA. Scores on RAN were significantly slower than Older TD controls for 3 of the seven children that fit the phonological DD profile (PD1, PD2, and PD7) as well as 4 of the seven children that fit the surface DD profile (SD1, SD2, SD4 and SD7). In addition, PD1, PD2 and SD7 were significantly slower than Younger TD controls for RAN. Scores on VA span were significantly lower than Older TD controls for 3 of the seven children that fit the phonological DD profile (PD1, PD2, and PD3) as well as 3 of the seven children that fit the surface DD profile (SD1, SD2, and SD5) as well as MD1. In addition, no child with DD scored significantly lower than Younger TD controls for VA span. Scores on WM were significantly lower than Older TD controls for 2 of the seven children that fit the phonological DD profile (PD2, and PD3) as well as 1 of the seven children that fit the surface DD profile (SD5). Only SD2 scored significantly lower than Younger TD controls for WM. Scores on VSSTM were significantly lower than Older TD controls for 1 of the seven children that fit the phonological DD profile (PD5). No children scored significantly lower than Younger TD controls for VSSTM.

In summary, three children (PD6, SD3 and SD6) showed no cognitive deficits relative to older TD controls; four children manifested with a single deficit (1 PA: PD4, 2 RAN: PD7 and SD4, 1 VA span: MD1), six children manifested with a double deficit (1 PA/ RAN: SD7, 1 PA/ VA span: PD3, 1 PA/ VSSTM: PD5, 2 RAN/ VA span: SD1, SD2, 1 VA span/ WM: SD5), one child manifested with a deficit in 3 domains (PA/ RAN/ VA span: PD1) and one child (PA/ RAN/ VA span/ WM: PD2) manifested with a deficit in 4 domains. Subsequently, 7 (46%) of the DD subgroup has selective deficits in RAN and VA span measures, 5 (33%) had deficits in PA, 3 (20%) had deficits in WM, and 1 (6%) had deficits in VSSTM.

TABLE 61: PREVALENCE OF PHONOLOGICAL, SURFACE, AND MIXED PROFILES

	Classic: Accuracy	Classic: RT	Classic: Nonword Accuracy/ Word RT	Regression: Older TD (Nonword Accuracy/Word RT)	Regression: Older TD (Nonword RT/ HF Word RT)
PD	3 20% 2.1%	1 6.7% 0.7%	3 20% 2.1%	7 47% 4.8%	3 20% 2.1%
SD	3 20% 2.1%	1 6.7% 0.7%	6 40% 4.1	7 47% 4.8%	4 27% 2.76%
MD	5 33% 3.5%	10 67% 6.9%	5 33% 3.5%	1 7% 0.7%	7 47% 4.8%
No Deficit	4 27% 2.76%	3 20% 2.1%	1 7% 0.7%	0	1 7% 0.7%

Table 61 above shows a summary of membership to different subgroups based on the different approaches used. Note: PD stands for Phonological subgroup, SD stands for Surface subgroup, MD stands for Mixed subgroup. The three numbers under each section represent the absolute number of cases, the relative percentage of cases and the absolute percentage of cases.

In a final consideration, the prevalence of DD within the current cohort was considered. While an accurate estimate of the prevalence of DD amongst Turkish-speaking children is beyond the scope of this study, there is utility in providing estimates to be further examined in future studies in reading development and disorder in Turkish. The overall prevalence of DD within this cohort was 10.34% using a 1.25 SD cut-off on scores of ORF. Using the regression procedure adopted by Niolaki, Terzopoulos and Masterson (2014), nearly half (47%) of the DD cohort manifested as Phonological Dyslexics, and nearly half (47%) of the DD cohort manifested as Surface Dyslexic. From Table 60, as above, the absolute estimates for each subtype of DD are between 0.7 and 4.8%. Considering only the seven children found in the initial cohort (as opposed to the children referred to the study), the prevalence of DD within this group was 5.11% using a 1.25 SD cut-off on scores of ORF. Furthermore, within this group, there were two (2) cases of Phonological DD (PD5 and PD6), four (4) cases of Surface DD (SD1, SD2, SD3, SD5) and 1 Mixed DD case (MD1). In terms of absolute percentages of the entire TD cohort, it can be deliberated that the prevalence of each subtype is 1.46% for Phonological DD, 2.92% for Surface DD and 0.73% for Mixed DD cases.

5.5 DISCUSSION

The current chapter explored the factors influencing reading, cognitive profiles and subtypes of DD in a group of Turkish-speaking children. In a similar vein to Chapter 4, reading was studied at both the word and text levels. As predictors, the effect of length, lexicality and frequency (and their interactions) were explored as well as considering the influence of a set of cognitive predictors that have been indicated to influence word and text reading such as phonological awareness, rapid naming, and visual attention span. In addition, the current chapter explored the presence of subtypes of DD within the current cohort.

5.5.1 GROUP COMPARISONS OF CHILDREN WITH DD AND OLDER AND YOUNGER TD CONTROLS

When the DD group was considered as a whole, the current investigation found evidence that children in the DD group were slower than TD children at reading at both the text and single word levels, although their word/pseudoword reading was relatively accurate. This finding is mostly congruent with results from a number of studies on transparent orthographies such as Greek (Nikolopoulos, Goulandris, & Snowling, 2003), Italian (Zoccolotti et al., 1999) and Spanish (Davis, Cuetos, & Glez-Seijas, 2007). Additionally, the DD group were less accurate than older TD children.

In addition, the current study found evidence of main effects of both lexicality and length with mixed evidence for frequency for both word/pseudoword reading accuracy and RT. The presence of length effects in a transparent orthography is considered to be reflective of the use of serial sublexical coding processes (Coltheart et al., 2001; Weekes, 1997) and is congruent with previous reports from Italian (Zoccolotti et al., 2005) and Spanish children (Davies et al., 2013). Furthermore, the current study found that word length effects were present in all three groups of children concerning their word/ pseudoword reading accuracy and RTs. Davies and colleagues (2013) propose that this finding is indicative of the role of sublexical processing in transparent orthographies as even the older TD children could not avoid the effect of word length on the time needed to utter words. In further consideration of word length effects, when the random intercept of length was removed from the random-effects model, a distinct group by length interaction was observed. This was seen to be suggestive that Older TD readers, manifested with a significantly reduced word length effect on their word reading RTs when compared to the DD group. That is, the use of or reliance on the sublexical coding route is sustained in children with DD relative to older TD readers. These findings lend support for H5 regarding the presence of sustained length effects in the DD cohort. Previous findings in other transparent alphabetic orthographies such as Albanian (Avdyli & Cuetos, 2012), Dutch (Verhoeven & Keuning, 2018), Italian (Burani et al., 2002; De Luca et al., 1999; 2002) and Spanish (Davies et al., 2013) provide further support for the current findings.

Following this, there was an observed main effect of Lexicality for word reading accuracy and RT. This finding is indicative of words being read more accurately and quickly than pseudowords and has previously been reported in Finnish (Hautala et al., 2012), Italian (De Luca et al., 2010) and Spanish (Davies et al., 2013). In addition, the group by lexicality interaction found in the word reading accuracy investigation was suggestive of statistically significant differences in accurately naming nonwords than words in the DD group when compared to both younger and older TD controls and broadly in line with previous studies in German (Landerl, Wimmer, Frith, 1997) and Italian children (Martelli et al., 2014). In line with previous research (e.g. concerning word reading in children learning to read in transparent orthographies, the observed length by lexicality interaction within the word RT data is suggestive of the availability of both the lexical and sublexical routes for visual word recognition and is one of the critical findings used in support of dual-route theories of reading aloud (Weekes, 1997). Also, the observed 3-way interaction of group by length by lexicality within the word reading RT data denoted that the greater length effect for nonwords was more marked in the DD group compared to older TD readers. Taken together, the main effects and interactions above suggest that although Turkish children with DD have both routes available to them, there is an overreliance on sublexical processing which results in slow word reading.

Regarding the mixed findings of word frequency effects amongst the current cohort, it is not clear if the same concerns, regarding the distribution of RT, explored in Chapter 4, are present here or if the observed null finding for frequency effects is indicative of the current state of word reading of the Turkish-speaking children within this cohort. For instance, the literature regarding word frequency effects in young readers of transparent orthographies presents a varied perspective. While there is some evidence of null frequency effects (e.g. Valle-Arroyo, 1989), the prevailing evidence suggests that, in fact, word frequency effects are reported in young readers of transparent orthographies (e.g. Barca et al., 2007; Burani, Marcolini, & Stella, 2002; Davies, Cuetos & Glez-Seijas, 2007). Furthermore, although small frequency effects have been previously reported in Turkish-speaking adults (I. Raman, 1999), there is an argument for the use of different word frequency metrics to fully understand the full extent of the complexity of frequency in agglutinating languages like Turkish (Bilgin, 2016). In a follow-up investigation, when the measure of RT was explored with an inverse Gaussian distribution, word frequency and group interactions revealed that the DD group was less sensitive to word frequency effects than TD controls indicating a reliance on the sublexical route. Taken together, the finding of a significant group by frequency interactions provides further supporting evidence for the availability of the lexical route for TD children but not the DD group. This finding is largely in line with previous reports of DD in Spanish-speaking children (Davies et al., 2013).

Another central topic addressed by the current study concerns the nature of the cognitive deficits associated with DD in Turkish-speaking primary school children. To this end, no group differences emerged when investigating reading accuracy and Inverse RT though PA was a significant predictor of word/ pseudoword reading accuracy and inverse RT across groups. In addition, when only word reading inverse RT were considered, RAN was the only significant predictor across groups. In a follow-up investigation, when the measure of RT was explored with an inverse Gaussian distribution, significant group differences emerged. For instance, consider both word/ pseudoword stimuli, there was a significant interaction between RAN and Group (DD vs Younger TD) indicating that RAN may be a more important predictor of letter string processing for the younger TD readers than for the DD group. Conversely, when only words were considered, there was a significant interaction between RAN and Group (DD vs older TD) indicating that RAN may be a more important predictor of word processing for the DD group than for older TD controls. All in all, when considering word/ pseudoword reading RT, RAN emerges as the only predictor of DD in Turkish-speaking children giving support to H4. In light of the literature, the results of the current study are largely congruent with previous studies of the importance of RAN to reading (Araújo et al., 2015; Song et al., 2016). However, the is also evidence that the relative influence of RAN in predicting reading fluency differs marginally between orthographies (Georgiou et al., 2016; Landerl et al., 2018; Moll et al., 2014; Vaessen et al., 2010; Ziegler et al., 2010).

5.5.2 Establishing DD subtypes in Turkish

As a further aim of the current study, the presence of subtypes of DD in a group of Turkish-speaking dyslexic children was explored. Using the dual-route theoretical framework, the current DD cohort was analysed using an approach implemented in Greek (Douklias et al., 2009; Niolaki, Terzopoulos, & Masterson, 2014) in order to investigate surface and phonological subtypes of DD in Turkish. Specifically, using nonword reading accuracy as a measure of sublexical processing, and word reading RT as a measure of lexical processing, the current study established distinct subtypes of DD in Turkish-speaking children. Overall, the outcomes of the current study lend support to the existence of distinct types of DD in Turkish. The presence of distinct subtypes in transparent orthographies is further supported by studies in Greek (Douklias et al., 2009; Niolaki, Terzopoulos, & Masterson, 2014), Italian (Zoccolotti et al. (1999) and Spanish (Jiménez, Rodríguez, & Ramírez, 2009). In conclusion, it is apparent that both surface and phonological subtypes of DD can be detected in transparent and opaque alphabetic orthographies (Hanley, 2017).

5.5.3 Prevalence of DD subtypes in Turkish

With reference to the experimental hypotheses of this study, the current investigation found supporting evidence for a reduced incidence of DD in Turkish speaking children relative to other orthographies. For instance, when the whole DD cohort (n = 15) was considered together, the incidence rate of 10.34% is largely in line with previously reported research in English (e.g. Snowling, 2000) and Danish (e.g. Elbro, Moller & Nielsen, 1995), both considered to be opaque orthographies. However, considering that the numbers in the DD group were artificially inflated due to referred cases, the correct approach regarding this line of enquiry would be only to consider cases found within the cohort study in Chapter 4. As such, the incidence of DD in this study was found to be 5.11% resulting in an acceptance of H1. Moreover, the current study is largely in agreement with the hypothesis of granularity and transparency (Wydell & Butterworth, 1999) since the incidence rate of phonological DD found in the present study (2.92%, when seven cohort cases were measured) is considerably lower than that found in studies of English-speaking children with DD (10%; Brunswick, McDougall, & de Mornay Davies, 2010) in line with H2. Further still, when considering the relative prevalence of each subtype, the current study is in congruence with previous reports in Spanish (Jiménez & Ramírez, 2002; Jiménez, Rodríguez, & Ramírez, 2009) in that the relative percentages of DD subtypes in Turkish and Spanish are highly similar. For instance, the current investigation found that the prevalence of Surface and Phonological subtypes were 57.1% and 28.6%, respectively. Jiménez, Rodríguez, & Ramírez (2009) reported prevalence rates of 45.7% and 22.8%, respectively. Similarly, Jiménez and Ramírez (2002) reported relative prevalence rates of 53% and 18% for surface and phonological subtypes of DD. The collective agreement of the above studies with the current investigation stands in contrast to studies carried out with English-speaking children. For example, in their seminal study, Castles and Coltheart (1993) report a relative prevalence rate of 30% for Surface DD and 55% for Phonological DD. Taken together, it appears that Phonological DD is less evident in transparent alphabetic orthographies such as Turkish and Spanish as predicted by the HGT (Wydell & Butterworth, 1999).

5.5.4 Cognitive Profiles of DD in Turkish

As a final consideration, the nature of the cognitive deficits associated with DD in Turkish-speaking children was considered using a multiple-case study design. To the best of the author's knowledge, this is the first reported study to systematically assess the different cognitive deficits among Turkish-speaking children with DD.

Generally, performance across the tasks used in this study was variable, with different children manifesting with deficits in tasks of phonological awareness, rapid naming, letter report, working memory and visuo-spatial short-term memory. Similar results have also been discovered for children with DD in several previous studies of DD in transparent orthographies (e.g. Brizzolara et al., 2006; Jimenez et al., 2009; Nikolopoulos, 1999; Tobia & Marzocchi, 2014). Furthermore, the current results lend support to the proposition that, at least in transparent orthographies, RAN and VA span are substantial reading-related cognitive deficits. Specifically, 46% of children with DD had selective deficits in RAN and VA span measures, respectively, 33% had deficits in PA, 3 20% had deficits in WM, and 6% had deficits in VSSTM.

The current study also investigated whether surface and phonological DD subtypes could be distinguished in terms of their underlying cognitive deficits. Previous studies suggest that phonological and surface DD subtypes manifest as developmental deviancy and lag, respectively (e.g., Manis et al., 1996; Stanovich et al., 1997). The findings of the current study lend support to this position as most children in the phonological DD subgroup performed poorer than younger TD on most measures of word/pseudoword reading accuracy and RT were as children in the surface DD subgroup were more reflective of a delay profile as they were not statistically different to younger TD in word/pseudoword reading performance. Finally, the findings of the current study suggest that the manifestation of DD in Turkish-speaking children is heterogeneous, and the majority of children in the DD subgroup exhibited either double or multiple deficits and therefore providing further support for the multiple-deficit hypothesis for Turkish developmental dyslexia.

5.6 LIMITATIONS

As with all experimental studies, there are a number of limitations to consider. Firstly, the present results are based on a cross-sectional design. While there is a need for this type of research design to better understand the role of these cognitive skills to word reading there is also a need for longitudinal studies in order to examine the relative contribution of these cognitive skills over time. Another notable limitation is the relatively small sample size of the current DD cohort. Future studies will need to recruit a larger representative sample of Turkish-speaking children with DD to explore several of the findings in the current chapter further. Separately, some of the constructs used within the current investigation were conceptualised using only one relevant measure, and therefore future studies will need to incorporate additional measures in order to increase construct validity (Landerl et al., 2013).

The mixed results regarding, and need for transformation of, the word frequency measures could be reflective of the complexity of word frequency measures in agglutinating languages such as Turkish. Consequently, there is a need to further define novel word frequency measures as the use of surface frequency may not be sufficient to characterise Turkish psycholinguistic data fully. Furthermore, there is a need to broaden the linguistic and cognitive domains under exploration. For example, reading comprehension, spelling and morphological awareness are excellent candidates for further investigation for predictors of DD in Turkish-speaking populations. For instance, the role of morphological awareness may be particularly important to investigate in agglutinative orthographies (Acha et al., 2010). In one of the few studies of morphological awareness in Turkish, Durgunoğlu (2003) proposed that the rich morphological structure of Turkish may be best addressed by the use of a left-to-right computational strategy. Additionally, Fowler, Feldman, Andjelkovic, and Öney (2003) suggest that phonological predictability could play a more crucial role than semantic relatedness in the acquisition of distinctive types of morphology. Finally, as this was a monolingual study, the extent to which the results can reflect comparability between the different orthographies is limited. For future studies, cross-language studies on DD are of particular importance.

5.6 CONCLUSION

The findings of this study contribute to a number of topics concerning the underlying cognitive and linguistic mechanisms of reading disorder in Turkish-speaking children. To the best of the author's knowledge, the current study represents the most comprehensive attempt to characterise reading disorder in Turkish at both group and individual levels. Taken together, for children with DD reading in Turkish, performance appears to be over-reliant on sublexical processes. Furthermore, the results obtained in the current study reveal that phonological awareness, rapid automatized naming, visual attention span, working memory and visuo-spatial short-term memory can all differentially contribute to the cognitive deficits associated with reading disorder in Turkish. In addition, it appears that RAN is essential for both decoding and to sight word reading and, along with VA span, were the most important cognitive predictors of DD in Turkish-speaking children learning to read. Considering the existence of subtypes of DD in Turkish, the current results show support for distinct profiles of developmental surface and phonological DD in Turkish speaking children. It is anticipated that by further identifying distinct cognitive profiles of DD, the goal of developing and applying screening and intervention measures that are tailored to the specific manifestation of each type of developmental dyslexia can be achieved.

The following chapter will introduce ongoing work aimed at developing a psychologically plausible computational model of visual word recognition in Turkish using the dual-route cascaded model (Coltheart et al., 2001).

CHAPTER 6: TOWARDS DEVELOPING COMPUTATIONAL MODELS OF VISUAL WORD RECOGNITION AND READING ALOUD IN TURKISH

6.1 PREFACE

In order to truly comprehend the complexity of visual word recognition, there is a need to consider the assortment of processes involved at both an individual level and as a whole. Over time, verbal theories are giving way to computational models that provide falsifiable predictions while also highlighting gaps in our theoretical understanding of complex phenomena. As such, computational modelling is fast becoming omnipresent within cognitive psychology. Further still, the recent movement beyond the monosyllabic constraints of older models, coupled with considerations of how learning could be plausibly incorporated into the reading process together highlight some interesting new directions in the field of computational modelling of visual word recognition and reading aloud.

This chapter aims to provide a review of the recent progress made in developmental computational models of visual word recognition while also highlighting contemporary theoretical considerations in computational modelling. Also, this chapter will introduce the groundwork for the ongoing development of a Turkish child version of the dual-route cascaded (DRC) model of reading aloud and word recognition. Finally, the newly created model will be evaluated in terms of fit to human reading data.

6.2 INTRODUCTION

Dating back to Morton's (1969) logogen model, models of visual word recognition have become essential tools for investigating the cognitive phenomena associated with both normal and disordered reading. Within his model, Morton (1969) introduced the concept of the logogen– a type of unit that activates during word recognition and contains information about the unique properties of a word including visual, phonological and semantic information (Besner & Swan, 1982). As such, the presentation of a word lowers the threshold of that word's logogen, consequently making it more accessible for future presentations. Due to these features, the logogen model was successful in simulating word frequency effects as high-frequency words would have a lower threshold than low-frequency words. However, the logogen model offered limited information about what was happening within the logogen system and the nature of information transmission through the model. According to Norris (2013), the development of computational models of visual word recognition in the early 1980s offered a novel approach to developing theoretical models.

Among the first and most influential of these models was the Interactive Activation (IA) model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Within the IA model, information from the visual stimulus moves through each of the feature, letter, and word levels of representation, i.e. “localist” structure. Each of the three levels of representation is made up of individual units or nodes, and the connections between the three adjacent levels of representation can be both excitatory and inhibitory. Additionally, information flows continuously (i.e. in a “cascade”; McClelland, 1979) through the levels of representation in a bidirectional manner. In contrast to the logogen model (Morton, 1969), information at one level of representation does not have to reach a threshold before being passed on to another level of representation. In order to select the word node candidate that provides the best fit to the letter string stimulus, the IA model implements a system of competition whereby inhibitory connections between word nodes enable the most active node to reduce the activation of alternative candidate nodes. Although the IA model represents an essential step in computational models of visual word recognition, several limitations have been noted (See Andrews & Davies, 1999). Most significantly is that because orthography was coded using four slots, the model was limited to processing four-letter stimuli. In addition, IA models generally cannot learn (though a recent attempt to overcome this limitation forms a central component of this chapter). The success of the IA model may be best marked by the success of its successors namely the dual-route cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; multiple readout model (MROM), (Grainger & Jacobs, 1996; Jacobs, Rey, Ziegler, & Grainger, 1998); and the CDP family of models (Perry, Ziegler, & Zorzi, 2007).

The following section will reintroduce three models of reading aloud, namely, the previously mentioned, dual-route cascaded model of visual word recognition and reading aloud (DRC) (Coltheart, Rastle, Perry, Langdon & Ziegler 2001), and the connectionist dual-process family of models (CDP, CDP+, CDP++) (Perry, Ziegler & Zorzi 2007) as well as the parallel distributed processing (PDP) model family (Seidenberg & McClelland 1989; Harm & Seidenberg 1999, 2004; Plaut, McClelland, Seidenberg & Patterson 1996). The motivation for presenting these three families of models is that they have been widely discussed in the computational modelling literature, and have been implemented as computational models, signifying that they are fully formulated. In addition, all three types of model have been highly successful in their attempt to simulate the cognitive phenomena associated with reading aloud single words (Hendrix, Ramscar, & Baayen, 2019). To provide a comprehensive account of these models however is beyond the scope of this chapter; instead, the sections below will present the underlying architecture of each of the models and will be focused on the role of word learning mechanisms, i.e. the development of orthographic knowledge.

Though the DRC was principally created to explain skilled reading performance, proponents of the model argue that the DRC can also be used to understand and evaluate the acquisition of the two routes that developing readers need to learn (Sheriston, Critten & Jones, 2016). Although the DRC is often criticized for being a static, non-learning model of skilled reading (Perry, Ziegler, & Zorzi, 2007; Seidenberg & Plaut, 2006; Snowling, Bryant, & Hulme, 1996), there is some evidence that static models may be informative regarding reading acquisition. For example, Castles and colleagues (2006), using simple regression modelling with a sample of 2136 children ranging from 6 to 15 years old, determined that a student's regular word reading performance could be accurately predicted by their irregular and nonword reading scores. The authors concluded that independent of a reader's age, the DRC model provides an exceptional account of children's reading performance. In support of this position, several studies have provided further evidence of the DRC models capability to simulate reading disorder (Jones, Castles, & Kohnen, 2011; Moore, Porter, Kohnen, & Castles, 2012). As stated previously, the DRC model posits that developmental phonological dyslexia is characterized by a specific difficulty in developing the nonlexical reading route, whereas developmental surface dyslexia is reflective of a specific difficulty in acquiring the lexical reading route (Castles & Coltheart, 1993).

More recently, in order to address the criticisms of the static nature of the DRC model, two novel learning mechanisms have been designed to supplement the DRC model (Pritchard, 2012). The first is grounded in an initial account of an algorithm for learning grapheme-phoneme correspondences (GPCs) (Coltheart, Curtis, Atkins, and Haller, 1993). Building on this, Pritchard and colleagues (2016) produced a GPC Learning (L-DRC) model that was able to effectively learn GPCs with mixed results. While the model was able to simulate word reading accuracy at a similar rate to the original DRC model (77.9% vs 83.7%), a number of issues concerning increased error rates when the model was trained with multi-morphemic words, when single- letter and multi-letter rules were learned in the same training phase, and the use of a token-based input corpus represent significant challenges for the L-DRC model (Pritchard, Coltheart, Marinus, & Castles, 2016). Aside from these challenges, the L-DRC model operation raises intriguing prospects regarding the incorporation of a DRC model for Turkish. Firstly, the L-DRC would allow for the creation of GPC rules for Turkish for the first time and would facilitate a computational exploration of Turkish visual word recognition. Secondly, the anticipated lack of multi-letter rules in Turkish would address one of the limitations stated above.

The second learning algorithm incorporated into the DRC concerns a computational implementation of the self-teaching hypothesis (Share, 1995). Share (1995) argues that beginning readers utilize their pre-existing knowledge of GPCs to ascertain the pronunciation of a newly encountered printed word and consequently create a scheme to self-teach a new orthographic representation. Proponents of the self-teaching hypothesis argue that orthographic learning is fundamental to reading. Pritchard and colleagues (2018) adopted a semi-supervised approach to learning, using the sub-lexical route to assist in training the lexical route. In a series of simulations, the findings validated the suitability of the Learning DRC (L-DRC) model in simulating self-teaching in order to build orthographic knowledge. However, the psychological plausibility of this absolute form of learning, where orthographic learning can occur after a single exposure to a word was raised by its creator (Pritchard, 2012). Still, the availability of such a learning mechanism with the DRC would allow for a finer examination of the verbal self-hypothesis theory, the exploration of vocabulary growth in developing readers as well as the potential to model age or grade-based accounts of reading development. The latter opportunity is of particular interest to the current investigation as it would allow for the development of both an average child model of the DRC in Turkish as well as grade-based accounts of reading. This implementation would allow for a further investigation of reading development and disorder in Turkish using a computational approach.

In summary, the DRC model provides a computational implementation of the dual-route theory of reading aloud. Its success is indexed by the model's ability to simulate a wide range of word reading phenomena observed in human readers as well as recent extensions of the model to differing orthographies. The recent incorporation of learning mechanisms opens up an exciting avenue of further investigation of reading development and reading disorder – which will be further explored in this chapter. However, before this, it is essential to consider alternative computational interpretations of the visual word recognition process, starting with another class of dual-route models, namely the Connectionist Dual Process (CDP) family of models.

One of the most recent implementations, the CDP++ model (Perry, Ziegler, & Zorzi, 2010) accounted for over 49% of the item-specific variance in naming latencies on the English Lexicon Project data using a selection of monomorphemic monosyllabic and disyllabic words (Yap & `, 2009). Specifically, the CDP++ model accounted for word length, frequency, neighbourhood, and consistency effects as well as disyllabic benchmark effects such as syllable number and stress regularity (Perry, Ziegler, & Zorzi, 2010). The CDP models have also been successfully extended to other alphabetic orthographies including French (Perry, Ziegler, & Zorzi, 2014a), German (Perry, Ziegler, Braun, & Zorzi, 2010) and Italian (Perry, Ziegler, & Zorzi, 2014b).

Criticism of the CDP model includes having a slow learning rate and the psychological implausibility of supervised learning being the only learning mechanism (Pritchard, 2012). This slow learning rate over voluminous trials is needed to avoid the risk of catastrophic interference (McCloskey & Cohen, 1989; Ratcliff, 1990)) where the connection weights that accurately reflect the initial information are altered by successive learning events in an incompatible way with the initial learning resulting in initially learned information being lost. Subsequently, delta rule training also undergoes catastrophic interference (Lewandowsky & Li, 1995), and requires many trials using a low rate of learning to minimize this interference. While progress is being made in addressing this limitation (see Parisi et al., 2019 for a recent review), the slow learning rate and risk of catastrophic forgetting pose challenges regarding the psychological plausibility of neural network accounts of visual word recognition. It is also argued that supervised learning is comparable to a beginning reader receiving direct instruction, which Share (1995) argues that direct instruction cannot be the primary avenue by which children acquire new orthographic knowledge. Though, the delta rule is used exclusively in the sublexical route of the CDP model, which would be more akin to systematic phonics teaching methods. This being said, the CDP sublexical training routine needs a large amount of training with whole words as input, rather than only explicit phonics training.

Despite the challenges faced by CDP models, the family of models have been highly successful in modelling reading development through self-teaching Ziegler, Perry & Zorzi, 2014) and Developmental Dyslexia (Perry, Zorzi, & Ziegler, 2019; Ziegler, Perry & Zorzi, 2014; Ziegler, Perry, & Zorzi, 2019). In a parallel development to the ST-DRC model, a computational version of the self-teaching hypothesis was also recently implemented within the CDP++model of reading aloud, (Ziegler, Perry, & Zorzi, 2014). Ziegler and colleagues (2014) simulated sublexical learning simultaneously with orthographic learning demonstrating how self-teaching of both routes might be bootstrapped from initial knowledge of only a few single-letter sublexical GPCs and spoken vocabulary, i.e. a developed phonological lexicon. In doing so, the study produced, arguably, the first developmentally plausible computational model of reading development and also subsequently addressed the inherent limitation of supervised learning, highlighted above. The overall model accuracy was 80% - given the inconsistent nature of English word pronunciation; this level of accuracy can be considered to be exceptional for a computational model of reading development.

Furthermore, Ziegler and colleagues (2014) demonstrated how this developmental reading model could be used to simulate reading disorder. For instance, in Simulations 4 and 5 of their study, Ziegler and colleagues (2014) determined that the model may explain how both visual and phoneme deficits affect orthographic development and nonword reading. In order to simulate visual deficits, the authors parametrically manipulated the probability that each letter in a word switched with the letter next to it leading to letter position errors. To simulate deficits in phonological awareness, the authors parametrically manipulated the probability that each time a correct word was activated in the phonological lexicon, the phonemes in the output of the TLA network were changed. While the simulations were promising, they were not compared to human developmental data. To address this, Perry, Zorzi, & Ziegler (2019) applied the newly created CDP++ developmental model to large-scale individual simulations of 622 children (62% dyslexic) to examine how the core deficits of developmental dyslexia could predict both individual learning outcomes and reading profiles. Using three component tasks, the study found that the models accurately simulated normal and impaired reading development. In terms of individual differences, models were found to capture human data variance between 63% and 72%, displaying a very good fit to data. Clearly, the application of models of reading development and disorder to large-scale data is a particularly attractive endeavour.

While it was initially intended to develop a CDP++ developmental reading model in Turkish within this chapter, the unavailability of the source code (C. Perry, personal communication, October 31, 2017) meant that this would not be feasible within the time frame of this thesis. The following section covers the final family of models, namely the Parallel Distributed Processing (PDP) or "triangle" models.

Regarding learning, the triangle models all use some form of the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986) to facilitate learning in the model. Backpropagation allows networks with multiple layers of processing units to be effectively trained, and such networks are more powerful than their older and simpler two-layer predecessors. Multi-layer networks employ intermediate layers of hidden units, which are not accessible to the external world. They receive activation from layers closer to the input and feed activation forward to layers at the output. However, similar to the criticisms of the delta rule used in CDP models, backpropagation is a form of supervised learning and has also been criticized as psychologically implausible due to a large amount of training that the triangle models need to endure to become skilled (Pritchard, 2012; Norris, 2006). The risk of catastrophic interference is mediated by setting the learning rate very low. As a consequence, the training corpus is typically presented for numerous epochs. For example, the model of reading described in Plaut et al. (1996) was trained with over 300 epochs.

Despite the challenges faced and in a similar vein to the other models described above, triangle models have also been applied to the simulation of reading development and disorder. For example, Powell, Plaut, and Funnell, (2006) evaluated the PMSP96 triangle model (Plaut et al., 1996) for use with behavioural, developmental data and found that like the child data, their model read more words than nonwords correctly at Time 1. However, at Time 2, the model read significantly more words than nonwords showing an opposite pattern of performance to the children. Furthermore, the types of errors made by the model were mostly exclusive to nonwords, whereas the developmental data suggested that errors were more common in word reading. Using a series of adaptations including incremental training on a frequency-based vocabulary and inclusion of GPCs during training, Powell, Plaut, and Funnell, (2006) were able to produce a near-exact match to the children's nonword reading performance.

Furthermore, Harm and Seidenberg's (1999) model (HS99) offers a connectionist account of developmental dyslexia subtypes. The model was first trained to develop phonological attractors which facilitated pronunciations for English words via interactive activation in the phonological network. Following this, a set of orthographic inputs was connected to the phonological network through a set of hidden units, and the back-propagation-through-time algorithm (see Werbos, 1990) was used to "teach" the model to generate the most suitable phonological outputs for the given orthographic input. To simulate a phonological reading deficit, Harm and Seidenberg (1999) added noise to the attractor units, which prohibited the model from producing efficient phonological representations. In contrast to the unimpaired model, which learned to read all orthographic input correctly, the impaired model successfully simulated phonological dyslexia - manifesting with difficulties in reading nonwords. In order to simulate surface dyslexia, conceptualized by Harm and Seidenberg (1999) as reading delay dyslexia, the model slowed down reading acquisition in several ways. These included less training and therefore simulating a lack of reading experience as well as decreasing the learning rate, degrading the orthographic input, and removing some of the hidden units (Peterson, Pennington & Olson, 2013). In each iteration, the HS99 model manifested with a pattern of relative surface dyslexia, that is, an impairment in both irregular word and nonword reading but could not simulate pure surface dyslexia. Peterson, Pennington & Olson (2013) contest that while the HS99 model conceptualizes surface dyslexia as a reading delay, several reading profiles in the study did not reflect word reading delays.

Ultimately, the competition between the computational approaches described above has progressed the understanding of the theoretical frameworks surrounding visual word recognition, reading acquisition and disorder. While the common driver of current computational models of reading has been the English language (e.g. Coltheart et al. (2001); Perry et al. (2010); Plaut et al. (1996)), there is a need to further consider various orthographies and writing systems in a move towards more universal conceptualizations of visual word recognition (Frost, 2012). Furthermore, while a comparison of several models would be beneficial, the lack of availability of the CDP source code coupled with the inability of triangle models to simulate pure types of dyslexia naturally dictates that a more singular approach is warranted within this thesis. As such, the DRC model will be used for the current purposes in providing a model for simulating typical reading as well as surface and phonological developmental dyslexia within the Turkish orthography. To this end, the following section describes the implementation of the

DRC model for use with Turkish-speaking children.

6.3 TURKISH CHILD DRC (DRC-TR-CHILD) MODEL CREATION

6.3.1 MODEL ARCHITECTURE

The architecture of the Turkish version of the DRC model is identical to that of the English version 1.2.3 (http://www.cogsci.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/) with a number of modifications. First, a new set of visual letter features was developed to reflect the properties of the Turkish alphabet. Using the 14-feature uppercase-letter font (Rumelhart and Siple, 1974), one additional critical feature was added to the letters C, G, O, S, U to make the letters Ç, Ğ, Ö, Ş Ü. Also, one critical feature was removed from İ to make the letter I. Finally, Q, X and W were removed from the letters list. In the lexical route, both the orthographic and phonological lexicons of the Turkish model contained 5000 words and their associated frequencies. The rationale for this was to produce a mental lexicon that reflected the average size vocabulary of an average Turkish-speaking child. To this end, while research into vocabulary size in Turkish-speaking children is largely absent, the choice to include 5000 words was a pragmatic one, balancing the need for psychological plausibility and computational capacity.

Additionally, in the non-lexical route, the English GPC rules were replaced by the Turkish rule set that was learned (See the section below for an overview of GPC rule learning integration into the current model. Further adjustments included:

1. increasing the number of units in each position of the visual feature and letter layers to 30 — one for each of the 29 Turkish letters, and an extra unit for the 'blank',
2. increasing the number of units in each position of the phoneme layer to 40 — one for each of the 39 Turkish phonemes, and one for the 'blank', and
3. increasing the number of positions in each of the visual feature, letter, and phoneme layers to 13 to accommodate long Turkish words.

Fundamentally, following the above changes, the Turkish DRC could handle both mono and polysyllabic words correctly, including stress assignment, as Turkish has a simple stress assignment rule that places primary stress on the final syllable of a word irrespective of the length of the word and weight of the syllables (Sezer, 1983) though instances of non-final stress exist. All following simulations were carried out with a PC Intel(R) Core (TM) i7-4770 CPU @ 3.4 GHz with 8.0 GB of RAM.

6.3.2 A COMPUTATIONAL TURKISH MODEL OF SUBLEXICAL ROUTE GPC LEARNING

In his doctoral thesis, Pritchard (2012) implemented a model that could provide a psychologically plausible computational account of GPC learning. Similar to the DRC model (Coltheart et al., 2001), GPC learning is based on the concept of rule learning as opposed to statistical learning. The model's procedure, also implemented within this chapter, involves two stages: First, an information-gathering stage, and secondly a rule consolidation stage. In the information-gathering stage, the model is presented with individual written word-spoken word pairs, and the model attempts to identify graphemes and GPCs in each input while also gathering information about GPC occurrence frequency. The learning model implemented by Pritchard and colleagues (Pritchard, 2012; Pritchard et al., 2016) classifies input into three types, namely, 1. number of letters equals the number of phonemes; 2. Number of letters is less than number of phonemes and 3. Words containing the letter X. The nature of written Turkish dictates that the only type of interest for this research is the first type. The result of this stage is a list of candidate GPCs, with an attached frequency of occurrence.

Following this, in the rule consolidation stage, the model inspects the list of candidate GPC and makes changes, including the removal of GPCs that are infrequent, i.e. low-frequency GPCs. The threshold for determining low-frequency cut-ff is set by the experimenter as a parameter choice. The subsequent step is the modification of GPCs that apply to the same grapheme by adapting them to form context-sensitive rules. In the formation of context-sensitive rules, GPCs are grouped into small sets organised by grapheme. For each small set, the GPC with the highest frequency of occurrence is regarded as the default. For the other GPCs in each set, if their frequency relative to the GPC with maximum frequency is less than a value determined by the experimenter and specified as a parameter choice, then they are dropped from the list. If any of the non-highest-frequency rules in a set are above the relative frequency cut-off, then the model will take the one with the highest frequency and attempt to form a context-sensitive rule for it. To do this, the model will go back and loop through the full list of inputs in the input corpus, looking for the instances when the GPC under consideration was identified. Whenever it finds an input where this is the case, it will take note of the letters that precede and follow the letter comprising the GPC and record how many times it notices the GPC with particular preceding and following letters. After doing this, the model will have a list of preceding and following letters, with a frequency count for each letter. If one particular letter is seen to precede or follow the rule with a frequency that dominates the frequency of the other preceding or following letters, then a context-sensitive rule is created. In the final phase, the newly learned rules were extrapolated by separating matching grapheme-phoneme pairs based on position, i.e. beginning of a word (the first grapheme), end of a word (the last grapheme), and the middle of a word (all other positions). For the learning DRC, the same grapheme corresponding to the same phoneme but in a different position is taken as a discrete GPC.

In order to model GPCs in Turkish, the METUbet phonetic transcription (Salor, Pellom, Çiloğlu, Hacıoğlu & Demirekler, 2002), described in detail in Chapter 2, was re-employed as the phoneme system for the Turkish DRC resulting in a 39-phoneme implementation. While the Grafofon phonetic system would have been a useful resource for this investigation, it was unavailable at the time of initial model development (O. Koşaner, personal communication, September 20, 2016).

A series of exploratory simulations were conducted to establish a working set of parameter values for the learning algorithm (Table 62). From the table below, Simulations 1,2 and 3 provided the best results in terms of the percentage of words named correctly. In order to differentiate between the models, their ability to accurately name the 5000 words was also assessed using only the sublexical route.

TABLE 62: SIMULATION PARAMETERS AND RESULTS FOR MODELS 1 TO 10.

	Parameters			Rules Learned					
Model	Absolute	Relative	Minimum	Single	Multi	cs	output	Total	Accuracy
1	1	0	1	37	0	16	0	53	99.24%
2	2	0	1	37	0	13	0	50	99.24%
3	3	0	1	35	0	13	0	48	99.24%
4	1	0	2	37	0	5	0	42	99.12%
5	2	0	2	37	0	3	0	40	99.12%
6	3	0	2	35	0	3	0	38	99.12%
7	1	1	1	37	0	0	0	37	98.62%
8	2	1	1	37	0	0	0	37	98.62%
9	3	1	1	35	0	0	0	35	98.62%
10	1	0.5	1	37	0	13	0	50	99.24%

The best results (Model 2) in terms of the percentage of words named correctly using the learned GPCs were obtained by using the following parameter settings:

Minimum absolute frequency threshold: 2
Minimum relative frequency threshold: 0
Minimum contextual dominance: 1

The resulting GPC rules using the above parameter settings were then used in all subsequent simulations. Briefly, the Turkish GPC route was best characterised by 50 rules: 37 single letter rules and 13 context-sensitive rules. A full list of the rules generated is provided in Appendix 5.

6.3.3 Model Optimisation 1: Average Child DRC

Perhaps the most significant changes needed to the original DRC model were a consideration of the parameters that drive the behaviour of the model. As there are 25 parameters in the model (excluding five noise and four decay parameters), model optimisation moves through 25-dimensional parameter space. This dimensional space, therefore, represents a vast potential area to explore. In order to reduce the computational load and time of exploring the dimensional space, a series of model optimisation decisions were taken.

First, the 25 parameters were reduced to nine parameters. A full consideration of these parameters is given below. This decision was based on theoretical motivations considering orthographic transparency, conversations with one of the DRC models principle designers (S. Saunders, Personal Communication, 5 July 2019) and the parameter changes made to the recent implementation of the Greek DRC model (Kapnoula et al., 2017). Though the Greek DRC was not explicitly designed to provide an optimised model, the movement through parameter space was systematic and as such a similar approach was adopted in the current chapter. In addition, a similar approach was adopted in parameter value selection where parameters were adjusted using a process of trial-and-error until the final model produced no errors on any of the 5000 vocabulary words in the lexicon, or any of a set of 500 nonwords, with the minimum reading phonology set to 0.9. As a consequence, a substantial penalty was applied to incorrect model responses so that the optimisation procedure would reject parameter sets that did not generate perfect accuracy. The following is a list of parameters that were modified in the current modelling work:

1. Letter to Orthographic lexicon Excitation: This parameter determines the excitatory strength of the connection from the letter layer to the orthographic lexicon. It provides a means of determining the relative balance of lexical and sublexical processing. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.02 and 0.07.

2. Letter to Orthographic lexicon Inhibition: This parameter determines the inhibitory strength of the connection from the letter layer to the orthographic lexicon. It provides a means of regulating or repressing stimuli-incompatible words in the orthographic lexicon. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.30 and 0.40. Values above 0.4 up to 0.9 led to identical predictor strength and as such was considered to be redundant.

3. Orthographic lexicon to phonological lexicon excitation: This parameter controls the feedforward of activation from the orthographic lexicon to the phonological lexicon. This

excitatory connection is involved in producing both frequency and late neighbourhood density effects in the DRC model. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.10 and 0.30.

4. Orthographic lexicon to Letter layer excitation: This parameter is a feedback excitation from the orthographic lexicon back to the letter layer. This excitatory connection is involved in early neighbourhood density effects. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.10 and 0.30.

5. Orthographic lexicon lateral inhibition: This parameter determines the level that units within the orthographic lexicon, i.e. words mutually inhibit each other. Within this layer, lateral inhibition is considered to be homogeneous in that each unit inhibits any other unit in the lexicon to the same degree, regardless of its identity. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.00 and 0.20.

6. Phonological lexicon lateral inhibition: This parameter determines the level that units within the phonological lexicon mutually inhibit each other. Within this layer, lateral inhibition is considered to be homogeneous in that each unit inhibits any other unit in the lexicon to the same degree, regardless of its identity. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.00 and 0.20.

7. GPC Onset: This parameter controls how many cycles pass before the first letter supplies activation to the nonlexical route and therefore also helps balance the relative strengths of the lexical and nonlexical routes. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0 and 20.

8. GPC Critical Phonology: This parameter determines the level of activation required to move serially on to the next letter when any phoneme in the right-most phoneme unit is

excited in the previous cycle. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.01 and 0.30.

9. GPC Phoneme Excitation: This parameter denotes the strength of the input from the GPC system to the phoneme system. The higher the GPC phoneme excitation value, the faster the activation builds within the phoneme system. Exploratory navigation of the parameter space established that the edges of the parameter values for the Turkish dataset were between 0.02 and 0.06.

To give a numerical indication of the computational capacity and time needed to move through a 9-parameter dimensional space, consider the following equation:

Number of potential model parameter sets
= Parameter Number (1 x 2 x 3 x 4 x 5 x 6 x 7 x 8 x 9)
= Parameter Space (6 x 11 x 21 x 21 x 21 x 21 x 21 x 21 x 5)
= 28,302,819,930

In addition, there were additional ad-hoc modifications to several other parameters, including adjusting the phoneme to phonological lexicon excitation value in the averaged model to 0.02. For the grade 2 and 3 models, phoneme to phonological lexicon excitation was set at 0.01, and the frequency scale parameter was set to 0.10. For the grade 4 model, phoneme to phonological lexicon excitation was set at 0.01, phoneme to phonological lexicon inhibition was set to 0.15, phoneme lateral inhibition was set to 0.15, and the frequency scale parameter was set to 0. For the grade 5 model, phoneme to phonological lexicon excitation was set at 0.01, phoneme to phonological lexicon inhibition was set to 0.1, phoneme lateral inhibition was set to 0.1, and the frequency scale parameter was set to 0.1.

Evidently, moving through the parameter space in a single step represents unreasonable computational times. As such a further modelling decision was made to move through the parameter space in a series of steps. First, the Letter layer parameters were moved through followed by the orthographic lexicon layer, the GPC layer and finally the phonological lexicon layer. While this presents an unavoidable limitation of the current work, confidence was taken from the resulting goodness of fit evaluations between the model and the RT data.

6.4 Turkish Child DRC (DRC-TR-Child) Grade-based Model Creation

Following on from the development of an average Child DRC model for Turkish, the subsequent line of enquiry revolved around the development of averaged child DRC models based on grade. In order to achieve this, the self-teaching algorithm incorporated into the L-DRC model (Pritchard, 2012) was utilised to simulate varying vocabularies with grade through orthographic learning. All the following created models used the same architecture as the average Turkish DRC model mentioned above.

To effectively simulate self-teaching in the DRC model, the L-DRC triggers activation in the visual feature and letter layers upon presentation of novel stimuli. Similar to the DRC-1.2.1, the sublexical route then generates a pronunciation according to GPC rules based on the letter layer activation. Subsequently, the phoneme layer is activated, leading to the interactive excitation of the phonological lexicon. Overall, this process can simulate “phonological recoding” through the activation of a spoken word representation without any prior knowledge of a written word representation in the orthographic lexicon (Pritchard, 2012). To replicate this in Turkish, a number of modelling decisions were made. First, the L-DRC introduced a number of new parameters associated with learning based on threshold levels of activation. The new parameters considered in this chapter were SpokenWordRecognisedThreshold, WrittenWordRecognisedThreshold and WrittenWordFrequencyMultiplier. While there were additional parameters related to context and semantics, they were switched off (set to zero) as they were beyond the theoretical scope of the current modelling work. It was also decided to leave the three parameters of interest with their default values. Further explorations of self-teaching in Turkish can better utilise these parameters in future studies.

Starting with a vocabulary of 5000 words for Grade 2, each subsequent grade was exposed to an additional 10000 words. This figure comes from a simple estimate of vocabulary growth in Turkish-speaking children learning to read in primary school. While there were no available vocabulary growth studies available in Turkish children, the growth rate of 10,000 new words per year comes from crude estimates based on previous studies with children. For example, it has been suggested that between the 3rd and 9th grades, English-speaking children learn approximately 3,000 new words per year (Nagy & Anderson, 1984). Additionally, Anglin (1993) found that in Grade 1, vocabulary size was about 3,100 root words to about 7,500 root words in Grade 5. Similarly, Biemiller and Slonim (2001) reported that in 2nd grade, the mean vocabulary size was 5,200 root words, increasing to roughly 8,400 root words by 5th grade. Given the productivity of Turkish morphology, it is feasible to assume that the vocabulary of Turkish children would be significantly higher than that of speakers in other non-agglutinative languages though the lack of empirical studies means that this area of research warrants further investigation. This was balanced against a need to reduce the size of the lexicons in the model and 10,000 new words per year was the conservative approximation of this balance. The size of each Grade-based DRC model is given in table 63 below.

TABLE 63: SUMMARY OF THE SELF-TEACHING MODELS CREATED TO SIMULATE GRADE-BASED DRC MODELS

Grade	Words presented	Words learned	% words learned	Difference
2	5000	3593	72%	-
3	15000	10612	71%	7019
4	25000	17817	71%	7205
5	35000	24784	71%	6967

6.4.1 Model Optimisation 2: Grade-based Child DRC

Following the creation of the four new models, described above, each model was optimised by moving through the same parameter space as the averaged Turkish DRC model. However, the criteria for model acceptance was relaxed in that the requirement to read all of the words in the lexicon and a set of 500 nonwords were removed with only the need to read the 80 words/ nonwords stimuli to perfect accuracy being maintained. Each grade-based model was compared to the averaged RT scores form the relevant grade.

6.5 TURKISH CHILD DRC (DRC-TR-CHILD) AVERAGE AND INDIVIDUAL PROFILES OF DD MODEL CREATION

In a final modelling consideration, developmental dyslexia (DD) was investigated in the context of the Turkish DRC model of reading. While previous computational investigations of acquired dyslexia using the DRC model have simulated surface and phonological dyslexia by introducing lesions to the lexical or sublexical route (Coltheart et al., 1996, 2001; Nickels et al., 2008), there is a need to make a distinction concerning the nature of developmental and acquired dyslexia. Several criticisms have been aimed towards the use of the adult DRC models use in interpreting developmental data (e.g. Snowling, 1983; Snowling, Bryant & Hulme, 1996). In order to mitigate against these criticisms, a similar approach was adopted to previous DD modelling attempts using the DRC framework (Ziegler et al., 2008; Ziegler, 2011) in that, the word/nonword reading profile of each child with DD was simulated by adding five noise parameters to the model parameter space search, following optimisation, at the Letter, Orthlex, Phonlex, Phoneme and GPC levels. Participant-based modelling approaches are relatively novel to the field of developmental dyslexia, and the use of noise parameters offers an alternative implementation of disorder into the reading system.

Each child-specific model started with the relevant Grade-based DRC model and then went through a series of optimisation steps to improve the variance accounted for the model. From Ziegler et al., (2008), noise was introduced into the model at each representational level and then added to the net input of the relevant unit (Equations 3 and 6 from Coltheart et al., 2001, pages 215–216). This method was reiterated for each unit at each processing cycle of the model. Using this approach, 17 models (one for each DD case described in Chapter 5 and two averaged DD subtype (phonological and surface) models were created.

The aim of this computational modelling work is three-fold: First to evaluate the utility of the DRC model framework for Turkish-speaking children's data, second to explore further the DD profiles found in the empirical portions of this thesis and third, to assess the suitability of current subtyping approaches to Turkish. All models were evaluated using both a factorial and regression approach. The factorial approach consists of analysing RTs using analyses of variance (ANOVA) to evaluate if the effects found in the human data were also significant in the model data. The regression method involves predicting item-level variance (Spieler &

Balota, 1997; Besner, 1999) and includes computing the proportion of variance (R^2) in human RTs that is accounted for by the model. Further optimisation of models was carried out and is covered further down in this chapter after further modelling approaches had been undertaken.

In an attempt to align the current computational study with contemporary practices within the literature, it was also decided to add variables of onset phoneme articulation (Balota et al., 2004). To achieve this, variables for voicing, place of articulation (i.e., labial, labiodental, palatoalveolar, alveolar, palatal, velar and glottal) and manner of articulation (i.e. a stop, fricative, approximant, trill and nasal) were dummy coded. For vowel phonemes, Turkish phonology stipulates that there are three dimensions to consider representing six further dummy codes (Front/ Back, Rounded/ Unrounded and Open/ Closed).

6.5.1 Model data collection

The stimuli used in Chapters 4 and 5 were provided as input to the model, and the phonological output of the model was recorded. Rather than accuracy measures, RT was the variable of interest in all modelling approaches, and as such, the number of processing cycles was used as a proxy measure for RT in the single word/pseudoword naming task. Now that the modelling approach has been outlined, the following section describes the findings of the simulations described above in the context of their fit to the relevant human RT data.

6.6 Word/ Nonword reading simulations

6.6.1 Simulation 1: Performance of the averaged Turkish DRC model (Model 1)

The first step in model selection was moving through the parameter space to identify models that could read all 80 words/ nonwords correctly as well as name all words in the 5000-word lexicon and 500 nonwords correctly when the *MinReadingPhonology* parameter was set to 0.9 to simulate casual reading. To this end, over 4000 parameter sets could successfully achieve this goal. Following this, the *MinReadingPhonology* parameter was set to 0.4 to simulate speeded reading, and the parameter space between the 4000 working models was explored.

The best performing Turkish DRC model pronounced all 5000 words in the lexicon and all but two of the 500 nonwords (0.4%) correctly. The two incorrectly named nonwords were both lexicalisation errors, i.e. nonwords being read as words. Appendix 6 provides a list of the parameter values of Model 1. The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.806; $p < 0.001$). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 68.1% of the variance, R^2=.681, F (10, 69) = 16.32, p<.0001 with the DRC model predicting 49.1% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 64.5% of the variance in children's RT, F (1, 78) = 144.45, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 64) and Length and Frequency (Table 65) as predictors were also carried out.

TABLE 64: LINEAR REGRESSION ANALYSIS OF AVERAGED CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	148.74***	345.69***
Lexicality	0.733***	0.926***
Length	0.508***	0.206***
R^2	78.90%	89.70%

From Table 64, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model overfitted lexicality effects and underfitted length effects. The DRC model also accounted for 10.80% more variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 65: LINEAR REGRESSION ANALYSIS OF AVERAGED CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	42.00***	2.00
Length	0.810***	0.313^
Frequency	-0.130	0
R^2	67.80%	4.9%

From Table 65, the DRC model provided a poor fit of the word RT data in its ability to capture both length and frequency effects. From the beta values reported, the DRC model underfitted length effects. The DRC model also accounted for 62.90% less variance than the RT data suggesting an overall underfitting of the DRC model. The failure of the DRC model to account for length effects in known words is anticipated as parallel processing through the lexical route explains the model's insensitivity to word length effects.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 66 and 67 below.

TABLE 66: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN AVERAGED CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	194.62***	0.911	742.05***	0.975
Length	144.35***	0.884	38.00***	0.667
Lexicality by length interaction	10.74**	0.361	8.32**	0.305

From Table 66, the DRC model provided an excellent fit of the RT data in its ability to capture both length and lexicality effects and their interaction. However, from the beta values reported, the DRC model again overfitted lexicality effects and underfitted length effects.

TABLE 67: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN AVERAGED CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	.876	0.089	.016	0.002
Length	199.06***	0.957	4.45	0.331
Frequency by length interaction	0.023	0.003	3.79	0.305

From Table 67, the DRC model provided a poor fit of the word RT data in its ability to capture both length effects. From the beta values reported, the DRC model underfitted both length and frequency effects while overfitting their interaction. Again, the failure of the DRC model to account for length effects in known words is anticipated as parallel processing through the lexical route explains the model's insensitivity to word length effects.

6.6.2 Simulation 2: Performance of the Grade-based models (Models 2, 3, 4 and 5)

In a similar line of enquiry, development of Grade-based models proceeded by moving through the parameter space to identify models that could read all 80 words/ nonwords correctly though the criteria to name all words in the lexicon and the 500 nonwords correctly was removed. Following this, the *MinReadingPhonology* parameter was set to 0.4 to simulate speeded reading, and the parameter space between the working models for each grade were explored.

Model 2: Grade 2

The best performing Turkish DRC Grade 2 model pronounced all but 16 of the 3593 words (0.43%) in the lexicon and all of the 500 nonwords correctly. The 16 incorrectly named words were all regularisation errors. Appendix 7 provides a list of the parameter values of Model 2. The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.728; $p < 0.001$). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 61.2% of the variance, R^2=.612, F (9, 70) = 13.47, p<.0001 with the DRC model predicting 33.1% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 52.4% of the variance in children's RT, F (1, 78) = 88.14, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 68) and Length and Frequency (Table 69) as predictors were also carried out.

TABLE 68: LINEAR REGRESSION ANALYSIS OF GRADE 2 CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	66.21***	275.97***
Lexicality	0.676***	0.876***
Length	0.419***	0.331***
R^2	62.30%	87.40%

From Table 68, the DRC Grade 2 model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model again overfitted lexicality effects and slightly underfitted length effects. The DRC model also accounted for 25.10 % more variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 69: LINEAR REGRESSION ANALYSIS OF GRADE 2 CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	15.56***	2.46
Length	0.666***	0.339*
Frequency	-0.118	-0.049
R^2	42.70%	7.0%

From Table 69, the DRC model provided a poor fit of the word RT data. While the model was able to capture both present length and absent frequency effects, the overall model was nonsignificant. From the beta values reported, the DRC model underfitted length effects. The DRC model also accounted for 35.70% less variance than the RT data suggesting an overall underfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 70 and 71 below.

TABLE 70: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN GRADE 2 CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	78.72***	0.701	237.72***	0.926
Length	44.58***	0.806	1016.03***	0.982
Lexicality by length interaction	0.204	0.011	81.58***	0.811

From Table 70, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, while there was no significant interaction in the RT data, there was a significant interaction in the DRC model. From the beta values reported, the DRC model overfitted both lexicality and length effects.

TABLE 71: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN GRADE 2 CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	.772	0.079	.010	0.011
Length	49.24***	0.845	4.54	0.335
Frequency by length interaction	0.581	0.061	2.10	0.189

From Table 71, the DRC model provided a poor fit of the word RT data in its ability to capture both length effects. From the beta values reported, the DRC model underfitted both present length and absent frequency effects while overfitting their interaction.

Model 3: Grade 3

The best performing Turkish DRC Grade 3 model pronounced all of the 10612 words in the lexicon and all but 8 of the 500 nonwords (1.6%) correctly. The eight incorrectly named nonwords were made of 6 regularisation errors and two LOWAC errors. Appendix 8 provides a list of the parameter values of Model 3. The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.715; $p < 0.001$). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 59.7% of the variance, R^2=.597, F (9, 70) = 3.39, p=.002 with the DRC model predicting 38.3% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 50.4% of the variance in children's RT, F (1, 78) = 81.42, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 72) and Length and Frequency (Table 73) as predictors were also carried out.

TABLE 72: LINEAR REGRESSION ANALYSIS OF GRADE 3 CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	58.67***	136.54***
Lexicality	0.658***	0.806***
Length	0.413***	0.362***
R^2	59.30%	77.40%

From Table 72, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model overfitted lexicality effects and underfitted length effects. The DRC model also accounted for 18.10% more variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 73: LINEAR REGRESSION ANALYSIS OF GRADE 3 CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	16.00***	1.86
Length	0.681***	0.237
Frequency	-0.023	0.187
R^2	43.50%	4.2%

From Table 73, the DRC model provided a poor fit of the word RT data in its ability to capture length effects. From the beta values reported, the DRC model underfitted present length effects and overfitted absent frequency effects. The DRC model also accounted for 39.30% less variance than the RT data suggesting an overall underfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 74 and 75 below.

TABLE 74: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN GRADE 3 CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	74.36***	0.796	373.72***	0.952
Length	27.15***	0.588	97.50***	0.837
Lexicality by length interaction	0.002	0	61.75***	0.765

From Table 74, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, the DRC model reported a significant lexicality by length interaction when none was present in the RT data. From the beta values reported, the DRC model again overfitted both lexicality and length effects.

TABLE 75: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN GRADE 3 CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	.037	0.004	1.42	0.137
Length	89.93***	0.909	2.35	0.207
Frequency by length interaction	0.018	0.002	1.84	0.170

From Table 75, the DRC model provided a poor fit of the word RT data in its ability to capture length effects. From the beta values reported, the DRC model underfitted length effects and overfitted frequency effects while also overfitting their interaction. Again, the failure of the DRC model to account for length effects in known words is anticipated as parallel processing through the lexical route explains the model's insensitivity to word length effects.

Model 4: Grade 4

The best performing Turkish DRC Grade 4 model pronounced all of the 17817 words in the lexicon and all of the 500 nonwords correctly. The eight incorrectly named nonwords were all regularisation errors. Appendix 9 provides a list of the parameter values of Model 4. The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.807; p < 0.001). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 62.4% of the variance, R^2=.62.4, F (9, 70) = 14.13, p<.05 with the DRC model predicting 47.7% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 64.7% of the variance in children's RT, F (1, 78) = 145.53, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 76) and Length and Frequency (Table 77) as predictors were also carried out.

TABLE 76: LINEAR REGRESSION ANALYSIS OF GRADE 4 CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	176.69***	95.58***
Lexicality	0.840***	0.703***
Length	0.340***	0.468***
R^2	81.60%	70.50%

From Table 76, the DRC model provided an excellent of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model overfitted length effects and underfitted lexicality effects. The DRC model also accounted for 11.10% less variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 77: LINEAR REGRESSION ANALYSIS OF GRADE 4 CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	7.39***	17.84
Length	0.457***	0.689***
Frequency	-0.277^	-0.126
R^2	24.70%	46.3%

From Table 77, the DRC model provided a poor fit of the word RT data as the overall model was not significant. The model was, however, able to capture both present length and absent frequency effects. From the beta values reported, the DRC model overfitted length effects. The

DRC model also accounted for 21.60% more variance than the RT data suggesting an overall overfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 78 and 79 below.

TABLE 78: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN GRADE 4 CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	119.88***	0.863	544.71***	0.966
Length	84.34***	0.816	94.78***	0.833
Lexicality by length interaction	2.40	0.112	71.03***	0.789

From Table 78, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, the DRC model reported a significant interaction when none was present in the RT data. From the beta values reported, the DRC model again overfitted lexicality and length effects.

TABLE 79: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN GRADE 4 CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	1.11	0.110	6.53*	0.420
Length	29.64***	0.767	10.35**	0.535
Frequency by length interaction	0.281	0.030	3.50	0.280

From Table 79, the DRC model provided a poor fit of the word RT data in its ability to capture frequency effects. From the beta values reported, the DRC model underfitted length effects and overfitted frequency effects.

Model 5: Grade 5

The best performing Turkish DRC Grade 5 model pronounced all but 25 of the 24784 words (0.10%) in the lexicon and all of the 500 nonwords correctly. The 25 incorrectly named words were all regularisation errors. Appendix 10 provides a list of the parameter values of Model 5. The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.899; $p < 0.001$). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 80% of the variance, R^2=.800, F (9, 70) = 32.65, p<.0001 with the DRC model predicting 65% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 80.6% of the variance in children's RT, F (1, 78) = 329.60, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 80) and Length and Frequency (Table 81) as predictors were also carried out.

TABLE 80: LINEAR REGRESSION ANALYSIS OF GRADE 5 CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	169.40***	163.71***
Lexicality	0.834***	0.751***
Length	0.344***	0.496***
R^2	81.00%	80.50%

From Table 80, the DRC model provided an excellent fit of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model overfitted length effects and underfitted lexicality effects. The DRC model also accounted for 0.50% less variance than the RT data.

TABLE 81: LINEAR REGRESSION ANALYSIS OF GRADE 5 CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	29.09***	8.05***
Length	0.781***	0.483***
Frequency	-0.028	-0.263
R^2	59.00%	26.40%

From Table 81, the DRC model provided a good fit of the word RT data in its ability to capture both present length and absent frequency effects. From the beta values reported, the DRC model underfitted length effects and overfitted frequency effects. The DRC model also accounted for 32.60% less variance than the RT data suggesting an overall underfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 82 and 83 below.

TABLE 82: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN GRADE 5 CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	300.66***	0.941	594.01***	0.969
Length	110.63***	0.853	101.70***	0.843
Lexicality by length interaction	24.84***	0.567	87.78***	0.822

From Table 82, the DRC model provided an excellent fit of the RT data in its ability to capture both length and lexicality effects and their interaction. However, from the beta values reported, the DRC model provides an excellent fit for the main effects while overestimating the effect of the lexicality by length interaction.

TABLE 83: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN GRADE 5 CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	0.053	0.006	5.00	0.357
Length	115.87***	0.928	10.25**	0.533
Frequency by length interaction	0.045	0.005	4.03	0.309

From Table 83, the DRC model provided a good fit of the word RT data in its ability to capture both present length effects and absent frequency effects. From the beta values reported, the DRC model underfitted length and overfitted frequency effects and their interaction.

Table 84, below, offers a summary of the findings from the comparisons between grade-based human and modelling data. From the table, when all stimuli were considered, length and lexicality effects were found in both the children's RTs and relevant DRC models. However, in contrast to the children's word reading RTs, two DRC models (2 and 3) were not sensitive to the effects of length for known words. In addition, one DRC model (4) found a frequency effect which was absent in the behavioural data and three of the grade-based DRC models (2,3 and 4) reported a significant interaction between Lexicality and length which were also absent in the behavioural data. Only model 5 accurately captured all of the effects observed in the behavioural data.

TABLE 84: SUMMARY OF GRADE-BASED CHILDREN'S RT AND DRC MODEL REGRESSION AND FACTORIAL ANALYSES

Grade ->	2		3		4		5	
	Human	DRC	Human	DRC	Human	DRC	Human	DRC
R^2 (All)	62.3	87.4	59.3	77.4	81.6	70.5	81	80.5
R^2 (Words)	42.7	7	43.5	4.2	24.7	46.3	59	26.4
Lexicality	✓	✓	✓	✓	✓	✓	✓	✓
Length (All)	✓	✓	✓	✓	✓	✓	✓	✓
Length (Words)	✓	✕	✓	✕	✓	✓	✓	✓
Frequency	✕	✕	✕	✕	✕	✓	✕	✕
Lex x Len	✕	✓	✕	✓	✕	✓	✓	✓
Freq x Len	✕	✕	✕	✕	✕	✕	✕	✕

6.6.3 Simulation 3: Performance of average DD subtype models (Models 6 and 7)

In the first consideration of DRC models of DD in Turkish, two new models (Model 6: Phonological Dyslexia and Model 7: Surface Dyslexia) were created by optimising parameters according to the average child DRC model (Model 1). Following this, five noise parameters were added to each model to simulate deficits at each level of processing in the DRC. The outcome of Models 6 and 7 in terms of noise (deficits) is shown below in Figure 19.

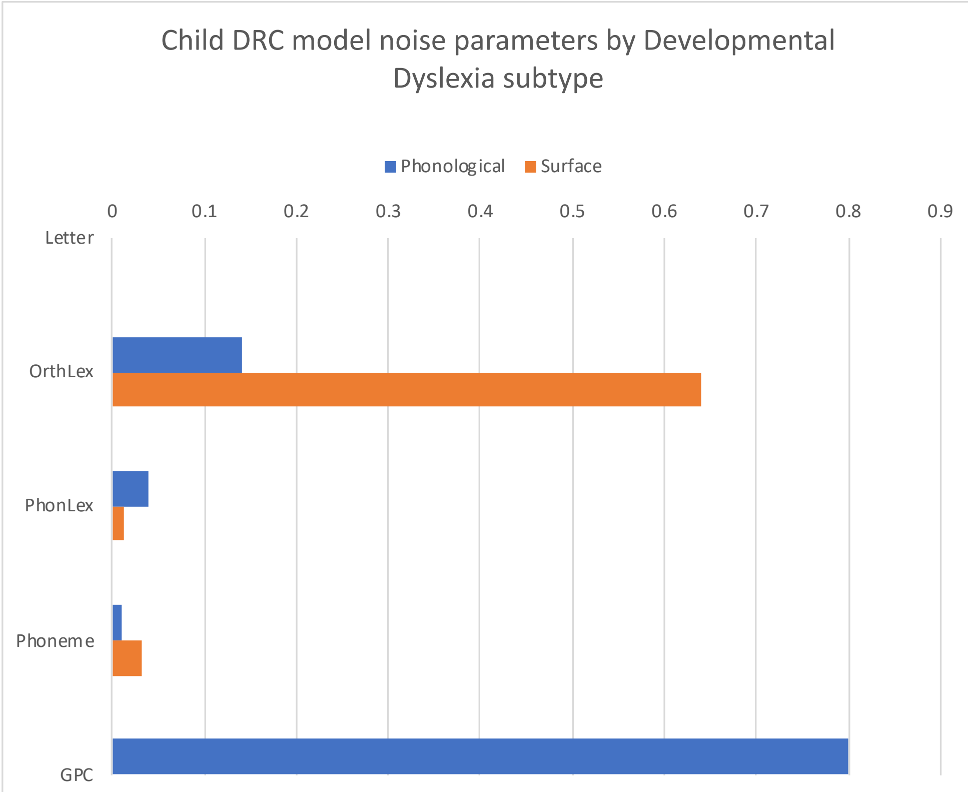

FIGURE 21: AVERAGE CHILD DRC NOISE PARAMETERS BY DD SUBTYPE

From Figure 19, phonological dyslexia was associated with a large deficit at the GPC level and smaller deficits in access to the orthographic and phonological lexicons as well as the phoneme level. Conversely, surface dyslexia was associated with a large deficit at the orthographic lexicon level and smaller deficits in access to the phonological lexicon as well as the phoneme level. Each DD subtype model was then considered independently.

Model 6: Average Phonological Developmental Dyslexia subtype Turkish DRC model

The correlation between the average Phonological DD child RTs and DRC model cycles was significant, (r (80) = 0.651; p < 0.001). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 46.6% of the variance, R^2=.466, F (10, 69) = 7.89, p<.0001 with the DRC model predicting 30.6% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 41.6% of the variance in children's RT, F (1, 78) = 57.25, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 85) and Length and Frequency (Table 86) as predictors were also carried out.

TABLE 85: LINEAR REGRESSION ANALYSIS OF PHONOLOGICAL DD CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	17.58***	71.08***
Lexicality	0.550***	0.737***
Length	0.107	0.325***
R^2	12.60%	64.00%

From Table 85, the Phonological DD DRC model provided a poor fit of the RT data in its ability to capture non-significant length effects, perhaps reflecting an overactivation of the sublexical route. However, the DRC model reported a significant length effect when none was present in the RT data. From the beta values reported, the DRC model overfitted both lexicality and length effects. The DRC model also accounted for 51.40% more variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 86: LINEAR REGRESSION ANALYSIS OF PHONOLOGICAL DD CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	1.58	17.59***
Length	0.232	0.050
Frequency	-0.158	0.696***
R^2	2.9%	46.00%

From Table 86, the DRC model provided a poor fit of the word RT data in its ability to capture frequency effects. From the beta values reported, the DRC model underfitted length effects and grossly overfitted frequency effects. The phonological DD DRC model also accounted for 43.10% more variance than the RT data suggesting an overall overfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 87 and 88 below.

TABLE 87: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN PHONOLOGICAL DD CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	48.18***	0.717	140.49***	0.881
Length	1.10	0.055	23.27***	0.550
Lexicality by length interaction	0.08	0.004	5.20**	0.3215

From Table 87, the DRC model provided a poor fit of the RT data in its ability to capture both length effects and the interaction between lexicality and length. From the beta values reported, the DRC model again overfitted lexicality and length effects as well as their interaction.

TABLE 88: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN PHONOLOGICAL DD CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	.888	0.090	.261	0.028
Length	1.34	0.130	35.70***	0.799
Frequency by length interaction	0.299	0.032	5.67*	0.386

From Table 88, the DRC model provided a poor fit of the word RT data in its ability to capture length effects. From the beta values reported, the DRC model underfitted frequency effects while overfitting length effects and their interaction.

Model 7: Average Surface Developmental Dyslexia subtype Turkish DRC model

The correlation between the average child RTs and DRC model cycles was significant, (r (80) = 0.748; p < 0.001). A hierarchical regression analysis was carried out in which acoustic articulation variables and the DRC model cycles were entered in a stepped fashion. The results of the regression analysis indicated that the selected predictors significantly explained 60.3% of the variance, R^2=.603, F (9, 70) = 14.33, p<.0001 with the DRC model predicting 31% of the unique variance in the model. When the DRC model was entered alone as a predictor into a linear regression model, DRC accounted for 55.4% of the variance in children's RT, F (1, 78) = 99.22, p<.0001. Comparison of linear regressions between the children's RT and DRC model using Length and lexicality (Table 89) and Length and Frequency (Table 90) as predictors were also carried out.

TABLE 89: LINEAR REGRESSION ANALYSIS OF SURFACE DD CHILDREN'S AND DRC MODEL'S RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	51.11***	524.85***
Lexicality	0.542***	0.822***
Length	0.526***	0.506***
R^2	55.90%	93.00%

From Table 89, the DRC model provided an excellent fit of the RT data in its ability to capture both length and lexicality effects. However, from the beta values reported, the DRC model overfitted lexicality effects and slightly underfitted length effects. The DRC model also accounted for 37.10% more variance than the RT data suggesting an overall overfitting of the DRC model.

TABLE 90: LINEAR REGRESSION ANALYSIS OF SURFACE DD CHILDREN'S AND DRC MODEL'S WORD RT

	Human (Std. Beta)	DRC (Std. Beta)
Model F	18.26***	62.39***
Length	0.702***	0.878***
Frequency	0.062	-0.006
R^2	47.00%	75.90%

From Table 90, the DRC model provided a good fit of the word RT data in its ability to capture both present length and absent frequency effects. From the beta values reported, the DRC model overfitted length effects and underfitted frequency effects. The DRC model also accounted for 28.90% more variance than the RT data suggesting an overall overfitting of the DRC model.

Turning to the factorial approach, a series of repeated measures ANOVAs were run to explore the effects of a number of psycholinguistic variables on human and DRC models data. The variables of interest and their interactions were length and lexicality for all stimuli and length and frequency for word stimuli only. The resulting ANOVAs are presented in Tables 91 and 92 below.

TABLE 91: FACTORIAL ANALYSIS OF LENGTH AND LEXICALITY EFFECTS IN SURFACE DD CHILDREN'S RT AND DRC MODEL

	Human		DRC	
Lexicality	68.94***	0.784	1330.20***	0.986
Length	27.96***	0.595	488.49***	0.963
Lexicality by length interaction	3.96	0.172	23.69***	0.555

From Table 91, the DRC model provided a good fit of the RT data in its ability to capture both length and lexicality effects. However, the surface DD DRC model reported a significant interaction where none was present in the RT data. From the beta values reported, the DRC model overfitted both lexicality and length effects.

TABLE 92: FACTORIAL ANALYSIS OF LENGTH AND FREQUENCY EFFECTS IN SURFACE DD CHILDREN'S WORD RT AND DRC MODEL

	Human		DRC	
Frequency	0.188	0.020	0.006	0.001
Length	39.76***	0.815	154.07***	0.945
Frequency by length interaction	0.320	0.034	0.947	0.095

From Table 92, the DRC model provided a good fit of the word RT data in its ability to capture length effects. From the beta values reported, the DRC model overfitted length effects while underestimating the frequency effect.

Table 93, below, offers a summary of the findings from the comparisons between DD subtype human and modelling data. From the table, when all stimuli were considered, lexicality effects but not length effects were found in the phonological DD children's RTs. In contrast, the relevant DRC model (model 6) found both length and lexicality effects as well as a significant interaction between them. When considering phonological DD children's word reading RTs, no main or interaction effects were significant. In contrast, model 6 found both length effects and a significant interaction between length and frequency. When the surface DD profiles were considered, surface DD children's RTs showed length and lexicality effects when all words were considered and only significant length effects when only words when studied. The relevant DRC model (model 7) was able to successfully simulate this pattern of results but also included a significant length by lexicality interaction which was not present in the children's RT data.

TABLE 93: SUMMARY OF AVERAGED DD SUBGROUP CHILDREN'S RT AND DRC MODEL REGRESSION AND FACTORIAL ANALYSES

	Phonological		Surface	
	Human	DRC	Human	DRC
All(R^2)	12.6	64	55.9	93
Length	×	✓	✓	✓
Lexicality	✓	✓	✓	✓
Length by Lex	×	✓	×	✓
Words (R^2)	2.9	46	47	75.9
Length	×	✓	✓	✓
Frequency	×	×	×	×
Length by Freq	×	✓	×	×

In summary, both the phonological and surface DD DRC models overfitted the models to the RT data. While both models were successful in accounting for the lexicality effect, only the surface DD DRC model accounted for the length effect found in the RT data. Furthermore, both models reported significant length by lexicality interaction were none were found in the RT data. Of the two models, the surface DD DRC model was the more successful in terms of the number of effects (main and interaction) accounted for (Surface: 5/6 vs Phonological: 2/6).

6.6.4 Simulation 4: Performance of Individual DD models (Models 8 – 23)

In a final consideration of DRC models of DD in Turkish, 15 new models were created by optimising parameters according to the appropriate grade-based child DRC models (models 2-5). While model accuracy was not of specific consequence to the current work, model parameters were selected only if the resulting models could name as accurately or better than the corresponding individual data but not worse. Using this approach, the correlation between overall accuracy scores of children and accuracy of the DRC models was significant $r(15) = .881$, $p < .001$. Furthermore, the overall DRC accuracy scores predicted 75.9% of the variance in the observed overall accuracy scores $F(1,13) = 45.11$, $p < .001$. Following this, five noise parameters were added to each model to simulate deficits at each level of processing in the DRC. The outcome of models 8 - 23 in terms of noise (deficits) is shown below in Table 94 for the regression analyses and Table 95 for the factorial ANOVAs.

TABLE 94: SUMMARY OF LINEAR REGRESSION ANALYSIS OF INDIVIDUAL DD DRC MODELS

ID ->	PD1		PD2		PD3		PD4		PD5		PD6		PD7		SD1		SD2		SD3		SD4		SD5		SD6		SD7		MD1	
ID->	901		904		902		601		3		277		903		159		162		552		900		82		906		604		120	
	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC
r	0.447		0.371		0.555		0.292		0.783		0.578		0.398		0.564		0.639		0.511		0.448		0.582		0.508		0.019		0.656	
R2	18		12.1		29.6		6.9		60.7		32.3		14.5		30.7		40		25.1		19		32.9		24.8		-1.4		42.2	
Len					✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
Lex	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Len					✓			✓	✓	✓		✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓
Freq									✓			✓				✓		✓				✓								✓

From Table 94, linear regression analysis between the RT data of the 15 children with DD on both words and nonwords as well as words only with the predictions of the corresponding DRC models was carried out. On average, the models accounted for 25.83% (SD: 15.45; Range: -1.40 – 60.70) of the variance. Closer inspection of the data revealed that two models (PD4 and SD7) performed particularly poorly. A reanalysis with these two models removed accounted for 29.38% (SD: 13.56; Range: 12.10 –

60.70) of the variance. Taken together the individual DD models correctly accounted for 73% of the predictor variables found in the RT data, while 23% of the predictor variables in the DRC models reported a false effect, i.e. finding an effect when none was present, and 3% of DRC predictor variables failed to account for a significant predictor found in the RT data. In addition, only three of the individual DD models (PD1, SD3 and SD5) accounted for all of the predictor variable findings in their corresponding RT data whereas nine individual models (PD2, PD3, PD5, PD7, SD1, SD2, SD4, SD6 and MD1) had one discrepancy between the predictor variables findings of the corresponding RT data. Further, two individual DD models (PD6 and SD7) had two discrepancies between their predictor variables and the corresponding RT data with one model (PD4) having three discrepancies between their predictor variables and the corresponding RT data. Encouragingly, all of the individual models accounted for at least one predictor variable that was present/ absent in the RT data.

TABLE 95: SUMMARY OF FACTORIAL ANALYSIS OF INDIVIDUAL DD DRC MODELS

ID (Chapter 5)->	PD1		PD2		PD3		PD4		PD5		PD6		PD7	
ID->	901^		904		902^2		601		3		277		903	
	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC
Length	-	0	0.124	0.041	0.85	.310**	0.315	.708***	.830***	.776***	.649**	.930***	0.115	.341**
Lexicality	-	.659***	0.613	0.632***	.906*	.771***	0.069	.816***	.938***	.976***	.717**	.839***	0.378	.867***
Len * Lex	-	0	0.001	0.003	0.846	.279*	0.016	.658***	.655**	.574***	0.065	0.003	0.001	0.181
Length	0.986	0.131	0.02	0.025	0.385	0.155	0.086	.736***	0.329	.559**	0.153	.922***	0.075	0.25
Frequency	0.122	0.108	0.016	0.112	0.004	0.058	0.305	0.162	.445*	0.006	0	.364*	0.026	0.111
Len * Freq	0.305	0.131	0.053	0.05	0.041	0.153	0.067	0.068	0.009	0.124	0.02	0.004	0.043	0.116

ID (Chapter 5)->	SD1		SD2		SD3		SD4		SD5		SD6		SD7		MD1	
ID->	159		162		552		900		82		906		604		120	
	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC	RT	DRC
Length	0.153	.665***	.687***	.911***	.446**	.903***	.774***	.808***	0.229	0.893***	0.062	.867***	0.205	.483***	.812**	.879***
Lexicality	.605***	.946***	.600***	.940***	.634***	.871***	0.255	.774***	.513***	0.934***	.407***	.938***	0.033	.834***	.459***	.910***
Len * Lex	.519***	.642***	0.06	.793***	0.013	0.18	0.072	.258**	0	.599***	.217^3	.834***	0.196	0.063	0.071	.752***
Length	.682*	.886***	.836***	.821***	.737***	.845***	.731***	.725***	0.287	.851***	0.007	0.066	0.067	.428***	.676***	.728***
Frequency	0.199	.403*	0.008	.637**	0	0.283	0.003	.480**	0.161	0.056	0.036	0.172	0.288	0.134	0.001	.441*
Len * Freq	0.107	0.04	0.025	0	0.02	0	0.01	0.083	0.113	0.285	0.002	0.002	0.003	0.063	0.056	0.158

From Table 95, the individual DD models correctly accounted for 67% of the effects (main and interaction) found in the RT data, while 32% of the significant effects in the DRC models reported a false effect, i.e. finding an effect when none was present, and 1% of DRC effects failed to account for a significant effect found in the RT data. In addition, only two of the individual DD models (PD1 and SD3) accounted for all of the effects in their corresponding RT data whereas two individual DRC models (PD2 and SD6) had one discrepancy between the effect findings and the corresponding RT data. Seven individual models (PD3, PD5, PD6, PD7, SD1, SD2 and MD1) had two discrepancies between the model's effects findings and the corresponding RT data.

Further, three individual DD models (SD4, SD5 and SD7) had three discrepancies between their predictor variables and the corresponding RT data with one model (PD4) having four discrepancies between their effect and the corresponding RT data. Encouragingly, all of the individual models accounted for at least two effects that were present/ absent in the RT data.

TABLE 96: NOISE PARAMETERS FOR INDIVIDUAL DD MODELS

ID->	PD1	PD2	PD3	PD4	PD5	PD6	PD7	SD1	SD2	SD3	SD4	SD5	SD6	SD7	MD1
Letter	0.03	0.02	0.04	0.02	0.04		0.04	0.05		0.07	0.02	0.02	0.02		0.01
OrthLex		0.40			0.20						0.10	0.70	0.06	0.20	0.20
PhonLex		0.02			0.01	0.04		0.03	0.02	0.02	0.02		0.02		0.02
Phoneme								0.02	0.02	0.02					0.04
GPC	0.60	0.60	0.20	0.20	0.19	0.80	0.05	0.34	0.20	0.32	0.60	0.40	0.60	0.80	0.30

From Table 96, the average Phonological Noise parameter was 0.50 (0.35) compared to the average Surface Noise parameter of 0.67 (0.21). In terms of the number of noise deficits, the average count for the phonological models was 2.57 (0.98) compared to the average count for the surface models 3.43 (0.79). A series of paired t-tests found no significant difference between phonological and surface subtypes in terms of noise parameters across all levels of representation in the DRC model. In terms of the number of noise deficits, letter level noise was found in 6 phonological models compared to 5 surface models, orthographic lexicon level noise was found in 2 phonological models compared to 4 surface models, phonological lexicon level noise was found in 3 phonological models compared to 5 surface models, phoneme level noise was found in no phonological models compared to 3 surface models, and GPC level noise was found in all DD models.

6.7 Discussion

This chapter assessed several versions of the DRC model by piloting a quantitative evaluation of the models' responses to those of Turkish-speaking children with and without DD. This computational modelling work aimed to evaluate the utility of the DRC model framework for Turkish-speaking children's RT data as well as to explore further the DD profiles and current subtyping approaches to Turkish. To achieve this, all 23 created models were assessed using both a factorial and regression approach. To this end, the approach taken during this chapter demonstrated that the DRC model could be successfully utilised to account for both group-based and individual DD profiles, with some limitations (discussed further below). The following section will further consider the simulations, starting with the group-based considerations before presenting an overview of the individual DD profiles.

6.7.1 Simulation of the average child's reading performance

The average Turkish child DRC model was able to simulate the human RT data of the word and nonword stimuli with 100% accuracy. However, the model was overly sensitive to lexicality effects in that nonword reading in the DRC model was roughly twice the number of cycles in the word reading condition. In contrast, the corresponding difference in the average child's RT data was equivalent to a 24% increase between words and nonwords. In addition, the DRC child model's sensitivity to length effects across stimuli and absent word frequency effect is also in line with the RT data. Conversely, when only considering the word stimuli, the DRC child model failed to report a significant length effect that was present in the RT data. As stated previously, the failure of the DRC model to account for length effects in words was predictable due to parallel processing in the lexical route. Similar findings have been reported in the Greek version of the DRC model (Kapnoula et al., 2017). The marked oversensitivity to lexicality effects and under sensitivity to length effects in known words may either reflect a divergence between the Turkish readers and the DRC architecture or be indicative that further parameter optimisation is needed. Considering the latter possibility, Kapnoula and colleagues (2017) carried out further optimisation of parameters in the face of a lack of reported length effects in known words within the DRC model. The optimisation procedure found a significant length effect in known words at the expense of accuracy. According to Kapnoula et al., (2017), the observed trade-off between the strength of the non-lexical route and naming accuracy may be informative regarding how humans read words or may alternatively reflect an underlying limitation of the DRC model. The authors conclude that these observed divergences between the RT and DRC data suggest that the DRC model (as a model built on English theoretical frameworks of reading) may overly rely on lexical processing though they argue that further simulations and model comparisons are needed to address this outstanding question. In light of this, further consideration of future modelling endeavours is considered below.

Separately, the use of the GPC learning algorithm for the development of the average Turkish child DRC model appears to have particular usefulness for Turkish, specifically and perhaps other transparent orthographies generally. Using this approach, simulations showed that the GPC Learning Model could successfully acquire GPCs to the degree that reflects ceiling level performance in terms of model accuracy. The use of only 50 GPC rules compared to the 104; Dutch, 226; English, 340; French, 130; German and 59; Italian versions of the DRC (Schmalz, Marinus, Coltheart, & Castles, 2015) further quantitatively communicate the degree of transparency of the Turkish orthography. This further confirms the usefulness of Turkish as an orthographic medium for further computational investigation of visual word recognition.

In sum, while the Turkish child DRC model provides a better fit to human RT data than previously reported DRC models in other orthographies such as English (Perry et al., 2007), French (Ziegler et al., 2008) and Russian (Ulicheva, 2015) a number of divergences between the RT and DRC data were observed. Coupled with data from adult Turkish (I. Raman, 1999; 2003) the outcome of simulation 1 seems to lend provisional support to dual-route accounts of reading without discarding the possibility that alternative theoretical frameworks of reading could accommodate the results reported in this chapter.

6.7.2 SIMULATION OF AVERAGE GRADE-BASED READING PERFORMANCE

A novel approach adopted in the current chapter was an attempt to develop average Grade-based models of reading using the DRC architecture. By adopting the self-teaching DRC model approach (Pritchard, 2012), four grade-based DRC models were created and evaluated against their corresponding RT data. While connectionist models have made a long and sustained contribution to our understanding of reading development (e.g. Monaghan & Ellis, 2010), the recent incorporation of the self-teaching hypothesis within dual-route models (Perry, Zorzi, & Ziegler, 2019; Pritchard, Coltheart, Marinus, & Castles, 2018; Ziegler, Perry, & Zorzi, 2014) represents an interesting advance for computational models of visual word recognition. The increasing availability of reading acquisition models provide further opportunity for exploring the developmental changes that accompany learning to read. While a comprehensive computational investigation of reading development was beyond the scope of the current study, there is much opportunity to build on the preliminary work reported in this chapter.

Overall, the development and findings of the four Turkish grade-based DRC models was encouraging but again presented with a number of divergences from the RT data. Firstly, all four grade-based Turkish child DRC models were able to simulate the human RT data of the word and nonword stimuli with 100% accuracy. On average the four models accounted for 62% of the variance in their corresponding RT datasets with the grade-based DRC models of older children (73%; grade 4 and 5) performing better than the grade-based DRC models of younger children (51%; grade 2 and 3). Additionally, when considering all word/ nonword stimuli, grade 2 and 3 models overfit the captured variance relative to the RT data, whereas the grade 4 model underfit the captured variance relative to the RT data. The Grade 5 DRC model provided a near-perfect match to the behavioural data concerning the captured variance when all stimuli were considered. The length by lexicality interaction is argued to be the hallmark feature of dual-route models (Weekes, 1997) as this interaction is interpreted as signifying the parallel lexical processing of words and serial processing of pseudowords. To this end, using the factorial approach, only the Grade 5 RT data and DRC model reported a significant interaction between length and lexicality. Therefore, based on averaged models, only the grade 5 RT and DRC model manifest with the full availability of both routes. This is clearly at odds with the mixed model analysis carried out in Chapter 4, and likely reflects a general limitation of the ANOVA approach. This particular limitation will be further considered following the discussion of the remaining models.

When considering only the word stimuli, grade 2 and 3 based DRC models failed to capture length effects that were present in the RT data. Interestingly, the grade 4 based model overfitted the word-only RT data by reporting a significant frequency effect when one was not present in the corresponding behavioural data. Again, only the grade 5 model provided an account of all of the effects found in the RT data though it captured under half of the variance in the word-only RT data. Further evaluation found that while Grade 2 and 3 models both underfit length effects and overfit lexicality effects, Grade 4 and 5 models showed the opposite pattern.

In summary, while all four models were able to capture present length and lexicality effects across the word/pseudoword stimuli, the failure of grade 2 and 3 DRC models to accommodate the word length effect found in the corresponding word RT data presents a challenge for modelling reading development using the current computational approach.

6.7.3 Simulation of DD subtypes

The following section reports an attempt to develop average DD subtype models in Turkish using the averaged RT of the two groups. Impaired reading of each of the DD subtypes was simulated by first adjusting model parameter values and then by adding noise to each level of representation in the DRC. When the DD cohort was again divided into the phonological and surface subtypes according to the regression procedure adopted in Chapter 5, the results suggest that phonological dyslexics as a group were primarily affected at the GPC level. Concerning the surface group, results suggest that they were primarily affected at the level of access to the orthographic lexicon. This apparent divergence lends support to dual-route accounts of DD, in that phonological dyslexia develops from distinct impairment to the sublexical route, whereas surface dyslexia develops from distinct impairment to the lexical route

(Peterson, Pennington & Olson, 2013). However, several additional observations raise a number of important issues. For instance, the modelling approach reported above also found that the surface DD model (model 7) had a larger phoneme deficit than the phonological DD model (model 6). Similar findings have been reported in behavioural data in transparent orthographies such as Spanish (Jiménez, Rodríguez, & Ramírez, 2009) as well as the opaque orthography of French (Sprenger-Charolles et al., 2000; Ziegler et al., 2008).

Additionally, the phonological DD model presented with a marked secondary deficit in the orthographic lexicon. While seemingly an odd finding, the role of the lexical route in the aiding sublexical processing can offer further insight into this finding (Coltheart & Leahy, 1996). That is, feedback from lexical access to the sublexical route (Coltheart et al., 1993; 2001) through analogy (Glushko, 1979) can explain the access to orthographic lexicon deficit observed in the average phonological DD model. Furthermore, the presence of deficits across multiple levels of representation within the DD subgroup DRC models lends further support to the multiple-deficit hypothesis of developmental dyslexia as well as indicating that the manifestation of DD in Turkish-speaking children is heterogeneous as neither the phonological nor surface DRC model was characterised by a single deficit.

Considering the individual performance of each model, both Model 6 (Phonological DRC) and Model 7 (surface DRC) accounted for a good amount of variance in the RT data but accounted for more variance than the RT data did when the effects of length and lexicality were considered. Model 6 overestimated the effect of both length and lexicality and also reported a present length effect when one was not present in the data. Model 7 successfully captured the effect of length and lexicality though overestimated the effect of length. Considering the words only analysis, Model 6 again reported a present length effect when one was not present in the data. Model 7 again successfully captured the present effect of length and the absent effect of frequency though overestimated the effect of length.

In summary, both the phonological and surface DRC models provide a reasonable fit to the RT data with the surface DRC model performing at a superior level in capturing the effects observed in the RT data. Although the averaged DD DRC models suggest that there are multiple deficits in both subtypes of DD when considered as a group, it is feasible that each individual had only a single deficit and that the observed pattern of deficits only appeared as a consequence of the averaging procedure. To further investigate this, individual models were created for the 15 DD cases first reported in Chapter 5.

6.7.4 Simulation of individual DD profiles

The principal aim of the current modelling work was to simulate normal and impaired reading with the newly created Turkish child version of the DRC model. The originality of the approach taken was to develop individual models for each of the children with DD reported in Chapter 5 by first optimising parameters based on grade-based models and then by systematically adding noise to each representational level to increase the captured variance. This work was primarily influenced by a previous DRC modelling endeavour carried out by Ziegler and colleagues (2008). However, while the previous attempt at modelling DD in the DRC model used a top-down approach in that noise parameters were altered according to the observed deficits in corresponding ancillary tasks, the approach adopted in this chapter adopted a bottom-up approach in that no a priori information beyond RT data of children with DD was used in optimising the parameters of the model. Model selection was, therefore, solely based on the best-fitting parameters for the model that were found during the duration of the computational work. The rationale for this is threefold, first, while the computational work carried out by Ziegler and colleagues (2008) builds upon a fully developed French DRC model of visual word recognition (Ziegler, Perry & Coltheart, 2003), the current chapter presents ongoing work towards the development of a DRC model in Turkish. Second, Ziegler et al. (2008) used word reading accuracy as their dependent variable were as the current approach used RT and cycles. Third, while a number of variables measured in this doctoral project could potentially be mapped onto the DRC model as ancillary tasks, this was not the explicit purpose of the current research.

According to the observed noise parameters across participants, the most substantial deficits were obtained for phonological processes. Also, further examination of the phonological noise parameters revealed that all 15 participants had deficits at the GPC (sublexical) level, 9/ 15 (3 phonological, 5 surface, 1 mixed) had deficits at the phonological lexicon level, and 4/15 (4 Surface) had deficits at the phoneme level. This finding lends support to both the phonological deficit hypothesis of DD (Stanovich & Siegel 1994; Snowling, 2000) and the double-deficit hypothesis (Wolf, 1999) in that the central deficits concerning DD are associated with phonological processes. Furthermore, the majority (12/15) of individual DD models also exhibited letter processing deficits. This finding is mostly coherent with a number of studies that highlight the importance of deficits in the parallel processing of letters (Bosse et al., 2007; Valdois et al., 2004; Zoubrinetzky, Bielle, Valdois, 2014). Finally, there was also evidence for the presence of deficits in the orthographic lexicon level in 7/15 DD models (3 phonological, 3 surface, 1 mixed). In sum, evaluation of the individual models suggests that as the majority of models exhibited multiple deficits across the DRC's representational levels, the strongest theoretical argument regarding the manifestation of DD, in models of Turkish-speaking children, is provided by the multiple deficit model of DD. The considerable heterogeneity in both the number and magnitude of deficits across subjects further lends support to the concept of multiple probabilistic risk factors (Pennington et al., 2012).

Additionally, when considering DD subtypes, the individual model approach adopted in this chapter stipulates that there were no meaningful differences in the underlying noise parameters and therefore argue that the classification of children with DD into qualitatively distinct subtypes presents an inadequate account of DD (Griffiths & Snowling, 2002). Instead, growing evidence suggests that phonological and surface subtypes may best be represented as two ends of a continuum (Castles et al., 1999; Griffiths and Snowling, 2002). A parallel finding was also reported by Ziegler et al. (2008).

6.7.5 Limitations and Future research

In addition to some of the limitations highlighted in Chapter 5 concerning the behavioural data such as the small sample size, there are also a number of limitations in the current computational work. Firstly, the use of averaged models and RT datasets reintroduces an issue previously addressed in Chapters 4 and 5. That is the reliance on averaging across participants or items. Coupled with the small sample size, the findings of the current study can be accepted as exploratory at best.

Perhaps the largest challenge in the current computational work was the suboptimal optimisation procedure used as a consequence of limited computational power and time. While several models captured a good amount of variance of their respective RT data, several also failed. In future, to achieve superior fitting models to individual participants will require data of greater precision at an individual level (Adelman & Brown, 2008). Similarly, future studies could address this limitation by adopting an alternative optimisation procedure such as the one reported in Adelman and Brown (2008). In an attempt to address the so-called optimal parameters problem, Adelman and Brown (2008) applied the Nelder-Mead simplex procedure to the DRC. As such, the simplex procedure is an example of a standard non-gradient local search method which inspects the corners of a region of interest, i.e. the parameter space of a model and subsequently expands into parameter spaces that represent improved model fit while also moving away from regions that fail to improve the model. In doing so, the DRC model can be optimised to a state where any observed effects or predictor variables are a consequence of a given theoretical framework rather than an artificial obstacle to falsification brought about by suboptimal approaches to model fitting.

Developmental models of reading represent a fascinating enterprise that has a rich history in connectionist approaches (see Chang, Monaghan & Welbourne, 2019 for a recent implementation) but has only recently received increasing interest in dual-route accounts of reading (Pritchard, 2012) There are also further prospects to carry out further research using both the GPC learning and self-teaching models (Pritchard, 2012). While the approach taken to self-teaching was akin to one-shot learning, a more psychologically plausible approach would be to incorporate a process of incremental training that starts with a small number of words and eventually builds a full representative vocabulary over time. In a similar line of enquiry, the recently implemented CDP++ developmental model of learning to read (Ziegler, Perry & Zorzi, 2014) has made some important progress in this direction. Using this approach, the authors achieved an accuracy rating of 80% for the 32,000 words that the model was exposed to after training. While the default parameters of the self-teaching model were used for model development in the current work, future studies need to address the distinct lack of vocabulary growth data in Turkish-speaking primary school children. Following this, a more fine-grained exploration of vocabulary change, orthographic learning and self-teaching in Turkish can take place. In line with this, future computational modelling work in Turkish needs to take a closer look at features of phonology as there is much debate surrounding this topic (see Chapter 2 of this thesis; Koşaner, Birant, & Aktaş, 2013). Finally, while there is a widely accepted view in the domain regarding the relevant dependent variables of interest being a function of orthographic transparency, i.e. transparent orthographies index RT whereas opaque orthographies index accuracy, there is also a need to take potential speed-accuracy trade-offs into account. Future behavioural and computational studies can adopt the response signal paradigm (Reed, 1973) in order to address this.

The recent move towards modelling individual differences (Ziegler, Perry & Zorzi, 2019) in order to increase our understanding of the heterogeneity of DD represents a critical line of enquiry with computational models. Using this approach, the authors were able to develop a model that could successfully simulate the individual learning trajectories of 622 children (388 with DD). The authors argue that only by adopting personalised computational models that allow for multiple deficits can the heterogeneity and individual differences in DD profiles be captured. Such models may also have significant implications on the early identification of DD as well as being informative regarding the outcomes of interventions.

In summary, this chapter provides an overview of the first computational study to evaluate a series of DRC models of visual word recognition and reading aloud with Turkish stimuli by directly measuring modelling alongside behavioural data. The computational work undertaken within this chapter, while exploratory, produced good fitting models of the average Turkish-speaking children's RT data as well as adequate models of grade-based average RT data and subtypes of DD. Overall, the DRC models produced were generally useful in simulating several aspects of the behavioural data; however, a number of limitations were also identified. In particular, several of the DRC models displayed sensitivity to length, lexicality and frequency effects (which were absent in the behavioural data). Certain parameter adjustments would expectedly lead to better-quality models. Despite the limitations of the current modelling work, the simulations carried out here provide a rich foundation for further work on computational models of reading, reading development and disorder among Turkish-speaking populations.

CHAPTER 7: GENERAL DISCUSSION

This chapter begins with a summary of the research carried out within this doctoral project intending to highlight both the key findings and the unique contribution made by the current research. Following this, a consideration of both the limitations and future avenues for research that builds on this work will be offered.

7.1 SUMMARY OF FINDINGS, LIMITATIONS AND FUTURE STUDIES

The primary aim of this doctoral research project was to apply contemporary cognitive, computational and psycholinguistic theories and methods to the exploration of reading, reading development and disorder using the Turkish orthography as the medium of choice. This choice was motivated by the underlying orthographic transparency of the Turkish writing system in a preliminary attempt to redress the largely Anglo-centric nature of research that has been carried out in reading research to date (Share, 2008).

7.1.1 HOW TRANSPARENT IS TRANSPARENT? A QUANTIFICATION OF THE TURKISH ORTHOGRAPHY

To this end, the work carried out in Chapter 2 bid to understand the orthographic transparency of Turkish from a quantitative perspective. Chapter 2 began with a review of the previous research on approaches to understanding the differences between alphabetic writing systems with a focus on recent conceptualisations and quantification attempts regarding orthographic transparency. Additionally, the properties of the Turkish orthographic and phonological system were described in detail. Surprisingly, while there is a degree of agreement between researchers in terms of qualitative descriptions of orthographic transparency, only a hand-full of studies have considered quantitative indices (e.g. Borgwaldt, Helliwig, & de Groot, 2004). It is the author’s view that the ongoing conversation surrounding universal and language-specific aspects of reading will only progress when a) more diverse orthographies and writing systems are included into current debate and b) more quantitative approaches are developed to try to capture the diversity of orthographies.

From a methodological perspective, entropy was adopted over calculating the percentage of dominant mappings as entropy measures factor in the relative proportions of non-dominant mappings. As such, variability in terms of entropy values (Borgwaldt, Hellwig & de Groot, 2004; Martensen, Maris, & Dijkstra, 2000; Protopapas & Vlahou, 2009) were generated for each word in the created corpus.

To the best of the author's knowledge, Chapter 2 provides the first quantitative indices of orthographic transparency for Turkish. Regarding this, the Turkish orthography was characterised as both highly predictable (h = 0.045) and simple (100% one-to-one GPC mapping) at the grapheme level. However, deviations from the alphabetic principle exist within the feedforward direction, i.e., the presence of irregular words.
The approach used in Chapter 2 also addresses previous limitations of using unrepresentative samples in the quantification of an orthography, such as using monosyllabic words or word-initial letters entropy measures (Protopapas & Vlahou, 2009; Schmalz, Marinus, Coltheart, & Castles, 2015).

Chapter 2 highlights an important consideration for future psycholinguistic research concerning orthographic transparency. Research carried out to date frequently identifies a number of Indo-European orthographies such as Dutch, German, Greek, Italian and Spanish as examples of transparent orthographies. When compared to Turkish, it is clear that any claims of overall transparency are relative. When considered in absolute terms, these orthographies, with the exception of Italian, would be considered to be of medium transparency. It is consequently the author's view that the psycholinguistic investigation of Turkish will benefit ongoing research in the domain by highlighting these important differences. With specific interest to the author, understanding where unpredictability and complexity exist, can aid in a greater understanding of literacy acquisition and disorder. This thesis then represents a theoretical starting point to begin to deconstruct Anglo-centric dominated reading research. If English is considered to be an outlier orthography, then Turkish represents the opposite side of the continuum as it closely adheres to the alphabetic principle.

One primary avenue for future research is to carry out a similar language entropy model development using the newly created SUBTLEX-TR and comparing this with the entropy models created in this thesis. Additionally, there is scope to utilise entropy as an item-level measure (feedforward/ back consistency) in future psycholinguistic investigations in Turkish. Item-level analysis of consistency/ regularity and predictability would be of particular interest in future studies of Turkish psycholinguistic research as they are currently absent from the literature. While orthographies are characterised on a continuum of transparency as a whole, a finer grained analysis of item-level variation within orthographies would be beneficial to our understanding of visual word recognition. Chapter 2, therefore offers a useful resource for such a study to be undertaken in Turkish.

Additionally, beyond the generation of well-controlled linguistic stimuli, the real value and utility of calculating a quantitative index of transparency lies in the ability to carry out cross-linguistic comparisons. At present, this comparison is restricted to analyses that have adopted the same approach to quantification and therefore, severally restrict the generalisability and applicability of the approach beyond the handful of orthographies that have been quantified. Furthermore, orthographic transparency, only represents one of many possible dimensions that capture the variability among the world's writing systems. As such, the results of Chapter 2 can, at this stage, only be compared to other alphabetic orthographies. However, there is also a need to develop a further framework that extends to non-alphabetic writing systems. For example, Shimron (2006) states that Hebrew depth is very different from English depth. Differences in graphic complexity (Chang, Chen, & Perfetti, 2017), for example, offer challenges for frameworks like orthographic depth that are based exclusively on alphabets. Additionally, Daniels and Share (2018) propose that in order to accommodate for the full spectrum of the world’s writing systems, at least ten dimensions of complexity need to be considered concerning reading development and disorder. These dimensions fall into one of three categories concerning oral language structure, visual shapes complexity, and the translation rules between the visual and phonological domains. Future attempts at developing quantitative indices of writing systems, therefore, need to consider these dimensions in order to truly develop an understanding of writing system variation.

7.1.2 SUBTLEX-TR: The Creation and Validation of a New Psycholinguistic Database for Turkish

Chapter 3 reviewed the currently available resources for psycholinguistic research in Turkish. Until recently, Turkish psycholinguistic researchers have relied solely on word stimuli that they have created themselves (e.g. Babayiğit & Stainthrop, 2007; I.Raman, 2011). The apparent lack of reproducibility by taking such an approach as well as the inefficiency of generating new stimuli for every new experiment reaffirmed the need to create a widely available psycholinguistic database for use in both Turkish-speaking children and adults. Highlighting a distinct lack of widely-available resources within the research area, Chapter 3 introduced the development of SUBTLEX-TR, a new Turkish-word database. The SUBTLEX-TR was validated with 72 participants completing a lexical decision task and confirmed the usefulness of the new frequency measures by comparing them with estimates derived from TS Corpus (Sezer & Sezer, 2013).

The findings of the lexical decision task were particularly interesting for theories of visual word recognition. Specifically, the presence of main effects of length, lexicality and frequency as we as a length by lexicality interaction provides evidence that both the lexical and sublexical processes are available to Turkish readers and conceivably reflects the flexibility of the reading system. This finding poses a challenge to the strong view of the Orthographic Depth Hypothesis and instead, lends support to the both the universal hypothesis and the PGST which argues that a transparent orthography operates both lexical and sublexical mappings together (Ziegler & Goswami, 2005). However, within the literature, several arguments have emerged in recent years that question the applicability of dual-route theory to transparent orthographies (Ardila & Cuetos, 2016) based on the argument that efficient reading only requires a sublexical approach given the simple mapping between orthography and phonology.

Taken together, this thesis argues that even in extremely transparent orthographies such as Turkish, evidence suggests the active use of the lexical route for reading. Interpreted in a dual-route framework, it could be postulated that the transparency of writing systems may facilitate a more congruent relationship between the lexical and sublexical routes given that the outcome of phonology is rarely ambiguous. That is, while some words a read predominantly via either the lexical or sublexical route, the majority of words can be read using either strategy.

Overall, there was a large advantage of SUBTLEX-TR over TS Corpus when words for which estimates given by the two corpora differed most were used as stimuli. With the increasing transition from printed to electronic media, Chapter 3 represents an important future direction of psycholinguistic investigation in Turkish. Chapter 3 also included the creation and validation of a new normative children's database for use in Turkish, representing an essential and much-needed direction in the literature. Additionally, Chapter 3 highlights the complex relationship between word frequency measures derived from distinct corpora and human performance on psycholinguistic tasks such as lexical decision. Ultimately, the SUBTLEX-TR corpus represents the first widely-available subtitle derived word database for Turkish. The database provides frequency and contextual diversity measures based on Turkish language subtitles. It is anticipated that the SUBTLEX-TR corpus will be a valuable resource in future psycholinguistic investigation in Turkish. It is also planned for this resource to be freely available online in the form of a web app.

Chapter 3 reported that Contextual Diversity (CD) measures accounted for less variance than SUBTLEX-TR word frequencies and the high degree of correlation between the two measures makes interpretation of this finding difficult. This sits in stark contrast with the literature, which frequently highlights the influence of CD on lexical decision times (e.g., Adelman, Brown, & Quesada, 2006; Perea, Soares, & Comesaña, 2013). There are several possible reasons why the current study failed to replicate the results of previous SUBTLEX studies. First, CD may have different effects in different orthographies, that is, CD effects may be language-specific. Second, while word frequency was the variable of interest, CD effects were only considered as an additional variable. Future studies in this area will need to explore this curious finding further by controlling for word frequency and manipulating CD as the variable of interest. While, as mentioned above, frequency and CD are highly correlated, and therefore only carefully designed factorial studies (with enough power) or the development of megastudies in Turkish (see below) could potentially address this finding.

Additionally, when considering the children's data, there was no advantage of the SUBTLEX-TR-child word frequencies over Children’s Language Corpus word frequencies. This null finding also extends to children's CD measures. The variance accounted for by the accuracy data is similar between children and adults. However, regarding RTs, the children's naming data accounted for significantly less variance than the adult lexical decision data. The non-significant findings of the SUBTLEX-TR-child corpus in comparison to the CLC database raise several important methodological issues regarding validation studies in children. Firstly, the small number of words used for this sub-investigation were not selected for their highly divergent estimates of word frequency and as such were highly correlated). In addition, the low variance captured for RT across the corpora suggests that naming tasks, particularly regarding children, are less informative than lexical decision tasks for use in validation studies. Furthermore, these methodological issues may be compounded by the reduced sensitivity to frequency effects in transparent orthographies. With these limitations considered, the small non-significant findings of a SUBTLEX-TR-child advantage over the child literature corpus (Acar, Zeyrek, Kurfali, & Bozsahin, 2016) word frequencies warrant further exploration taking into consideration the methodological issues stated above. Future iterations of the SUBTLEX-TR are already underway and will contain measures of Parts-of-Speech, CV type, lemma frequencies and initial phoneme, and further validation will take place in the form of a lexical decision mega study. New releases of the opensubtitles.org subtitle data will provide the opportunity to update the current database. A planned mega study will directly build upon the work undertaken in this thesis to elucidate further the variables that influence visual word recognition in Turkish.

7.1.3 The Development of Reading in a Highly Transparent Orthography

Chapter 4 began with an overview of the literature regarding reading development with a focus on transparent orthographies such as Italian, Spanish, Greek, Finnish. The limited research into the development of reading skills in Turkish was highly informative with regards to the rapid development of phonological awareness skills (Öney, & Durgunoğlu, 1997; Durgunoğlu., & Öney, 1999). Beyond this, there was little evidence of comprehensive investigations into visual word recognition and reading skill development in Turkish-speaking children. This gap in the literature, along with the methodological concerns of previous studies such as the lack of measures of RAN (Durgunoğlu & Öney, 1997) and use of word lists over discrete words (Babayiğit & Stainthrop, 2007) motivated the experimental work carried out in Chapter 4. The Chapter contributes to several topics concerning the underlying cognitive and linguistic mechanisms of reading development in Turkish children.

Taken together, for the 131 typically developing children learning to read in Turkish, performance was dependent on a mixture of both lexical and sub-lexical knowledge. Furthermore, the results obtained in the current study reveal that phonological awareness, rapid automatised naming and visual attention span differentially influence reading ability. As reading accuracy reaches ceiling quickly, the focus shifts toward developing reading speed by automating a superior method of decoding and progressively developing the lexical reading route. The investigation of reading development in Turkish raised some important theoretical implications concerning our current understanding of visual word recognition. First, the ceiling-level finding of accuracy highlights that oral reading accuracy has received disproportionate attention within the reading literature. Where transparent orthographies are considered, it is largely an irrelevant issue given that ceiling-level word reading accuracy is achieved soon after the beginning of reading instruction. Related to this, when word reading fluency was explored, the effect of PA disappeared. It appears that phonological skills are related to literacy skills that involve decoding as indexed by a significant finding in accuracy measures and lack of a significant finding in word reading speed. While the role of phonological awareness on reading development is crucial, this thesis argues that for transparent orthographies, RAN may be the more central measure to consider.

Furthermore, the availability of both lexical and sublexical routes to children further highlights the arguments above concerning the adult lexical decision data. In addition to this, it appears that the transparent nature of the orthography facilitates the rapid development of phonological skills and subsequently the development of an efficient sublexical route. It is currently hypothesised that this ultimately leads to the establishment of orthographic representations.

In addition, RAN (Wolf & Bowers, 1999) was related both to decoding and to sight word reading as indexed by the significant findings of RAN's role in both word and nonword reading speed. As word and nonword naming speed are considered to tap into separate skills, it could be considered that RAN may be involved in both the essential function of fluently converting visual stimuli into their corresponding phonological representations, i.e., phonological decoding as well as providing rapid access to the mental lexicon (Bowers & Wolf, 1993). These findings, while requiring further specific enquiry, tentatively add support to the growing literature that RAN is involved in both word and nonword reading as reported in several orthographies such as Dutch (van den Boer, de Jong, & Haentjens-van Meeteren, 2013), German (Moll et al., 2009) and Spanish (Onochie-Quintanilla, Defior, & Simpson, 2017).

Additionally, Visual Attention (VA) span appears to play a vital role in decoding speed, as has been reported in studies in additional languages, such as Dutch (Van Den Boer, Van Bergen, & de Jong, 2015) and Basque (Antzaka, Acha, Carreiras, & Lallier, 2019). Conceivably this mechanism may function by processing multiple letter-clusters as single units. While the literature concerning VA span's influence in reading development is still in its infancy, the findings of this thesis highlight the importance of this skill in reading development. It is postulated by the author that the agglutinative nature of Turkish, in particular, requires the rapid development of VA span in order to facilitate the accurate and rapid recognition and pronunciation of increasingly long words. In sum, as a result of the agglutinative nature of Turkish writing, reading could feasibly require increased sensitivity to morphological and syntactic structures at the word level. The specific influence of VA span on agglutinative writing systems remains an outstanding question in the domain.

Additionally, the inclusion of a measure of oral reading fluency was warranted as the measure was able to differentiate good from poor readers using a single index measure. Using, an arbitrary SD cut-off value of 1.25, Oral Reading Fluency (ORF), as a global index of reading ability, allowed for the distinction to be made between good and poor readers. Furthermore, Chapter 4 found that PA, RAN and VA Span were all significant predictors of ORF. Additionally, the results of Chapter 4 reported that increased phonological awareness skills predict reading fluency irrespective of age and as such offer support for the view that phonological processes continue to contribute to the efficiency of word recognition processes even in fluent readers (e.g., Rayner et al. 2012). When considering the influence of RAN on ORF, the findings of Chapter 4 indicate that RAN is a robust predictor of ORF (Christo & Davis, 2008; Papadopoulos, Spanoudis, & Georgiou, 2016). Furthermore, the parallel development of reading automaticity, as measured by RAN and ORF may be indicative of shared mental resources, i.e., domain-general factors such as serial processing and articulation (Georgiou, Aro, Liao, & Parrila, 2016).

Chapter 4 also signified that ORF is a reliable and valid measure of reading in Turkish grades 2–5 (aged from 86 to 151 months) and is relatively easy and fast to administer. As such, ORF measures may contribute to the early identification of Turkish students at risk for reading difficulties, and this discovery marks an important novel finding of the current thesis. Additionally, the findings of Chapter 4 seem to corroborate the position that reading speed is a superior index of reading than accuracy in transparent orthographies. Taken together, the results of Chapter 4 extend the predominantly European alphabetic findings of the influence of orthographic transparency on reading development.

The study in Chapter 4 was conducted with children attending several different schools from one district, and as participation was voluntary, and the participation rate was moderately low, a selection bias cannot be excluded. The findings of the current study need to be replicated with a feasibly, more representative sample. Second, all of the participants taking part in the current study already mastered the alphabetic principle before the start of this study; in order to fully explore earlier relationships between predictors and the Turkish outcome variables, future studies that include younger children are needed. Related to this, as the present study was concurrent, the causal role of how lexical and cognitive factors develop over time need to be addressed using a longitudinal study design that follows students over several years. Additionally, the lack of standardised measures led to the development of measures for each cognitive skill that was hypothesised to be involved in Turkish reading. While this approach was appropriate for this study, future experimental research could address this limitation following the standardisation of the measures developed in this work. This represents perhaps the principal future direction of the current doctoral work.

Also, related to this, the findings of the current study are limited to the specific set of measures used, i.e., it is entirely plausible that a different set of measures for the constructs measured would yield different results and therefore warrants further investigation. The null findings of Phonological Short-Term Memory (PSTM) are challenging to reconcile, given that, they could be due to a poorly conceived measure design through digit span tasks are widely used in the literature. Alternatively, the findings could be reflective of ceiling level phonological processing development in Turkish children (see Babayiğit & Strainthrop, 2007, for more details). Within the literature, there are mixed findings regarding the role of PSTM. For example, Parrila et al., (2004) and Torgesen et al. (1997) found that when considered along with PA and RAN, PSTM was only weakly associated with reading measures. Conversely, Swanson and colleagues (Swanson & Alexander,1997; Swanson & Howell, 2001) report that the contribution of PSTM to reading was significant. Separately, some of the constructs used within the current investigation were conceptualised using only one relevant measure. Therefore, future studies will need to incorporate additional measures in order to increase construct validity (Landerl et al., 2013). Future research would benefit from including well-designed measures of morphological awareness in studies exploring reading development in Turkish. The bottom-up approach used for the Linear Mixed Model (LMM) analysis was one of personal preference if the alternative top-down approach had been adopted, it is feasible that results would be different.

The mixed results regarding, and need for transformation of, the word frequency measures could be reflective of the complexity of word frequency measures in agglutinating languages such as Turkish. Consequently, there is a need to further define novel word frequency measures as the use of surface frequency may not be sufficient to characterise Turkish psycholinguistic data fully. That is, while surface frequency is suitable for the investigation of isolating writing systems such as English, the psycholinguistic investigation of Turkish likely requires sublexical considerations of frequency such as root or base frequency, given the rich morphological structure of agglutinative orthographies. This, in turn, would provide further clues as to the structure of the mental lexicon concerning the structure of stored lexical entries. This particular topic would also provide a rich platform for the exploration of different models of reading and reading development.

Furthermore, there is a need to broaden the linguistic and cognitive domains under exploration. For example, morphological awareness and word form prevalence (Keuleers, Stevens, Mandera, & Brysbaert, 2015) are excellent candidates for further investigation for predictors in Turkish-speaking populations. For instance, the role of morphological awareness may be particularly important to investigate in agglutinative orthographies (Acha et al., 2010). In one of the few studies of morphological awareness in Turkish, Durgunoğlu (2003) proposed that the rich morphological structure of Turkish may be best addressed by the use of a left-to-right computational strategy. Additionally, Fowler, Feldman, Andjelkovic, and Öney (2003) suggest that phonological predictability could play a more crucial role than semantic relatedness in the acquisition of distinctive types of morphology. Stimulatingly, it will be important to gain insight into whether being exposed to morphologically complex writing systems effects the development of visual attentional resources and word reading strategies. Finally, as this was a monolingual study, the extent to which the results can reflect comparability between the different orthographies is limited. For future studies, cross-language studies are of particular importance. Furthermore, the presence of vowel harmony and agglutination as characteristics of the Turkish language (and writing system) offer interesting avenues of investigation that can potential contribute to our current understanding of word level processing.

7.1.4 Developmental Dyslexia in Turkish-speaking children

Chapter 5 examined differing theoretical accounts of Developmental Dyslexia (DD), such as the Phonological Deficit Hypothesis (PDH) (Bradley & Bryant, 1983) and the visual attention span deficit hypothesis (Bosse et al., 2007), taking into consideration the influence of orthographic transparency on reading development. Subsequently, the Chapter reviewed both cognitive predictors and the diverse conceptualisations of DD subtypes. Comparisons between DD children and chronological and younger TD controls were reported using both a group and multiple case study approach. Chapter 5 also explored the factors influencing reading, cognitive profiles and subtypes of DD in a group of Turkish-speaking children. In a similar vein to Chapter 4, reading was studied at both the word and text levels. As predictors, the effect of length, lexicality and frequency (and their interactions) were explored as well as considering the influence of a set of cognitive predictors that have been indicated to influence word and text reading such as phonological awareness, rapid naming, and visual attention span. To this end, the current study found evidence of main effects of both lexicality and length with mixed evidence for frequency for both word/pseudoword reading accuracy and RT. The presence of length effects in a transparent orthography is considered to be reflective of the use of serial sublexical coding processes (Coltheart et al., 2001; Weekes, 1997) and is congruent with previous reports from Italian (Zoccolotti et al., 2005) and Spanish children (Davies et al., 2013). Furthermore, the current study found that word length effects were present in all three groups of children concerning their word/ pseudoword reading accuracy and RTs. Davies and colleagues (2013) propose that this finding is indicative of the role of sublexical processing in transparent orthographies as even the older Typically Developing (TD) children could not avoid the effect of word length on the time needed to utter words. Chapter 5 results also revealed that phonological awareness (PA), rapid automatized naming (RAN), visual attention span (VA Span), working memory and visuo-spatial short-term memory could all significantly differentially contribute to the cognitive deficits associated with reading disorder in Turkish. Also, it appears that RAN along with VA span were the most critical cognitive predictors of DD in Turkish-speaking children learning to read.

In addition, Chapter 5 explored the presence of subtypes of DD within the current cohort. Considering the existence of subtypes of DD in Turkish, overall, the outcomes of the current study lend support to the existence of distinct behavioural types of DD in Turkish. The presence of subtypes in transparent orthographies is further supported by studies in Greek (Douklias et al., 2009; Niolaki, Terzopoulos, & Masterson, 2014), Italian (Zoccolotti et al. (1999) and Spanish (Jiménez, Rodríguez, & Ramírez, 2009). However, the individual cognitive profiles of DD explored in this thesis paint a different picture. That is, there seems to be a mismatch in terms of the behavioural and cognitive manifestation of DD. Overall, no distinct cognitive profile of deficit or delay emerged during the subtyping investigation. Rather while both surface and phonological subtypes of DD can be detected in transparent and opaque alphabetic orthographies (Hanley, 2017) the findings of Chapter 5 indicate that the manifestation of DD in Turkish-speaking children is heterogeneous, and the majority of children in the DD subgroup exhibited either double or multiple deficits and therefore providing further support for the multiple-deficit hypothesis for Turkish developmental dyslexia.

To this end, the results underlined the independent role of RAN and phonological skills in predicting reading accuracy and speed. This is in line with the double-deficit theory of dyslexia (Wolf & Bowers, 1999), where RAN and phonological skills are considered to be two separate sources of reading difficulties. Further to this, the findings of Chapter 5 contribute to a number of topics concerning the underlying cognitive and linguistic mechanisms of reading disorder in Turkish-speaking children. To the best of the author's knowledge, the current study represents the most comprehensive attempt to characterise reading disorder in Turkish at both group and individual levels. When the DD group was considered as a whole, the current investigation found evidence that children in the DD group were slower than TD children at reading at both the text and single word levels, although their word/pseudoword reading was relatively accurate. This finding is mostly congruent with results from a number of studies on transparent orthographies such as Greek (Nikolopoulos, Goulandris, & Snowling, 2003), Italian (Zoccolotti et al., 1999) and Spanish (Davis, Cuetos, & Glez-Seijas, 2007).

Chapter 5 was also based on a cross-sectional design. While there is a need for this type of research design to understand better the role of these cognitive skills to word reading in DD, there is also a need for longitudinal studies in order to examine the relative contribution of these cognitive skills over time. Another notable limitation is the relatively small sample size of the current DD cohort. Future studies will need to recruit a larger representative sample of Turkish-speaking children with DD to explore several of the findings in the current Chapter further.

While it is proposed that group-based comparisons in DD research can be biased against specific subtypes of dyslexia (Wybrow, 2014), the further development of multiple-case designs in Turkish using large sample sizes will go some way to address this bias. There is also a wide range of methods currently utilised in DD research to identify subtypes such as the regression-outlier and use of z-scores, both of which were incorporated in the current thesis. However, there is relatively little consensus as to how research can differentiate typical from atypical reading performance, and as such, the comparability of different studies is often meaningless. While this work has mainly been Anglo- and Euro-centric to date, it is anticipated that input from a large variety of different orthographies can help facilitate advances in this important domain.

Finally, as this was a monolingual study, the extent to which the results can reflect comparability between the different orthographies is limited. For future studies, cross-language studies on DD are also of particular importance.

7.1.5 TOWARDS DEVELOPING COMPUTATIONAL MODELS OF VISUAL WORD RECOGNITION AND READING ALOUD IN TURKISH

Chapter 6 assessed several versions of the DRC model (e.g., child DRC model, DD subtype models) by piloting a quantitative evaluation of the models' responses to those of Turkish-speaking children with and without DD. This computational modelling work aimed to evaluate the utility of the DRC model framework for Turkish-speaking children's RT data as well as to explore further the DD profiles and current subtyping approaches to Turkish. Chapter 6 provides an overview of the first computational study to evaluate a series of DRC models of visual word recognition and reading aloud with Turkish stimuli by directly measuring modelling alongside behavioural data. The computational work undertaken within this Chapter, while exploratory, produced good fitting models of the average Turkish-speaking children's RT data as well as adequate models of grade-based average RT data and subtypes of DD. Overall, the DRC models produced were generally useful in simulating several aspects of the behavioural data.

However, a number of limitations were also identified. In particular, several of the DRC models displayed sensitivity to length, lexicality and frequency effects (which were absent in the behavioural data). It is feasible that the false-positive findings in the models were reflective of non-optimised model parameters or reflect a more critical issue in that the DRC model, being developed for English, may offer a poor fit to non-English behavioural data. Despite the limitations of the current modelling work, the simulations carried out in Chapter 6 provide a rich foundation for further work on computational models of reading, reading development and disorder among Turkish-speaking populations.

In addition to some of the limitations highlighted in Chapter 5 concerning the behavioural data such as the small sample size, there are also a number of limitations in Chapter 6. Firstly, the use of averaged models and RT datasets reintroduces an issue previously addressed in Chapters 4 and 5. That is the reliance on averaging across participants or items. Coupled with the small sample size, the findings of the current study can be accepted as exploratory at best. Perhaps the most substantial challenge in the current computational work was the suboptimal optimisation procedure used as a consequence of limited computational power and time. While several models captured a good amount of variance of their respective RT data, several also failed.

In future, to achieve superior fitting models to individual participants will require data of greater precision at an individual level (Adelman & Brown, 2008). Similarly, future studies could address this limitation by adopting an alternative optimisation procedure such as the one reported in Adelman and Brown (2008). In an attempt to address the so-called optimal parameters problem, Adelman and Brown applied the Nelder-Mead simplex procedure to the DRC. The Nelder-Mead algorithm (Nelder and Mead, 1965) is an example of a standard non-gradient local search method which inspects the corners of a region of interest, i.e., the parameter space of a model and subsequently expands into parameter spaces that represent improved model fit while also moving away from regions that fail to improve the model. In doing so, the DRC model can be optimised to a state where any observed effects or predictor variables are a consequence of a given theoretical framework rather than an artificial obstacle to falsification brought about by suboptimal approaches to model fitting.

Developmental models of reading represent a fascinating enterprise that has a rich history in connectionist approaches (see Chang, Monaghan & Welbourne, 2019 for a recent implementation) but has only recently received increasing interest in dual-route accounts of reading (Pritchard, 2012) There are also further prospects to carry out further research using both the GPC learning and self-teaching models (Pritchard, 2012). While the approach taken to self-teaching was akin to one-shot learning, a more psychologically plausible approach would be to incorporate a process of incremental training that starts with a small number of words and eventually builds a full representative vocabulary over time. In a similar line of enquiry, the recently implemented CDP++ developmental model of learning to read (Ziegler, Perry & Zorzi, 2014) has made some important progress in this direction. Using this approach, the authors achieved an accuracy rating of 80% for the 32,000 words that the model was exposed to after training.

While the default parameters of the self-teaching model were used for model development in the current work, future studies need to address the distinct lack of vocabulary growth data in Turkish-speaking primary school children. Following this, a more fine-grained exploration of vocabulary change, orthographic learning and self-teaching in Turkish can take place. In line with this, future computational modelling work in Turkish needs to take a closer look at features of phonology as there is much debate surrounding this topic as mentioned above (see Chapter 2 of this thesis; Koşaner, Birant, & Aktaş, 2013). Finally, while there is a widely accepted view in the domain regarding the relevant dependent variables of interest being a function of orthographic transparency, i.e. transparent orthographies index RT whereas opaque orthographies index accuracy, there is also a need to take potential speed-accuracy trade-offs into account. Future behavioural and computational studies can adopt the response signal paradigm (Reed, 1973) in order to address this.

The recent move towards modelling individual differences (Ziegler, Perry & Zorzi, 2019) in order to increase our understanding of the heterogeneity of DD represents a critical line of enquiry with computational models. Using this approach, the authors were able to develop a model that could successfully simulate the individual learning trajectories of 622 children (388 with DD). The authors argue that only by adopting personalised computational models that allow for multiple deficits can the heterogeneity and individual differences in DD profiles be captured. Such models may also have significant implications on the early identification of DD as well as being informative regarding the outcomes of interventions.

7.3 Theoretical Implications and Conclusion

The present doctoral research has addressed some outstanding questions regarding visual word recognition across a number of domains in the cognitive sciences. Using a computational linguistic method, this thesis explored current definitions of orthographic transparency and novel means of quantifying orthography, extending this approach to Turkish. The models produced stipulate that Turkish is more transparent than any other alphabetic orthography that has been quantified to date. To this end, it is important to consider orthographic transparency as a multifaceted construct and, in absolute terms, further research in extremely transparent orthographies can be used to extend reading research beyond the narrow Anglo- and Eurocentric research that has been dominated the domain to date. The extreme orthographic transparency of Turkish, therefore, serves as an excellent medium to test theories of visual word recognition, and any universal framework would need to account for the variation found in the writing systems of the Turkic family. Additionally, this thesis examined the currently available resources for Turkish psycholinguistic research and in response to the discovery of a lack of resources in the domain, has led to the creation of a Turkish lexicon database. It is anticipated that this much needed resource will stimulate further exploration of Turkish psycholinguistics in a structured and focussed way.

Overall the studies included in this thesis aimed to add to the growing field of reading, reading development and disorder of Turkish-speaking monolingual children and adults; however, this thesis has generated additional questions that demand exploration. For example, the finding of significant length by lexicality effects on both adults and children indicates that two routes of reading are available to Turkish-speakers. This finding lends general support to dual-route theories of reading and their applicability to extremely transparent orthographies. Related to this, Turkish presents a number of challenges for cross-linguistic theories of reading. Specifically, the presence of two routes to reading in children and adults contests the strong position of the ODH as this thesis presents a strong case for both lexical and sublexical reading strategies in Turkish.

In addition, a number of interactions emphasise both a diminishing effect of word length as a function of age and persistent effects of word length across children. Taken together, this provides evidence of a gradual shift from sublexical to lexical reading strategies without the full transition to lexical reading. This finding then suggests that the transparency of writing systems may modulate the nature of the relationship between the lexical and sublexical routes. In Turkish, it is hypothesised that while some words are read predominantly via either the lexical or sublexical route, the majority of words can be read using either strategy. As such, Turkish reading may not demand the simultaneous activation of both the lexical and the sublexical routes to generate phonology and instead, is suggestive of a highly flexible criterion account of reading strategy. While this thesis makes no specific claim as to what this account is, it does stress that the DRC model of reading would need to be modified to accommodate these findings. In line with this, Besner and others (O'Malley and Besner, 2008; Reynolds and Besner, 2008) argue that the DRC model should feature a threshold mode of processing in addition to a separate route-change control mechanism.

In a separate finding, orthographic (and phonological) neighbourhood density were inhibitory in Turkish lexical decision in contrast with the majority of the literature concerning this effect. This was hypothesised to be driven by the complex morphology of Turkish in which new words are formed with the addition of suffixes. This is particularly pertinent for models that incorporate the IA model into their lexical route i.e. DRC and CDP models. Within these models the faciliatory neighbourhood density effect is attributed to downstream feedback activation from word to features. Future modelling endeavours will need to explore if the direction of activation is reversable within these models. The failure of dual-route models to accommodate such a finding in Turkish would demand either a revision of the lexical route framework or seeking out alternative modelling approaches.

The findings of the reading development chapter indicate that increased phonological awareness skills predict reading fluency irrespective of age and as such offers general support for phonological theories of reading development as well as the likely universal role of phonological awareness in reading. As such, the findings of the rapid development of phonology in Turkish as well as the use of two distinct strategies in single-word reading lend support to the weak versions of the phonological and orthographic depth hypothesis of reading. Taken together, the results of the current study extend the predominantly European alphabetic findings of the influence of orthographic transparency on reading development.

Considering the role of cognitive predictors in reading development, this thesis made a number of interesting discoveries. For example, RAN, PA and VA Span were found to be exclusively influential on reading speed. However, when considering word reading speed only, the effect of PA disappeared. It appears that phonological skills are related to literacy skills that involve decoding as indexed by a significant finding in nonword reading speed and lack of a significant finding in word reading speed. This finding highlights both the universal aspects of cognitive predictors for reading development which appear to be present in all orthographies studied to date and the language-specific variation in these predictors in terms of their predictive value. VA span's role in ORF appears to diminish in older students. That is, for Turkish, it appears that the contribution of VA span to reading beyond the single-word level may be time-limited. It is feasible to postulate that this may be due to a language-specific feature of Turkish such as clearly defined syllable boundaries or agglutination or even a combination of both. In transitioning from sublexical to lexical reading strategies, the need to develop a VA span beyond three (the largest syllable length) would serve little advantage for Turkish children learning to read. Additionally, the mediation analysis within this thesis, argues that VA span may be driven by domain-general skills such as WM and Visuo-Spatial STM indicating the cognitive limitations on visual attention.

The current work was also extended to the examination of reading disorder, i.e., Developmental Dyslexia in Turkish children. When the DD group was considered as a whole, the current investigation found evidence that children in the DD group were slower than TD children at reading at both the text and single word levels, although their word/pseudoword reading was relatively accurate. This finding emphasises that accuracy is relatively well preserved in children learning to read in a transparent orthography and that reading speed and fluency are more sensitive measures to differences between TD and DD readers. Additionally, when the random intercept of length was removed from the random-effects model, a distinct group by length interaction emerged. This was seen to be suggestive that older TD readers, manifested with a significantly reduced word length effect on their word reading RTs when compared to the DD group. Taken together, the main effects and interactions found suggest that although Turkish children with DD have both routes available to them, there is an overreliance on sublexical processing.

It also emerged that both surface and phonological subtypes of DD can be detected in Turkish at the behaviour level though this is not supported at a cognitive level. That is, when considering subtypes, no distinct cognitive profile that distinguished between the groups emerged. This therefore suggests that while there is support for the use of subtyping approaches at the behavioral level, the utility of such approaches in understanding the cognitive developmental profile of children with DD is limited. Instead, the findings of the current study suggest that the manifestation of DD in Turkish-speaking children is heterogeneous, and the majority of children in the DD subgroup exhibited either double or multiple deficits and therefore providing further support for the multiple-deficit hypothesis for Turkish developmental dyslexia. There is a need then to move beyond the phonological-surface continuum and explore alternative accounts that better characterise the cognitive manifestation of DD.

This thesis also supplemented the behavioural data with the development of a computational model of visual word recognition in Turkish, the first of its kind. This is particularly important as an essential consideration of any modelling approach has to examine whether it can account for findings across languages. While the Turkish child DRC model provides a better fit to human RT data than previously reported DRC models in other orthographies, a number of divergences between the RT and DRC data were observed. This was not particularly surprising given the extreme transparency of the Turkish orthography.

The models particularly struggled with a marked oversensitivity to lexicality effects and under sensitivity to length effects in known words and may reflect a divergence between Turkish readers and the DRC architecture. The models performed particularly poorly when only words were considered which again highlights the limitations of applying IA modelling approaches to Turkish words via the lexical route. It may be that given the agglutinative nature of Turkish that the structure of the mental lexicon would be distinct in that whole word representations would not by psychologically plausible and instead maintain a structure that reflects a combined index of root words and suffixes.

The evaluation of the individual models suggests that as the majority of models exhibited multiple deficits across the DRC's representational levels, the strongest theoretical argument regarding the manifestation of DD, in models of Turkish-speaking children, is provided by the multiple deficit model of DD. The considerable heterogeneity in both the number and magnitude of deficits across subjects further lends support to the concept of multiple probabilistic risk factors. The combined computational work in this thesis, therefore, further confirms the usefulness of Turkish as an orthographic medium for further computational investigation of visual word recognition.

Overall, it is expected that this thesis will provide a stimulus for further research into the processes involved in reading, reading development and disorder in Turkish. This, as stated above, will help inform psycholinguistic theory in a broader sense and contribute to the realisation of suitable assessments of reading development and identification of reading disorder in Turkish.

References

Aaron, P. G., & Joshi, R. M. (2006). Written language is as natural as spoken language: A biolinguistic perspective. *Reading Psychology*, *27*(4), 263-311.

Acar, E. A., Zeyrek, D., Kurfali, M., & Bozsahin, C. (2016). A Turkish Database for Psycholinguistic Studies Based on Frequency, Age of Acquisition, and Imageability. In *LREC*.

Acha, J., Laka, I., & Perea, M. (2010). Reading development in agglutinative languages: Evidence from beginning, intermediate, and adult Basque readers. Journal of Experimental Child Psychology, 105(4), 359-375.

Adali, K., & Eryiğit, G. (2014, April). Vowel and diacritic restoration for social media texts. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)* (pp. 53-61).

Adams, M. J. (1990). Beginning to read: Learning and thinking about print. Beginning to Read: Learning and Thinking about Print,

Adelman, J. S., & Brown, G. D. (2008). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. Journal of Memory and Language, 59(4), 524-544.

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological science*, *17*(9), 814-823.

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H., Atasoy, G., Öz, S., Yildiz, I. & Kurtoglu, Ö. (2012, May). Construction of the Turkish National Corpus (TNC). In *LREC* (pp. 3223-3227).

Akyol, H., & Temur, T. (2006). İlköğretim üçüncü sınıf öğrencilerinin okuma düzeyleri ve sesli okuma hataları. *Ekev Akademi Dergisi*, *29*(25), 9-274.

Akyol, H., & Yıldız, M. (2010). Okuma bozukluğu olan bir öğrencinin okuma ve yazma becerisinin geliştirilmesine yönelik bir durum çalışması. 9. *New World Sciences Academy*, *5*(4), 1690-1700.

Allor, J. H. (2002). The relationships of phonemic awareness and rapid naming to reading development. *Learning Disability Quarterly*, *25*(1), 47-57.

Altinok, D. (2016). Towards Turkish ASR: Anatomy of a rule-based Turkish g2p. arXiv preprint arXiv:1601.03783.

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2018). Reading prosody development in Spanish children. Reading and Writing, 31(1), 35-52.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439-461.

Andrews, S., & Davis, C. (1999). Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap?. *Brain and Language*, *68*(1-2), 355-361.

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. Monographs of the society for research in child development, i-186.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. Psychological review, 105(4), 678.

Ansari, D., & Karmiloff-Smith, A. (2002). Atypical trajectories of number development: A neuroconstructivist perspective. *Trends in cognitive sciences*, *6*(12), 511-516.

Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. Annals of Dyslexia, 57(2), 113-137.

Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. Reading and Writing, 25(6), 1283-1305.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: a meta-analysis. *Journal of Educational Psychology*, *107*(3), 868.

Arduino, L. S., & Burani, C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. Journal of psycholinguistic research, 33(1), 75-95.

Arnell, K. M., Joanisse, M. F., Klein, R. M., Busseri, M. A., & Tannock, R. (2009). Decomposing the relation between Rapid Automatized Naming (RAN) and reading ability. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *63*(3), 173.

Aro, M., Huemer, S., Heikkilä, R., & Mönkkönen, V. (2011). Sujuva lukutaito suomalaislapsen haasteena. *Psykologia 46 (2011): 2-3*.

Arslan, D., & Dirik, M. Z. (2008). İlköğretim ikinci sınıf öğrencilerinin okuma güçlüklerini gidermede okumayı geliştirici duyuşsal yaklaşımın etkisi. Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 21(1), 1-18.

Avdyli, R., Castejón, L., & Cuetos, F. (2014). Lexical effects in word naming in Spanish children. The Spanish journal of psychology, 17.

Avdyli, S. R., & Cuetos, S. F. (2013). SUBTLEX-AL: Albanian word frequencies based on film subtitles. *ILIRIA International Review*, *3*(1), 285-292.

Awadh, F. H., Phénix, T., Antzaka, A., Lallier, M., Carreiras, M., & Valdois, S. (2016). Cross-language modulation of visual attention span: an Arabic-French-Spanish comparison in skilled adult readers. *Frontiers in psychology*, *7*, 307.

Baayen, R. H. (2001). *Word frequency distributions* (Vol. 18). Springer Science & Business Media.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. International Journal of Psychological Research, 3(2), 12-28.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, 59(4), 390-412.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174-1220.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The {CELEX} lexical data base on {CD-ROM}.

Babayiğit, S., (2006) The longitudinal and concurrent predictive role of phonological awareness in literacy development: evidence from Turkish. Unpublished Doctoral thesis, Institute of Education, University of London.

Babayiğit, S., & Konedrali, G. (2009). KKTC'deki ilk okuma-yazma öğretimi ile ilgili uygulamalar ve bunlara ilişkin öğretmen görüş ve değerlendirmeleri [Teacher views and evaluations related to the elementary reading instruction practices in the Turkish Republic of Northern Cyprus]. http://eprints.uwe.ac.uk/8426/

Babayiğit, S., & Stainthorp, R. (2007). Preliterate phonological awareness and early literacy skills in Turkish. Journal of Research in Reading, 30(4), 394-413.

Babayiğit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. Reading and Writing, 23(5), 539-568.

Babayiğit, S., & Stainthorp, R. (2011). Modeling the relationships between cognitive–linguistic skills and literacy skills: New insights from a transparent orthography. Journal of Educational Psychology, 103(1), 169.

Babur, N., Haznedar, B., Ercetin, G., Ozerman, D., & Erdat-Cekerek, E. (2011). T¨urkc¸e'de kelime okuma bilgisi testinin (KOBIT) gelis¸tirilmesi [Development of theword reading knowledge test in Turkish (KOBIT)]. Bogˇazic¸i U¨ niversitesi Egˇitim Dergisi [Boğaziçi University Journal of Education], 28, 1–15.

Backman, J., Bruck, M., Hebert, M., & Seidenberg, M. S. (1984). Acquisition and use of spelling-sound correspondences in reading. Journal of Experimental Child Psychology, 38(1), 114-133.

Bacon, A. M., Parmentier, F. B., & Barr, P. (2013). Visuospatial memory in dyslexia: Evidence for strategic deficits. Memory, 21(2), 189-209.

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of cognitive neuroscience*, *4*(3), 281-288.

Baddeley, A. (1993). Working memory or working attention?. *Selection, Awareness, and Control*, 152-170.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.

Baddeley, A. (2003). Working memory and language: An overview. Journal of communication disorders, 36(3), 189-208.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In Psychology of learning and motivation (Vol. 8, pp. 47-89). Academic press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 28-61). New York, NY, US: Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575-589.

Baddeley, A., Eysenck, M., & Anderson, M. (2009). Memory hove.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal and the relation with reporting errors and quality of psychological research. *PLoS One*, *9*(7), e103360.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of psycholinguistics*, *2*, 285-375.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B. & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445-459.

Baluch, B. (1993). Lexical decisions in Persian: A test of the orthographic depth hypothesis. *International Journal of Psychology*, *28*(1), 19-29.

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 644.

Barca, L., Burani, C., Di Filippo, G., & Zoccolotti, P. (2006). Italian developmental dyslexic and proficient readers: Where are the differences?. *Brain and Language*, *98*(3), 347-351.

Barca, L., Ellis, A. W., & Burani, C. (2007). Context-sensitive rules and word naming in Italian children. Reading and Writing, 20(5), 495-509.

Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly*, 335-345.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Basutçu, M., (2008). "Turkey. Consolidation", Cahiers du cinéma (Special issue 2008: World Cinema Atlas): 97, ISSN 0008-011X

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. Memory & Cognition, 29(7), 986-999.

Baydık, B. (2002). Okuma güçlüğü olan ve olmayan öğrencilerin sözcük okuma becerilerinin karşılaştırılması. Yayınlanmamış doktora tezi, Ankara Üniversitesi, Ankara.

Beech, J. R. (2005). Ehri's model of phases of learning to read: A brief critique. *Journal of research in reading*, *28*(1), 50-58.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). Regression diagnostics: Identifying influential data and sources of collinearity. New York, NY: Wiley.

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 353.

Berg, T. (2014). On the relationship between type and token frequency. *Journal of Quantitative Linguistics*, *21*(3), 199-222.

Berg, T. (2016). The multiplanar nature of frequency. *Glottotheory*, *7*(1), 1-19.

Berninger, V. W., & Abbott, R. D. (1994). Multiple orthographic and phonological codes in literacy acquisition: An evolving research program. The varieties of orthographic knowledge (pp. 277-319) Springer.

Berninger, V. W., & Wolf, B. (2009). Helping students with dyslexia and dysgraphia make connections: Differentiated instruction lesson plans in reading and writing. ERIC.

Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. Journal of Psycholinguistic Research, 39(2), 141-163.

Besner, D. (1999). Basic Processes in Reading: Multiple. *Converging*

methods for understanding reading and dyslexia, *413*.
Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*(5), 467-472.
Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 468.
Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *34*(2), 313-325.
Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary?.
Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of educational psychology*, *93*(3), 498.
Bilgin, O. (2016). *Frequency effects in the processing of morphologically complex Turkish words* (Master's thesis. Boğaziçi University).
Bingöl, A. (2003). Ankara'daki ilkokul 2. ve 4. sınıf öğrencilerinde gelişimsel disleksi oranı. İnsan Kaynakları Yönetimi, Beta Yayınları. Baskı, İstanbul, 322.
Blaiklock, K. E. (2004). The importance of letter knowledge in the relationship between phonological awareness and reading. Journal of Research in Reading, 27(1), 36-57.
Boets, B., De Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., & Ghesquiere, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. British Journal of Developmental Psychology, 28(1), 5-31.
Boets, B., Wouters, J., Van Wieringen, A., & Ghesquiere, P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory. *Neuropsychologia*, *45*(8), 1608-1620.
Borgwaldt, S. R., Hellwig, F. M., & de Groot, A. M. (2004). Word-initial entropy in five languages: Letter to sound, and sound to letter. *Written Language & Literacy*, *7*(2), 165-184.
Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. M. (2005). Onset entropy matters–Letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*(3), 211-229.
Bosse, M. L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: a cross-sectional study. Journal of Research in Reading, 32(2), 230-253.
Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198-230.
Bosse, M., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. Journal of Research in Reading, 32(2), 230-253.
Bowers, P. G., & Newby-Clark, E. (2002). The role of naming speed within a model of reading acquisition. *Reading and Writing*, *15*(1-2), 109-126.
Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. Reading and Writing, 5(1), 69-85.
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). Oxford: Blackwell.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of experimental child psychology*, *92*(3), 203-219.

Bozkurt, T. N. (2017). The Effects of word length and suffixation on eye movement control in Turkish reading (Master's thesis, Middle East Technical University).

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature*, *271*(5647), 746-747.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. Nature, 301(5899), 419-421.

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), 96-99.

Brizzolara, D., Chilosi, A., Cipriani, P., Di Filippo, G., Gasperini, F., Mazzotti, S., Pecini, C. and Zoccolotti, P. (2006). Do phonologic and rapid automatized naming deficits differentially affect dyslexic children with and without a history of language delay? A study of Italian dyslexic children. *Cognitive and Behavioral Neurology*, *19*(3), 141-149.

Brown, L. A., Forbes, D., & McConnell, J. (2006). Short Article: Limiting the use of Verbal Coding in the Visual Patterns Test. Quarterly Journal of Experimental Psychology, 59(7), 1169-1176.

Brown, W. (1910) Some Experimental Results in the Correlation of Mental Abilities", *British Journal of Psychology, 3,* 296–322.

Brunswick, N., McDougall, S., & de Mornay Davies, P. (Eds.). (2010). *Reading and dyslexia in different orthographies*. Psychology Press.

Bryant, P., & Impey, L. (1986). The similarities between normal readers and developmental and acquired dyslexics. *Cognition*, *24*(1-2), 121-137.

Brysbaert, M., & Diependaele, K. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods*, *45*(2), 422-430.

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: are early-learned words more resilient?. *Aphasiology*, *30*(11), 1240-1263.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41*(4), 977-990.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. Journal of Cognition, 1(1).

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology*.

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, *2*.

Buchanan, M. A. (1927). *A graded Spanish word book* (Vol. 3). The University of Toronto Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523.

Bundesen, C. (1998). A computational theory of visual attention. Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences, 353(1373), 1271-1281.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological review*, *112*(2), 291.

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography?. Brain and language, 81(1-3), 568-586.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 272-277.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425-455.

Bybee, J. (2001). Frequency effects on French liaison. *Typological studies in language*, *45*, 337-360.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS one*, *5*(6), e10729.

Cangöz, B. (1999). Türkiye'de biliflsel psikoloji çal› flmalar› nda kullan› lan sözel malzemeler. *Türk Psikoloji Yaz› lar*, *2*(4).

Çapan, S. (1989). A linguistic study of reading and writing disorders in Turkish, an agglutinative language. In *Reading and writing disorders in different orthographic systems* (pp. 191-202). Springer Netherlands.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. Journal of memory and language, 45(4), 751-774.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M. and Seidlová-Málková, G. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological science*, *23*(6), 678-686.

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of experimental child psychology*, *92*(2), 107-139.

Carello, C., Lukatela, G., & Turvey, M. (1988). Rapid naming is affected by association but not by syntax. Memory & Cognition, 16(3), 187-195.

Carreiras, M., Álvarez, C. J., & Devega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of memory and language*, *32*(6), 766-780.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in cognitive sciences*, *18*(2), 90-98.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. Cognition, 47, 149–180.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read?. Cognition, 91(1), 77-111.

Castles, A., & Friedmann, N. (2014). Developmental dyslexia and the phonological deficit hypothesis. Mind & Language, 29(3), 270-285.

Castles, A., Bates, T., & Coltheart, M. (2006). John Marshall and the developmental dyslexias. *Aphasiology*, *20*(9), 871-892.

Castles, A., Datta, H., Gayan, J., & Olson, R. K. (1999). Varieties of developmental reading disorder: Genetic and environmental influences. Journal of experimental child psychology, 72(2), 73-94.

Çayci, B., & Demır, M. K. (2006). Okuma ve anlama sorunu olan öğrenciler üzerine karşilaştirmali bir çalişma. *Journal of Turkish Educational Sciences*, *4*(4).

Chang, L. Y., Chen, Y. C., & Perfetti, C. A. (2017). GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages. Behavior Research Methods, 1-23.

Chang, Y. N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language*, *108*, 104025.

Chen, C., Schneps, M. H., Masyn, K. E., & Thomson, J. M. (2016). The effects of visual attention span and phonological decoding in reading comprehension in dyslexia: a path analysis. *Dyslexia*, *22*(4), 322-344.

Christo, C., & Davis, J. (2008). Rapid naming and phonological processing as predictors of reading and spelling. *The California School Psychologist*, *13*(1), 7-18.

Clements, G. N., & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. *The structure of phonological representations*, *2*, 213-255.

Cohen, J. (1977). Statistical power for the behavioral sciences (revised edition).

Coltheart, M. (1978). Lexical access in simple reading tasks. *Strategies of information processing*, 151-216.

Coltheart, M. (1987). Varieties of developmental dyslexia: A comment on Bryant and Impey. *Cognition, 27*(1), 97–101.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1), 5-17.

Coltheart, M., & Leahy, J. (1996). Assessment of lexical and nonlexical reading abilities in children: Some normative data. *Australian Journal of Psychology*, *48*(3), 136-140.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological review, 100(4), 589.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, N.J: Lawrence Erlbaum, 1977.

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology*, *35*(3), 469-495.

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.

Compton, D. L., & Carlisle, J. F. (1994). Speed of word recognition as a distinguishing characteristic of reading disabilities. *Educational Psychology Review*, *6*(2), 115-140.

Computational statistics & data analysis, 52(12), 5186-5201.

Connelly, V. (2002). Graphophonemic awareness in adults after instruction in phonic generalisations. *Learning and Instruction*, *12*(6), 627-649.

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes*, *23*(2), 296-326.

Constantinidou, F., & Evripidou, C. (2012). Stimulus modality and working memory performance in Greek children with reading disabilities: Additional evidence for the pictorial superiority hypothesis. *Child Neuropsychology*, *18*(3), 256-280.

Cossu, G. (1999). The acquisition of Italian orthography. In M. Harris & G. Hatano (Eds.), Learning to read and write. A cross-linguistic perspective. Cambridge, U.K: Cambridge University Press.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied psycholinguistics*, *9*(1), 1-16.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482-486.

Cronin, V. S. (2013). RAN and double-deficit theory. Journal of Learning Disabilities, 46(2), 182-190.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2012). SUBTLEX-ESP: frecuencias de las palabras espanolas basadas en los subtitulos de las peliculas. *Psicológica*, *32*(2), 133-144.

Dave, H., (2011) Frequency Word Lists.

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & cognition*, *41*(2), 297-311.

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.

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.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior research methods*, *37*(1), 65-70.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*(4), 665-671.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. Journal of educational psychology, 91(3), 450.

De Jong, P. F., & Van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. Brain and Language, 80(3), 617-626.

De Luca, M., Burani, C., Paizi, D., Spinelli, D., & Zoccolotti, P. (2010). Letter and letter-string processing in developmental dyslexia. *Cortex*, *46*(10), 1272-1283.

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.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., . . . Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. Science, 330(6009), 1359-1364.

del Prado Martín, F. M., Ernestus, M., & Baayen, R. H. (2004). Do type and token effects reflect different mechanisms? Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and language*, *90*(1), 287-298.

Delic, E. (2004). Présentation du Corpus de référence du français parlé. *Recherches sur le français parlé*, *18*, 11-42.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. Neuropsychologia, 37(10), 1189-1199.

Démonet, J. F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *The Lancet*, *363*(9419), 1451-1460.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized'naming (RAN): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 14(4), 471-479.

Denckla, M. B., & Rudel, R. G. (1976a). Naming of object-drawings by dyslexic and other learning disabled children. Brain and Language, 3(1), 1-15.

Denckla, M. B., & Rudel, R. G. (1976c). Rapid 'automatized' naming (R.A.N.): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 14(4), 471-479. doi:http://dx.doi.org.v-ezproxy.brunel.ac.uk:2048/10.1016/0028-3932(76)90075-0

Denckla, M. B., Rudel, R. G., & Broman, M. (1981). Tests that discriminate between dyslexic and other learning-disabled boys. Brain and Language, 13(1), 118-129.

Dimensional Semantic Space. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 660-665. Hillsdale, NJ: Erlbaum.

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: The case of Greek. *Frontiers in psychology*, *1*.

Douklias, S. D., Masterson, J., & Hanley, J. R. (2009). Surface and phonological developmental dyslexia in Greek. *Cognitive neuropsychology*, *26*(8), 705-723.

Dulay, K. M., & Hanley, J. R. (2015). Stress errors in a case of developmental surface dyslexia in Filipino. *Cognitive neuropsychology*, *32*(1), 29-37.

Durgunoğlu, A. Y. (2002). Cross-linguistic transfer in literacy development and implications for language learners. Annals of Dyslexia, 52(1), 189-204.

Durgunoğlu, A. Y. (2003). Recognizing morphologically complex words in Turkish. In *Reading complex words* (pp. 81-92). Springer, Boston, MA.

Durgunoğlu, A. Y., & Öney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. Reading and Writing, 11(4), 281-299.

Duyck, W., Anseel, F., Szmalec, A., Mestdagh, P., Tavernier, A., &

Hartsuiker, R. J. (2008). Improving accuracy in detecting acoustic onsets. Journal of Experimental Psychology: Human Perception and Performance, 34(5), 1317.

Education Statistical Yearbook (2019) Ministry of National Education and Culture, TRNC.

Ehri, L. C. (1991). The development of reading and spelling in children: An overview. Dyslexia: Integrating Theory and Practice, , 63-79.

Ehri, L. C. (1992). *Reconceptualizing the development of sight word reading and its relationship to recoding.* Lawrence Erlbaum Associates, Inc.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. *Learning to Spell: Research, Theory, and Practice Across Languages, 13*, 237-268.

Ehri, L. C., & Soffer, A. G. (1999). Graphophonemic awareness: Development in elementary students. *Scientific Studies of Reading*, *3*(1), 1-30.

Eklund, K. M., Torppa, M., & Lyytinen, H. (2013). Predicting reading disability: early cognitive risk and protective factors. Dyslexia, 19(1), 1-10.

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. Reading and Writing, 8(6), 453-485.

Elbro, C., Møller, S., & Nielsen, E. M. (1995). Functional reading difficulties in Denmark. *Reading and Writing*, *7*(3), 257-276.

Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate* (No. 14). Cambridge University Press.

Ellis, A. W., McDougall, S. J., & Monk, A. F. (1996). Are dyslexics different? I. A comparison between dyslexics, reading age controls, poor readers and precocious readers. *Dyslexia*, *2*(1), 31-58.

Ellis, N.C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V.H., Polyzoe, N., Tsipa, M.L. and Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading research quarterly*, *39*(4), 438-468.

Erden, G., Kurdoğlu, F., & Uslu, R. (2002). Ilköğretim okullarina devam eden Türk çocuklarinin sinif düzeylerine göre okuma hizi ve yazim hatalari normlarinin geliştirilmesi. *Türk Psikiyatri Dergisi*.

Eren, E. (2014). *Frequency and length effects of words on fixation durations in Turkish sentence reading* (Unpublished Master's thesis).

Ergenç, İ. (1995). *Konuşma dili ve Türkçenin söyleyiş sözlüğü: bir deneme* (Vol. 8). Simurg.

Ergenç, İ. (2002). *Spoken Language and Dictionary of Turkish Articulation*. Multilingual.

Erten, B., Bozsahin, C., & Zeyrek, D. (2014, May). Turkish Resources for Visual Word Recognition. In LREC (pp. 2106-2110).

Eryigit, G. (2014). ITU Turkish NLP Web Service. In *EACL* (pp. 1-4).

Esen, A. ve Çifçi İ. (1998). Sınıf öğretmenlerinin öğrenme yetersizliğiile ilgili bilgilerinin belirlenmesi. Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 8, 95-101.

Everatt, J., Ocampo, D., Veii, K., Nenopoulou, S., Smythe, I., al Mannai, H., & Elbeheri, G. (2010). Dyslexia in biscriptal readers. In N. Brunswick, S. McDougall & P. de Mornay Davies (Eds.), (pp. 221-245). New York, NY US: Psychology Press.

Everatt, J., Ocampo, D., Veii, K., Nenopoulou, S., Smythe, I., al Mannai, H., & Elbeheri, G. (2010). 11 Dyslexia in biscriptal readers. *Reading and dyslexia in* *different orthographies*, 221.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behavior research methods, 41(4), 1149-1160.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M. & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488-496.

Fischbach, A., Könen, T., Rietz, C. S., & Hasselhorn, M. (2014). What is not working in working memory of children with literacy disorders? Evidence from a three-year-longitudinal study. *Reading and Writing*, *27*(2), 267-286.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627-635.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior research methods, instruments, & computers, 35(1), 116-124.

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract?. *Trends in cognitive sciences*, *15*(6), 272-279.

Fowler, A., Feldman, L. B., Andjelkovic, D., & Oney, B. (2003). Morphological and phonological analysis by beginning readers: Evidence from Serbian and Turkish. In *Reading complex words* (pp. 53-80). Springer, Boston, MA.

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives*, 1-37.

Frith, U. (1985). Beneath the surface of developmental dyslexia. Surface dyslexia, 32, 301-330.

Frith, U., Hulme, C., & Snowling, M. (1997). Dyslexia: Biology, cognition, and intervention. *Brain, Mind, and Behaviour. London: Whurr*.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology-Learning Memory and Cognition*, *20*(1), 116-129.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. Psychological Bulletin, 123(1), 71.

Frost, R. (2012). A universal approach to modeling visual word recognition and reading: Not only possible, but also inevitable. *Behavioral and brain sciences*, *35*(5), 310-329.

Frost, R., & Bentin, S. (1992). Processing phonological and semantic ambiguity: Evidence from semantic priming at different SOAs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 58.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific studies of reading, 5(3), 239-256.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language*, *28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition?. European Journal of Psychology of Education, 8(3), 259.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. Journal of experimental child psychology, 93(3), 265-281.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. Developmental psychology, 40(2), 177.

Genard, N., Mousty, P., Alegria, J., Leybaert, J., & Morais, J. (1998). Methods to establish subtypes of developmental dyslexia. In *Problems and interventions in literacy development* (pp. 163-176). Springer, Dordrecht.

Georgiou, G. K., Aro, M., Liao, C. H., & Parrila, R. (2016). Modeling the relationship between rapid automatized naming and literacy skills across languages varying in orthographic consistency. *Journal of Experimental Child Psychology*, *143*, 48-64.

Georgiou, G. K., Parrila, R., & Liao, C. H. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and writing*, *21*(9), 885.

Germano, G. D., Reilhac, C., Capellini, S. A., & Valdois, S. (2014). The phonological and visual basis of developmental dyslexia in Brazilian Portuguese reading children. *Frontiers in psychology*, *5*, 1169.

Gernsbacher A.M. (1984). Resolving 20 Years of Inconsistent Interactions Between Lexical Familiarity and Orthography, Concreteness and Polysemy. *Journal of Experimental Psychology: General,* **113** (2), 256-281.

Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior research methods*, *48*(3), 963-972.

Giofrè, D., Toffalini, E., Provazza, S., Calcagnì, A., Altoè, G., & Roberts, D. J. (2019). Are children with developmental dyslexia all the same? A cluster analysis with more than 300 cases. *Dyslexia*, *25*(3), 284-295.

Giovagnoli, G., Vicari, S., Tomassetti, S., & Menghini, D. (2016). The role of visual-spatial abilities in dyslexia: Age differences in children's reading?. *Frontiers in psychology*, *7*, 1997.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading

Gökçearslan, S., & Seferoglu, S. S. (2016). The use of the Internet among middle school students: Risky behaviors and opportunities. Kastamonu Eğitim Dergisi / Kastamonu Education Journal, 24(1), 383-401.

Göksel, A., & Kerslake, C. (2004). *Turkish: A comprehensive grammar*. Routledge.

González, J. E. J., & Santana, G. R. (2002). Identifying subtypes of reading disability in the Spanish language. *The Spanish Journal of Psychology*, *5*(1), 3-19.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language*, **24**, 631-645.

Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, *6*(2), 133-151.

Goswami, U. (2002). Phonology, reading development, and dyslexia: A cross-linguistic perspective. *Annals of dyslexia*, *52*(1), 139-163.

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. Trends in cognitive sciences, 7(12), 534-540.

Goswami, U. (2008). The development of reading across languages. Annals of the New York Academy of Sciences, 1145(1), 1-12.

Goswami, U. & Bryant, P.E. (1990). Phonological skills and learning to read. Hove, East Sussex: Psychology Press.

Goswami, U., & Bryant, P. (1990). Phonological skills and learning to read Wiley Online Library.

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*(1), 19-52.

Goswami, U., Porpodas, C., & Wheelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education*, *12*(3), 273-292.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. Annals of Dyslexia, 30(1), 179-196.

Goulandris, N. E. (2003). *Dyslexia in different languages: Cross-linguistic comparisons*. Whurr Publishers.

Göz, İ. (2003). *Yazılı türkçenin kelime sıklığı sözlüğü* (Vol. 823). Türk Dil Kurumu.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of memory and language*, *29*(2), 228-244.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological review*, *103*(3), 518.

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of educational psychology*, *94*(1), 34.

Gustafson, S., Ferreira, J., & Rönnberg, J. (2007). Phonological or orthographic training for children with phonological or orthographic decoding deficits. *Dyslexia*, *13*(3), 211-229.

Güven, S., & Friedmann, N. (2019). Developmental letter position dyslexia in Turkish, a morphologically rich and orthographically transparent language. *Frontiers in psychology*, *10*, 2401.

Guzmán, R., Jiménez, J.E., Ortiz, M.R., Valle, I.H., Estévez, A., Rodrigo, M., García, E., Díaz, A. and Hernández, S. 2004). Evaluación de la velocidad de nombrar en las dificultades de aprendizaje de la lectura. *Psicothema*, *16*(3), 442-447.

Habib, M. (2000). The neurological basis of developmental dyslexia: an overview and working hypothesis. *Brain*, *123*(12), 2373-2399.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis (Vol. 6). Upper Saddle River, NJ: Pearson

Handler, S. M., & Fierson, W. M. (2011). Learning disabilities, dyslexia, and vision. *Pediatrics*, *127*(3), e818-e856.

Hanley, J. R. (2017). Is there just one dyslexic reader? Evidence for the existence of distinct dyslexic sub-groups. *Current Developmental Disorders Reports*, *4*(4), 101-107.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological review*, *106*(3), 491.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, *111*(3), 662.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. *Handbook of metacognition in education*, 131-153.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher, 59(7), 636-644.

Hasko, S., Groth, K., Bruder, J., Bartling, J., & Schulte-Körne, G. (2013). The time course of reading processes in children with and without dyslexia: an ERP study. Frontiers in human neuroscience, 7, 570.

Hautala, J., Aro, M., Eklund, K., Lerkkanen, M. K., & Lyytinen, H. (2013). The role of letters and syllables in typical and dysfluent reading in a transparent orthography. *Reading and Writing*, *26*(6), 845-864.

Hautala, J., Hyönä, J., Aro, M., & Lyytinen, H. (2011). Sublexical effects on eye movements during repeated reading of words and pseudowords in Finnish. *Psychology of Language and Communication*, *15*(2), 129-149.

Heim, S., Tschierse, J., Amunts, K., Wilms, M., Vossel, S., Willmes, K., Grabowska, A. and Huber, W. (2008). Cognitive subtypes of dyslexia. *Acta neurobiologiae experimentalis*, *68*(1), 73.

Hendrix, P., Ramscar, M., & Baayen, H. (2019). NDRA: A single route model of response times in the reading aloud task based on discriminative learning. *PloS one*, *14*(7), e0218802.

Henmon, V. A. C. (1924). *A French word book based on a count of 400,000 running words* (No. 3).

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.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of memory and language*, *63*(3), 259-273.

Ho, C. S. H., Chan, D. W. O., Lee, S. H., Tsang, S. M., & Luan, V. H. (2004). Cognitive profiling and preliminary subtyping in Chinese developmental dyslexia. *Cognition*, *91*(1), 43-75.

Hofmann, M. J., Stenneken, P., Conrad, M., & Jacobs, A. M. (2007). Sublexical frequency measures for orthographic and phonological units in German. *Behavior Research Methods*, *39*(3), 620-629.

Holmes, J. M. (1973). Dyslexia: A neurolinguistic study of traumatic and developmental disorders of reading University of Edinburgh Unpublished Ph. *D thesis*.

Holopainen, L., Ahonen, T., & Lyytinen, H. (2001). Predicting delay in reading achievement in a highly transparent language. Journal of learning disabilities, 34(5), 401-413.

Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, *52*(12), 5186-5201.

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how?. *The Reading Teacher*, *58*(8), 702-714.

Huestegge, L., Heim, S., Zettelmeyer, E., & Lange-Küttner, C. (2012). Gender-specific contribution of a visual cognition network to reading abilities. British Journal of Psychology, 103(1), 117-128.

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. Scientific studies of reading, 9(4), 351-365.

Human Development Report. (2006). United Nations. Retrieved August 18, 2013, from http://hdrstats.undp.org/countries/data_sheets/cty_ds_TUR.html

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, *30*(2), 225.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages?. *Cognition*, *91*(3), 273-296.

Ikeda, K., Ueno, T., Ito, Y., Kitagami, S., & Kawaguchi, J. (2017). An Extension of a Parallel-Distributed Processing Framework of Reading Aloud in Japanese: Human Nonword Reading Accuracy Does Not Require a Sequential Mechanism. *Cognitive Science*, *41*, 1288-1317.

Inkelas, S. (1999). Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. *The prosody-morphology interface*, *134*, 187.

Inkelas, S., Küntay, A., Orgun, O., & Sprouse, R. (2000). Turkish electronic living lexicon (TELL). *Turkic Languages*, *4*, 253-275.

Inkelas, S., Küntay, A., Orgun, O., & Sprouse, R. (2000). Turkish electronic living lexicon (TELL). *Turkic Languages*, *4*, 253-275.

International Association for the Evaluation of Educational Achievement. (2001). Progress in International Reading Literacy Study.

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of memory and language, 59(4), 434-446.

Janssen, M., Bosman, A. M., & Leseman, P. P. (2013). Phoneme awareness, vocabulary and word decoding in monolingual and bilingual Dutch children. Journal of Research in Reading, 36(1), 1-13.

Jerman, O., Reynolds, C., & Swanson, H. L. (2012). Does growth in working memory span or executive processes predict growth in reading and math in children with reading disabilities? Learning Disability Quarterly, 35(3), 144-157.

Jimenez Gonzalez, J. E., Alvarez Gonzalez, C. J., Estevez Monzo, A., & Hernandez-Valle, I. (2000). Onset-rime units in visual word recognition in Spanish normal readers and children with reading disabilities. *Learning Disabilities Research & Practice*, *15*(3), 135-141.

Jiménez, J. E., Rodríguez, C., & Ramírez, G. (2009). Spanish developmental dyslexia: Prevalence, cognitive profile, and home literacy experiences. Journal of Experimental Child Psychology, 103(2), 167-185.

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America*, *132*(2), EL74-EL80.

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*, 703.

Jones, K., Castles, A., & Kohnen, S. (2011). Subtypes of developmental reading disorders: Recent developments and directions for treatment. *ACQ*, *13*(2), 79-83.

Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Attention, Perception, & Psychophysics*, *81*(5), 1189-1196.

Jørgensen, J. N. (2003). Linguistic construction and negotiation of social relations among bilingual Turkish-speaking adolescents in north-western Europe. *Journal of Multilingual and Multicultural Development*, *24*(1-2), 1-11.

Kaplan, B. J., Wilson, B. N., Dewey, D., & Crawford, S. G. (1998). DCD may not be a discrete disorder. *Human movement science*, *17*(4-5), 471-490.

Kapnoula, E. C., Protopapas, A., Saunders, S. J., & Coltheart, M. (2017). Lexical and sublexical effects on visual word recognition in Greek: comparing human behavior to the dual route cascaded model. *Language, Cognition and Neuroscience*, *32*(10), 1290-1304.

Karadağ, Ö., Keskin, H. K., & Arı, G. (2019). Mediatory Effect of Reading Skill in the Rapid Automatized Naming/Reading Comprehension Relationship. *Egitim ve Bilim*, *44*(197).

Karaman, D., Tūrkbay, T., Cöngöloğlu, A., & Gökçe, F. S. (2006). Özgül öğrenme bozukluğu ile binişik psikiyatrik bozukluklar. *Türkiye'de Psikiyatri*.

Kargın, T., Güldenoğlu, B., & Ergül, C. (2017). Anasınıfı çocuklarının erken okuryazarlık beceri profili: Ankara örneklemi. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi, 18(01), 61-87.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 157.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(1), 157.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in psychology*, *94*, 67-84.

Katzir, T., Schiff, R., & Kim, Y. S. (2012). The effects of orthographic consistency on reading development: A within and between cross-linguistic study of fluency and accuracy among fourth grade English-and Hebrew-speaking children. *Learning and Individual Differences*, *22*(6), 673-679.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of memory and Language*, *44*(4), 592-617.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. Journal of Memory and Language, 47(1), 145-171.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. Behavior research methods, 42(3), 627-633.

Keuleers, E., & Brysbaert, M. (2011). Detecting inherent bias in lexical decision experiments with the LD1NN algorithm. *The Mental Lexicon*, *6*(1), 34-52.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior research methods*, *42*(3), 643-650.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono-and disyllabic words and nonwords. *Frontiers in Psychology*, *1*.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287-304.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1665-1692.

King, W. M., Lombardino, L. L., & Ahmed, S. (2005). Accuracy and speed of orthographic processing in persons with developmental dyslexia. *Perceptual and motor skills*, *101*(1), 95-107.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, *45*(3), 341-362.

Kılıç, A. (2008). Avşar Ağızlarındaki Alıntı Kelimelerde Görülen Ses Değişmeleri. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *28*(3).

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of experimental psychology: General*, *133*(3), 355.

Kokten, O., & Raman, I. (2008). Differential effects of number of letters on word and nonword naming latency: revisited.

Koller, M. (2016). robustlmm: an R package for robust estimation of linear mixed-effects models. *Journal of statistical software*, *75*(1), 1-24.

Koşaner, Ö., Birant, Ç. C., & Aktaş, Ö. (2013). Improving Turkish language training

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Külekçi, M. O., & Oflazer, K. (2006). An infrastructure for Turkish prosody generation in text-to-speech synthesis.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: tests in linear mixed effects models. R package version 2.0-20. Vienna: R Foundation for Statistical Computing.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive psychology, 6(2), 293-323.

Lallier, M., Abu Mallouh, R., Mohammed, A. M., Khalifa, B., Perea, M., & Carreiras, M. (2018). Does the visual attention span play a role in reading in Arabic?. *Scientific Studies of Reading*, *22*(2), 181-190.

Lallier, M., Acha, J., & Carreiras, M. (2016). Cross-linguistic interactions influence reading development in bilinguals: a comparison between early balanced French-Basque and Spanish-Basque bilingual children. *Developmental science*, *19*(1), 76-89.

Lallier, M., Carreiras, M., Tainturier, M. J., Savill, N., & Thierry, G. (2013). Orthographic transparency modulates the grain size of orthographic processing: behavioral and ERP evidence from bilingualism. *Brain research*, *1505*, 47-60.

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*(3), 287-294.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. Applied psycholinguistics, 21(2), 243-262.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: an 8-year follow-up. *Journal of educational psychology*, *100*(1), 150.

Landerl, K., Freudenthaler, H.H., Heene, M., De Jong, P.F., Desrochers, A., Manolitsis, G., Parrila, R. and Georgiou, G.K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading*, *23*(3), 220-234.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J. and Kunze, S. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686-694.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. Cognition, 63(3), 315-334.

Lee, & O. J. L. Tzeng (Eds.), *The extraordinary brain series. Dyslexia across languages: Orthography and the brain–gene–behavior link* (p. 169–183). Paul H Brookes Publishing.

Lee, C. J. (2003). Evidence-based selection of word frequency lists. Journal of Speech-Language Pathology and Audiology, 27(3), 172-175.

Leech, G., Rayson, P., & Wilson, A. (2001). Word frequencies in written and spoken English: Based on the British National Corpus. London: Longman.

Leinonen, S., Müller, K., Leppänen, P. H. T., Aro, M., Ahonen, T. and Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading.. *Reading and Writing*, 14: 265–296.

Leong, C. K. (1999). Phonological and morphological processing in adult students with learning/reading disabilities. Journal of Learning Disabilities, 32(3), 224-238.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, No. 8, pp. 707-710).

Lewandowsky, S., & Li, S. C. (1995). Catastrophic interference in neural networks: Causes, solutions, and data. In *Interference and inhibition in cognition* (pp. 329-361). Academic Press.

Lewis, G. (1999). The Turkish Language Reform: A Catastrophic Success: A Catastrophic Success. OUP Oxford.

Leybaert, J., & Content, A. (1995). Reading and spelling acquisition in two different teaching methods: A test of the independence hypothesis. Reading and writing, 7(1), 65-88.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. DOI: https://doi. org/10.1016/j.jesp.2013.03.013

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. *Orthography, reading, and dyslexia*, 137-153.

Liceralde, V. R. T., & Gordon, P. C. (2018, October 11). Consequences of using power transforms as a statistical solution in linear mixed-effects models of chronometric data. Retrieved from osf.io/vyu65

Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *LREC*.;

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology*, *6*, 1171.

Lobier, M., Dubois, M., & Valdois, S. (2013). The role of visual processing speed in reading speed development. *PloS one*, *8*(4), e58097.

Logacev, Ö. Ü., Fuchs, S., & Żygis, M. (2014). Soft ‘g’in Turkish: Evidence for Sound Change in Progress. *Journal of the International Phonetic Association*.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(2), 229-247.

Logie, R. H. (1995). *Visuo-spatial working memory*. Psychology Press.

Lorusso, M. L., Cantiani, C., & Molteni, M. (2014). Age, dyslexia subtype and comorbidity modulate rapid auditory processing in developmental dyslexia. *Frontiers in human neuroscience*, *8*, 313.

Lukatela, G., Popadić, D., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, *8*(2), 124-132.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. Behavior research methods, 49(4), 1494-1502.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203-208.

Lyon, R., & Watson, B. (1981). Empirically derived subgroups of learning disabled readers: Diagnostic characteristics. *Journal of Learning Disabilities*, *14*(5), 256-261.

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A. M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly (1982-)*, 514-546.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. Psychological Review, 89(5), 483.

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). SUBTLEX-PL: subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, *47*(2), 471-483.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of reading*, *3*(2), 129-157.

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition*, *58*(2), 157-195.

Marinus, E., & de Jong, P. F. (2010). Size does not matter, frequency does: Sensitivity to orthographic neighbors in normal and dyslexic readers. *Journal of experimental child psychology*, *106*(2), 129-144.

Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. Reading research: Advances in theory and practice, 3, 199-221.

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech, Language, and Hearing Research*.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of psycholinguistic research*, *2*(3), 175-199.

Martelli, M., De Luca, M., Lami, L., Pizzoli, C., Pontillo, M., Spinelli, D., & Zoccolotti, P. (2014). Bridging the gap between different measures of the reading speed deficit in developmental dyslexia. *Experimental brain research*, *232*(1), 237-252.

Martens, V. E., & De Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading*, *31*(1), 40-54.

Martensen, H., Maris, E., & Dijkstra, T. (2000). When does inconsistency hurt? On the relation between phonological consistency effects and the reliability of sublexical units. *Memory & Cognition*, *28*(4), 648-656.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1107.

materials: Grapheme-to-phoneme conversion for adding phonemic transcription into dictionary entries and course books. *Procedia-Social and Behavioral Sciences*, *103*, 473-484.

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological review*, *86*(4), 287.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, *88*(5), 375.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation* (Vol. 24, pp. 109-165). Academic Press.

Meisinger, E. B., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of dyslexia*, *60*(1), 1-17.

Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children’s vocabulary development. *Psychological science*, *23*(10), 1092-1098.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffino, M. and Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. Neuropsychologia, 48(4), 863-872.

Menghini, D., Finzi, A., Carlesimo, G. A., & Vicari, S. (2011). Working memory impairment in children with developmental dyslexia: Is it just a phonological deficity?. *Developmental neuropsychology*, *36*(2), 199-213.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J. & Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *science*, *331*(6014), 176-182.

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P.H., Lohvansuu, K. and Tóth, D. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. Learning and Instruction, 29, 65-77.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*(4), 506-525.

Moore, D. M., Porter, M. A., Kohnen, S., & Castles, A. (2012). Detecting different types of reading difficulties: A comparison of tests. *Australasian Journal of Special Education, The*, *36*(2), 112.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously?. *Cognition*, *7*(4), 323-331.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. British Journal of Psychology, 91(2), 167-180.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review,

Müller, K., & Brady, S. (2001). Correlates of early reading performance in a transparent orthography. *Reading and Writing*, *14*(7-8), 757-799.

Mullis, I. V., Martin, M. O., Gonzalez, E. J., & International Association for the Evaluation of Educational Achievement. (2003). *PIRLS 2001 international report*. International Study Center.

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.

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1998). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of experimental child psychology*, *71*(1), 3-27.

Nag, S. (2017) Assessment of literacy and foundational learning in developing countries: Final Report. Health and Education Advice and Resource Team, xiii, 94p

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language universal theory of learning to read. *Scientific Studies of Reading*, *16*(5), 404-423.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English?. *Reading research quarterly*, 304-330.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313.

New, B. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic bulletin & review*, *13*(1), 45-52.

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, *51*(4), 568-585.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied psycholinguistics*, *28*(4), 661-677.

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.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516-524.

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S., & Tree, J. J. (2008). Computational modelling of phonological dyslexia: How does the DRC model fare?. *Cognitive Neuropsychology*, *25*(2), 165-193.

Nikolopoulos, D., Goulandris, N., & Snowling, M. J. (2003). Developmental dyslexia in Greek. Whurr Publishers.

Niolaki, G. Z., Terzopoulos, A. R., & Masterson, J. (2014). Varieties of developmental dyslexia in Greek children. *Writing Systems Research*, *6*(2), 230-256.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian

Norris, D. (2013). Models of visual word recognition. Trends in Cognitive Sciences, 17(10), 517-524. doi:http://dx.doi.org/10.1016/j.tics.2013.08.003

Nunnally, J. (1978). Psychometric methods. New York: McGraw-Hill

Odabaş, H., Odabaş, Y., & Polat, C. (2008). Üniversite öğrencilerinin okuma alışkanlığı: Ankara Üniversitesi örneği. *Bilgi Dünyası*, *9*(2), 431-465.

Oflazer, K., & Inkelas, S. (2006). The architecture and the implementation of a finite state pronunciation lexicon for Turkish. *Computer Speech & Language*, *20*(1), 80-106.

Oktay, A., & Aktan, E. (2002). A Cross-linguistic Comparison of Phonological Awareness and Word Recognition in Turkish and English Comparaison. International Journal of Early Years Education, 10(1), 37-48.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113.

Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual and developmental differences in reading disability. Reading research: advances in theory and practice, Vol. 4. - San Diego. S. 1-64

Öney, B. (1990). Context effect on word recognition in a writing system with regular letter-sound correspondences: Evidence from Turkish. Unpublished PhD thesis, University of California, Santa Barbara.

Öney, B., & Durgunoğlu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. Applied psycholinguistics, 18(1), 1-15.

Oney, B., & Goldman, S. R. (1984). Decoding and comprehension skills in Turkish and English: Effects of the regularity of grapheme-phoneme correspondences. Journal of Educational Psychology, 76(4), 557.

Oney, B., Peter, M., & Katz, L. (1997). Phonological processing in printed word recognition: Effects of age and writing system. Scientific Studies of Reading, 1(1), 65-83.

Özmen, R. G. (2005). Öğrenme güçlüğü olan öğrencilerin okuma hızlarının metinlerde karşılaştırılması. *Eğitim ve Bilim*, *30*(136).

Ozturk, S. A., Sevim, N., & Eroglu, E. (2006). Leisure book reading and purchasing: an analysis of Turkish consumers. *International journal of consumer studies*, *30*(4), 378-388.

Pacheco, A., Reis, A., Araújo, S., Inácio, F., Petersson, K. M., & Faísca, L. (2014). Dyslexia heterogeneity: cognitive profiling of Portuguese children with dyslexia. *Reading and Writing*, *27*(9), 1529-1545.

Pacht, J. M., & Rayner, K. (1993). The processing of homophonic homographs during reading: Evidence from eye movement studies. *Journal of Psycholinguistic Research*, *22*(2), 251-271.

Pagliuca, G., & Monaghan, P. (2009). A connectionist model of reading for Italian. In *Connectionist models of behaviour and cognition II* (pp. 301-311).

Pagliuca, G., Arduino, L. S., Barca, L., & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: The lexicality effect in Italian. Language and Cognitive Processes, 23(3), 422-433.

Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in psychology*, *7*, 1217.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, *113*, 54-71.

Park, H. R., & Uno, A. (2015). Cognitive Abilities Underlying Reading Accuracy, Fluency and Spelling Acquisition in Korean Hangul Learners from Grades 1 to 4: A Cross-Sectional Study. *Dyslexia*, *21*(3), 235-253.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development?. *Scientific studies of reading*, *8*(1), 3-26.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M. and Pesenti, S. (2000). A cultural effect on brain function. *Nature neuroscience*, *3*(1), 91-96.

Pehlivan, A., Serin, O., & Serin, N. B. (2010). Determining reading interests and habits of candidate teachers (TRNC Sample). *Procedia-Social and Behavioral Sciences*, *9*, 869-873.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413.

Pennington, B. F., & Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child development*, *72*(3), 816-833.

Pennington, B.F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D.R., Samuelsson, S., Byrne, B., Willcutt, E.G. and Olson, R.K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of abnormal psychology*, *121*(1), 212.

Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: A review. *Psicológica*, *21*(2).

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37-44.

Perfetti, C. A. (2003). The universal grammar of reading. Scientific Studies

of Reading, 7(1), 3-24.
Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In *Learning to read across languages* (pp. 25-50). Routledge.
Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. Language Learning and Development, 9(4), 296-316.
Perfetti, C. A., & Liu, Y. (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. Reading and Writing, 18(3), 193-210.
Perfetti, C. A., & Zhang, S. (1995). *The universal word identification reflex.* In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 33* (p. 159–189). Academic Press.
Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: some implications of research on Chinese for general theories of reading. *Psychological review*, *112*(1), 43.
Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a “universal” phonological principle. *Advances in psychology*, *94*, 227-248.
Perfetti, C., Cao, F., & Booth, J. (2013). Specialization and universals in the development of reading skill: How Chinese research informs a universal science of reading. *Scientific Studies of Reading*, *17*(1), 5-21.
Perin, D. (1983). Phonemic segmentation and spelling. *British journal of Psychology*, *74*(1), 129-144.
Perry, C., Ziegler, J. C., & Coltheart, M. (2002). A dissociation between orthographic awareness and spelling production. *Applied Psycholinguistics*, *23*(1), 43-73.
Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological review*, *114*(2), 273.
Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive psychology*, *61*(2), 106-151.
Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. Cognitive Science, 37(5), 800-828.
Perry, C., Ziegler, J. C., & Zorzi, M. (2014). CDP++. Italian: Modelling sublexical and supralexical inconsistency in a shallow orthography. PloS one, 9(4), e94291.
Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological Science*, *30*(3), 386-395.
Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *The lancet*, *379*(9830), 1997-2007.
Peterson, R. L., Arnett, A. B., Pennington, B. F., Byrne, B., Samuelsson, S., & Olson, R. K. (2018). Literacy acquisition influences children's rapid automatized naming. *Developmental science*, *21*(3), e12589.
Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: Testing the predictions of the dual-route and connectionist

frameworks. *Cognition*, *126*(1), 20-38.

Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. J. (2014). Longitudinal stability of phonological and surface subtypes of developmental dyslexia. *Scientific Studies of Reading*, *18*(5), 347-362.

Peynircioglu, Z. F., Durgunoglu, A. Y., & Oney-Küsefoglu, B. (2002). Phonological awareness and musical aptitude. Journal of Research in reading, 25(1), 68-80.

Peyrin, C., Démonet, J. F., N'Guyen-Morel, M. A., Le Bas, J. F., & Valdois, S. (2011). Superior parietal lobule dysfunction in a homogeneous group of dyslexic children with a visual attention span disorder. *Brain and language*, *118*(3), 128-138.

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-Plus, 3-56.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, *12*(5-6), 765-806.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56.

Pneuman, S. (2009). *Defining the early indicators of dyslexia: providing the signposts to intervention* (Doctoral dissertation, University of Bath).

Porpodas, C. (2006). Literacy acquisition in Greek: Research review of the role of phonological and cognitive factors. *Handbook of orthography and literacy*, 189-199.

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of greek. Journal of Learning Disabilities, 32(5), 406-416.

Powell, D., Plaut, D., & Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child?. *Journal of Research in Reading*, *29*(2), 229-250.

Pritchard, S. C. (2012). Incorporating learning mechanisms into the dual-route cascaded (DRC) model of reading aloud and word recognition. (Unpublished Doctoral Thesis, Macquarie University)

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2016). Modelling the implicit learning of phonological decoding from training on whole-word spellings and pronunciations. *Scientific studies of reading*, *20*(1), 49-63.

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, *42*(3), 722-770.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behavior research methods, 39(4), 859-862.

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior research methods*, *41*(4), 991-1008.

Provazza, S., Adams, A. M., Giofrè, D., & Roberts, D. J. (2019). Double trouble–visual and phonological impairments in english dyslexic readers. *Frontiers in psychology*, *10*, 2725.

Puolakanaho, A., Poikkeus, A-M., Ahonene, T., Tolvanen, A., & Lyytinen, H. (2004). Emerging phonological awareness differentiates children with and without familial risk for dyslexia after controlling for general language skills. *Annals of Dyslexia, 54*, 221-243.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rack, J. P. (1994) Dyslexia: The phonological deficit hypothesis. In A. Fawcett and R. Nicholson (Eds.) Dyslexia in children: multidisciplinary perspectives. London: Harvester Wheatsheaf, pp. 5-37.

Raman, I. (1999). Single word naming in a transparent orthography. *Unpublished Ph. D. thesis, Middlesex University, England*.

Raman, I. (2003). Lexicality Effects in Single-word Naming in Alphabetic Turkish. *Literacy Acquisition: The Role of Phonology, Morphology and Orthography*, *353*, 83.

Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? Visual Cognition, 13(7-8), 1044-1053. doi:10.1080/13506280500153200

Raman, I. (2011). The role of age of acquisition in picture and word naming in dyslexic adults. British Journal of Psychology, 102(3), 328-339.

Raman, I., & Weekes, B. S. (2005). Acquired dyslexia in a Turkish-English speaker. *Annals of Dyslexia*, *55*(1), 79-104.

Raman, I., Baluch, B., & Besner, D. (2004). On the control of visual word recognition: Changing routes versus changing deadlines. Memory & cognition, 32(3), 489-500.

Raman, I., Baluch, B., & Sneddon, P. (1996). What is the cognitive system's preferred route for deriving phonology from print?. European Psychologist, 1(3), 221-227.

Raman, I., Raman, E., & Mertan, B. (2014). A standardized set of 260 pictures for Turkish: Norms of name and image agreement, age of acquisition, visual complexity, and conceptual familiarity. Behavior Research Methods, 46(2), 588-595.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction?. *Current opinion in neurobiology*, *13*(2), 212-218.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841-865.

Ramus, F., White, S., & Frith, U. (2006). Weighing the evidence between competing theories of dyslexia. *Developmental Science*, *9*(3), 265-269.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, *5*(2), 277-282.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. Journal of Experimental Psychology: Human Perception and Performance, 28(2), 307.

Rastle, K., Havelka, J., Wydell, T. N., Coltheart, M., & Besner, D. (2009). The cross-script length effect: further evidence challenging PDP models of reading aloud. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 238.

Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological review*, *97*(2), 285.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological bulletin*, *114*(3), 510.

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92-105.

Raven, J. (1986). Court, JH, & raven, J.(1986). Raven's Progressive Matrices and Raven's Coloured Matrices,

Ray, N. J., Fowler, S., & Stein, J. F. (2005). Yellow filters can improve magnocellular function: motion sensitivity, convergence, accommodation, and reading. *Annals of the New York Academy of Sciences*, *1039*(1), 283-293.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. Psychological Science in the Public Interest, 2(2), 31-74.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). *Psychology of reading*. Psychology Press.

Rayson, P., Wilson, A., & Leech, G. (2001). Grammatical word class variation within the British National Corpus sampler. *Language and Computers*, *36*(1), 295-306.

Razon, N. (1976) A Special Reading Disorder (Dyslexia). İ.Ü.E.F.Press No: 2031, İstanbul, pg: 99-102

Razon, N. (1980) Okuma Bozuklukları ve Nedenleri. Pedagoji Dergisi Sayı I, Edebiyat Fakültesi Basımevi, İstanbul.

Razon, N. (1982). Okuma güçlükleri. *Eğitim ve Bilim*, *7*(39).

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. Science, 181(4099), 574-576.

Reid, G. (2001). Biological, cognitive and educational dimensions of dyslexia: current scientific thinking. *Dyslexia: Successful Inclusion in the Secondary School*, 10.

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*(3), B71-B80.

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.

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. Journal of Experimental Child Psychology,102,96–113.

Róna-Tas, A. (2015). Genetic Question. *The Turkic Languages*, 67.

Rønneberg, V., & Torrance, M. (2017). Cognitive predictors of shallow orthography spelling accuracy and time course. *Reading & Writing*.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. Toward a psychology of reading, 66.

Rudel, R. G., & Denckla, M. B. (1976). Relationship of IQ and reading score to visual, spatial, and temporal matching tasks. *Journal of Learning Disabilities*, *9*(3), 42-51.

Rumelhart, D. E., & Hinton, G. E. Williams, R. J. (1986). Learning Internal Representations by Error Propagation in Parallel Distributed Processing*, The MIT Press: Cambridge, Massachusetts*.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological review*, *81*(2), 99.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). Parallel distributed processing, vol. 1: Foundations. *Explorations in the Microstructure of Cognition*.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Sequential thought processes in PDP models. *Parallel distributed processing: explorations in the microstructures of cognition*, *2*, 3-57.

Şahin, M. (2014). *Itumorph-Türkçe İçin Daha Geniş Kapsamlı Ve Başarılı Bir Biçimbilimsel Çözümleyici* (Doctoral dissertation, Fen Bilimleri Enstitüsü).

Sahin, M., Sulubacak, U., & Eryigit, G. (2013). Redefinition of Turkish morphology using flag diacritics. In *Proceedings of The Tenth Symposium on Natural Language Processing (SNLP-2013), Phuket, Thailand, October*.

Sak, H., Güngör, T., & Saraçlar, M. (2007). Morphological disambiguation of Turkish text with perceptron algorithm. *Computational Linguistics and Intelligent Text Processing*, 107-118.

Salor, Ö., Pellom, B. L., Ciloglu, T., Hacioglu, K., & Demirekler, M. (2002). On developing new text and audio corpora and speech recognition tools for the turkish language. In *INTERSPEECH*.

Sambai, A., Uno, A., Kurokawa, S., Haruhara, N., Kaneko, M., Awaya, N., Kozuka, J., Goto, T., Tsutamori, E., Nakagawa, K. and Wydell, T.N.(2012). An investigation into kana reading development in normal and dyslexic Japanese children using length and lexicality effects. Brain and Development, 34(6), 520-528.

Sarıpınar, E. G., & Erden, G. (2010). Okuma güçlüğünde akademik beceri ve duyusal-motor işlevleri değerlendirme testlerinin kullanılabilirliği. *Türk Psikoloji Dergisi*, *25*(65), 56-66.

Savage, R., & Carless, S. (2005). Phoneme manipulation not onset-rime manipulation ability is a unique predictor of early reading. Journal of Child Psychology and Psychiatry, 46(12), 1297-1308.

Savasir, I., & Sahin, N. (1988). Wechsler Cocuk Zeka Olcegi (WISC-R)(Wechsler Intelligence Scale for Children). *Ankara, Turkey: Milli Egitim Basy mevi*.

Scerri, T. S., & Schulte-Körne, G. (2010). Genetics of developmental dyslexia. *European child & adolescent psychiatry*, *19*(3), 179-197.

Scharlipp, W. E. (1994). *Introduction to the Old Turkish runic inscriptions*. Kantzilaris Centre.

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: more evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658-672.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic bulletin & review*, *22*(6), 1614-1629.

Sebastián-Galles, N. (1991). Reading by analogy in a shallow orthography. *Journal of Experimental Psychology: Human perception and performance*, *17*(2), 471.

Sebastian-Gallés, N., & Vacchiano, A. P. (1995). The development of analogical reading in Spanish. Reading and Writing, 7(1), 23-38.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1-30.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy?. Coltheart, Max (Ed). (**1987**). Attention and performance 12: The psychology of reading , (pp. 245-263). Hillsdale, NJ, US

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological review, 96(4), 523.

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: data fitting versus theory building. *From inkmarks to ideas: current issues in lexical processing.* Hove, UK: Psychology Press

Serrano, F., & Defior, S. (2008). Dyslexia speed problems in a transparent orthography. *Annals of dyslexia*, *58*(1), 81.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143-174.

Sezer, E. (1981). On non-final stress in Turkish. *Journal of Turkish studies*, *5*, 61-69.

Sezer, T., Sezer, B. S. (2013). TS Corpus: Herkes için Türkçe derlem. In *Proceedings of the 27th National Linguistics Conference. Antalya. Hacettepe University, Linguistics Department* (pp. 217-225).

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain: a journal of neurology*, *104*(3), 413-429.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10-21.

Share, D. L. (1995). Phonological recoding and self-teaching:< I> sine qua non of reading acquisition. Cognition, 55(2), 151-218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of experimental child psychology*, *72*(2), 95-129.

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an" outlier" orthography. Psychological Bulletin, 134(4), 584.

Share, D. L. &. Daniels, P. T (2016): Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). Writing Systems Research 8.1: 17-31.

Shaywitz, S. E., & Shaywitz, B. A. (2003). Dyslexia (specific reading disability). Pediatrics in Review, 24(5), 147-153.

Sheriston, L., Critten, S., & Jones, E. (2016). Routes to Reading and Spelling: Testing the Predictions of Dual-Route Theory. *Reading Research Quarterly*, *51*(4), 403-417.

Shimron, J. (2006). *Reading Hebrew: The language and the psychology of reading it*. Routledge.

Sidekli, S. (2010). İlköğretim 5. sinif öğrencilerinin okuma ve anlama becerilerini geliştirme (Eylem araştirmasi). *Unpublished doctoral dissertation). Gazi University, Institute of Educational Sciences, Ankara*.

Siegel, L. S. (1990). IQ and learning disabilities. Learning Disabilities: Theoretical and Research Issues, , 111.

Siegel, L. S., Swanson, H. L., & Keogh, B. (1990). IQ and learning disabilities. *Learning disabilities: Theoretical and research issues*, 23-39.

Siegel, L., & Hurford, D. (2019). The case against discrepancy models in the evaluation of dyslexia. *Perspectives of Language and Literacy*, *45*(1), 23-28.

Sinclair, J. (2005). Corpus and text-basic principles. *Developing linguistic corpora: A guide to good practice*, 1-16.

Snowling, M. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review*, *3*(1), 4-11.

Snowling, M. J. (1983). The comparison of acquired and developmental disorders of reading: A discussion. *Cognition, 14*(1), 105–118.

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. Journal of Research in Reading, 18(2), 132-138.

Snowling, M. J. (2000). *Dyslexia*. Blackwell publishing.

Snowling, M. J. (2001). From language to reading and dyslexia1. Dyslexia, 7(1), 37-46.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, *81*(1), 1-23.

Snowling, M. J., & Hulme, C. (2012). Annual Research Review: The nature and classification of reading disorders–a commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry*, *53*(5), 593-607.

Snowling, M. J., Bryant, P. E., & Hulme, C. (1996). Theoretical and methodological pitfalls in making comparisons between developmental and acquired dyslexia: Some comments on A. Castles & M. Coltheart (1993). *Reading and Writing*, *8*(5), 443-451.

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., Comesaña, M. & Perea, M. (2015). On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *The Quarterly Journal of Experimental Psychology*, *68*(4), 680-696.

Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of experimental psychology*, *43*(3), 195.

Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How well do phonological awareness and rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-analysis. *Scientific Studies of Reading*, *20*(2), 99-123.

Soriano, M., & Miranda, A. (2010). Developmental dyslexia in a transparent orthography: A study of Spanish dyslexic children. In *Literacy and Learning*. Emerald Group Publishing Limited.

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216.

Spearman, C. (1910). Correlation calculated from faulty data. *British journal of psychology*, *3*(3), 271.

Spencer, K. (2007). Predicting children's word-spelling difficulty for common

English words from measures of orthographic transparency, phonemic and graphemic length and word frequency. *British Journal of Psychology*, *98*(2), 305-338.

Sperling, A. J., Lu, Z. L., Manis, F. R., & Seidenberg, M. S. (2003). Selective magnocellular deficits in dyslexia: a “phantom contour” study. *Neuropsychologia*, *41*(10), 1422-1429.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*(6), 411-416.

Sprenger-Charolles, L., & Siegel, L. S. (1997). A longitudinal study of the effects of syllabic structure on the development of reading and spelling skills in French.

Sprenger-Charolles, L., Colé, P., Lacert, P., & Serniclaes, W. (2000). On subtypes of developmental dyslexia: evidence from processing time and accuracy scores. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*(2), 87.

Sprenger-Charolles, L., Siegel, L. S., & Bonnet, P. (1998). Reading and spelling acquisition in French: The role of phonological mediation and orthographic factors. *Journal of experimental child psychology*, *68*(2), 134-165.

Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling: A four-year longitudinal study. Journal of experimental child psychology, 84(3), 194-217.

Sprenger-Charolles, L., Siegel, L. S., Jimenez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading*, *15*(6), 498-521.

Stanovich, K. E. (1988). *Children's reading and the development of phonological awareness*. Wayne State University Press.

Stanovich, K. E. (1991). Word recognition: Changing perspectives.

Stanovich, K. E. (1996). Toward a more inclusive definition of dyslexia. Dyslexia, 2(3), 154-166.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of educational psychology*, *86*(1), 24.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of experimental child psychology*, *38*(2), 175-190.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of educational psychology*, *89*(1), 114.

Stein, J. (1989). Visuospatial perception and reading problems. The Irish Journal of Psychology, 10(4), 521-533.

Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, *7*(1), 12-36.

Stoodley, C. J., Hill, P. R., Stein, J. F., & Bishop, D. V. (2006). Auditory event-related potentials differ in dyslexics even when auditory psychophysical performance is normal. *Brain Research*, *1121*(1), 190-199.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition, 30*(2), 139-181.

Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2014). *Orthographic learning in Spanish children. Journal of Research in Reading, 39(3), 292–311.*

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological review*, *88*(2), 135.

Swanson, H. L., & Alexander, J. E. (1997). Cognitive processes as predictors of word recognition and reading comprehension in learning-disabled and skilled readers: Revisiting the specificity hypothesis. *Journal of Educational Psychology*, *89*(1), 128.

Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology*, *93*(4), 720.

Swanson, H. L., Cooney, J. B., & McNamara, J. K. (2004). Learning disabilities and memory. In *Learning about learning disabilities* (pp. 41-92). Academic Press.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research*, *73*(4), 407-440.

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.

Sze, W. P., Liow, S. J. R., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*(1), 263-273.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). Using multivariate statistics (Vol. 5). Boston, MA: Pearson.

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition*, *20*(3), 303-313.

Taft, M. (1979). Lexical access-via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 21-39.

Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics*, *5*(2), 167-169.

Tallal, P. E., Galaburda, A. M., Llinás, R. R., & von Euler, C. E. (1993). *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia*. New York Academy of Sciences.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. Journal of Educational Psychology, 89(2), 276.

TDK Yazim Kilavuzu, (2013) Türk Dil Kurumu Yayınları.

Temple, E., Poldrack, R.A., Protopapas, A.A., Nagarajan, S., Salz, T., Tallal, P., Merzenich, M.M. and Gabrieli, J.D.E. (2000). Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proceedings of the National Academy of Sciences*, *97*(25), 13907-13912.

Thorndike, E. L. (1921). The teacher’s word book.

Thorndike, E. L., & Lorge, I. (1944). The teacher's wordbook of 30,000 words. New York: Columbia University, Teachers College.

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.

Tilanus, E. A., Segers, E., & Verhoeven, L. (2013). Diagnostic profiles of children with developmental dyslexia in a transparent orthography. *Research in Developmental Disabilities*, *34*(11), 4194-4202.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*.

Tobia, V., & Marzocchi, G. M. (2014). Predictors of reading fluency in Italian orthography: Evidence from a cross-sectional study of primary school students. *Child neuropsychology*, *20*(4), 449-469.

TopbaŞ, S., Kaçar-Kütükçü, D., & Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical linguistics & phonetics*, *28*(7-8), 602-616.

Torgesen, J. K., Rashotte, C. A., & Alexander, A. N. N. E. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. *Dyslexia, fluency, and the brain*, 333-355.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific studies of reading*, *1*(2), 161-185.

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.

Torunoglu, D., & Eryigit, G. (2014). A cascaded approach for social media text normalization of Turkish. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL* (pp. 62-70).

Treiman, R., Fowler, C. A., Gross, J., Berch, D., & Weatherston, S. (1995). Syllable structure or word structure? Evidence for onset and rime units with disyllabic and trisyllabic stimuli. *Journal of Memory and Language*, *34*(1), 132-155.

Tressoldi, P. E., Lorusso, M. L., Brenbati, F., & Donini, R. (2008). Fluency remediation in dyslexic children: does age make a difference?. *Dyslexia*, *14*(2), 142-152.

Tressoldi, P. E., Stella, G., & Faggella, M. (2001). The development of reading speed in Italians with dyslexia: A longitudinal study. *Journal of learning disabilities*, *34*(5), 414-417.

Trick, L. M., Mutreja, R., & Hunt, K. (2012). Spatial and visuospatial working memory tests predict performance in classic multiple-object tracking in young adults, but nonspatial measures of the executive do not. Attention, Perception, & Psychophysics, 74(2), 300-311.

Ugurel-Semin, R. (1987) In: Ugurel-Semin R, editor. The adaptation of Stanford-Binet Intelligence Scale in children who live in İstanbul. Istanbul: The printing house of The Istanbul, 1987.

Ulicheva, A. (2015). The influence of phonotactics on oral reading in Russian. *HKU Theses Online (HKUTO)*.

Ulicheva, A., Coltheart, M., Saunders, S., & Perry, C. (2016). Phonotactic constraints: Implications for models of oral reading in Russian. *Journal of experimental psychology: Learning, memory, and cognition*, *42*(4), 636.

Uno, A., Wydell, T. N., Haruhara, N., Kaneko, M., & Shinya, N. (2009). Relationship between reading/writing skills and cognitive abilities among Japanese primary-school children: normal readers versus poor readers (dyslexics). Reading and Writing, 22(7), 755-789.

Vaessen, A. and Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological science*, *21*(4), 551-559.

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. Journal of Educational Psychology, 102(4), 827.

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia*, *10*(4), 339-363.

Valdois, S., Bosse, M. L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing*, *16*(6), 541-572.

Valdois, S., Peyrin, C., Lassus-Sangosse, D., Lallier, M., Demonet, J. F., & Kandel, S. (2014). Dyslexia in a French–Spanish bilingual girl: behavioural and neural modulations following a visual attention span intervention. *Cortex*, *53*, 120-145.

Valle-Arroyo, F. (1989). Reading errors in Spanish. In *Reading and writing disorders in different orthographic systems* (pp. 163-175). Springer, Dordrecht.

Van Daal, V., & Van der Leij, A. (1999). Developmental dyslexia: Related to specific or general deficits?. *Annals of dyslexia*, *49*(1), 71-104.

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the length effect: Specifying the relation with visual and phonological correlates of reading. *Scientific Studies of Reading*, *17*(4), 243-256.

Van den Boer, M., van Bergen, E., & de Jong, P. F. (2014). Underlying skills of oral and silent reading. *Journal of experimental child psychology*, *128*, 138-151.

Van Den Boer, M., Van Bergen, E., & de Jong, P. F. (2015). The specific relation of visual attention span with reading and spelling in Dutch. *Learning and Individual Differences*, *39*, 141-149.

Van den Bosch, A., Content, A., Daelemans, W., & De Gelder, B. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*(3), 178-188.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190.

Van Orden, G. C., & Kloos, H. (2005). The question of phonology and reading. *The science of reading: A handbook*, 61-78.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. Psychological Review, 97(4), 488.

Vander Beke, G. E. (Ed.). (1932). *French word book* (Vol. 15). Macmillan.

Vanlı, L. (1988). Dislektik çocuklarda algısal bozukluklar ve okuma yazma hataları. *Psikoloji Dergisi*, *6*(22), 36-40.

Velan, H., Frost, R., Deutsch, A., & Plaut, D. C. (2005). The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes*, *20*(1-2), 169-206.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? Journal of Child Psychology and Psychiatry, 45(1), 2-40.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*(4), 601.

Venezky, R. (2006). Foundations for studying basic processes in reading. Handbook of Orthography and Literacy, , 735-758.

Verhagen, W., Aarnoutse, C., & Van Leeuwe, J. (2008). Phonological awareness and naming speed in the prediction of Dutch children's word recognition. *Scientific Studies of Reading*, *12*(4), 301-324.

Verhoeven, L., & Keuning, J. (2018). The nature of developmental dyslexia in a transparent orthography. Scientific Studies of Reading, 22(1), 7-23.

Von Karolyi, C., Winner, E., Gray, W., & Sherman, G. F. (2003). Dyslexia linked to talent: Global visual-spatial ability. *Brain and language*, *85*(3), 427-431.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wagner, R. K., & Muse, A. (2012). Short-term memory deficits in developmental dyslexia. In Working Memory and Neurodevelopmental Disorders (pp. 55-72). Psychology Press.

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.

Wang, J. J., Bi, H. Y., Gao, L. Q., & Wydell, T. N. (2010). The visual magnocellular pathway in Chinese-speaking children with developmental dyslexia. *Neuropsychologia*, *48*(12), 3627-3633.

Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children’s and adults’ use of spelling-sound information in three reading tasks. Memory & Cognition, 12(3), 293-305.

Weber, P., Kozel, N., Purgstaller, C., Kargl, R., Schwab, D., & Fink, A. (2013). First and second language in the brain: Neuronal correlates of language processing and spelling strategies. Brain and Language, 124(1), 22-33.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. The Quarterly Journal of Experimental Psychology Section A, 50(2), 439-456.

Wells, J. C. (1995). Computer-coding the IPA: a proposed extension of SAMPA. *Revised draft*, *4*(28), 1995.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550-1560.

Whiting, C., Shtyrov, Y., & Marslen-Wilson, W. (2014). Real-time functional architecture of visual word recognition. *Journal of Cognitive Neuroscience*, *27*(2), 246-265.

Widrow, B., & Hoff, M. E. (1960). Adaptive Switching Circuits. Institute of

Radio Engineers, Western Electronic Show and Convention Record, Part 4 (pp. 96-104). New York:

Wilkins, A. (2003). Reading through colour: How coloured filters can reduce reading difficulty, eye strain, and headaches.

Willems, G., Jansma, B., Blomert, L., & Vaessen, A. (2016). Cognitive and familial risk evidence converged: A data-driven identification of distinct and homogeneous subtypes within the heterogeneous sample of reading disabled children. *Research in developmental disabilities*, *53*, 213-231.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied psycholinguistics*, *14*(1), 1-33.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*(1), 91-103.

Wimmer, H., & Hummer, P. (1990). How German-speaking first graders read and spell: Doubts on the importance of the logographic stage. Applied Psycholinguistics, 11(4), 349-368.

Wimmer, H., & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. *Journal of educational Psychology*, *94*(2), 272.

Wimmer, H., & Schurz, M. (2010). Dyslexia in regular orthographies: manifestation and causation. Dyslexia, 16(4), 283-299.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of educational psychology*, *92*(4), 668.

Winner, E., von Karolyi, C., Malinsky, D., French, L., Seliger, C., Ross, E., & Weber, C. (2001). Dyslexia and visual-spatial talents: Compensation vs deficit model. *Brain and language*, *76*(2), 81-110.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. Journal of Educational Psychology, 91, 415–438.

Wolf, M., & Bowers, P. G. (2000). Naming-speed processes and developmental reading disabilities: An introduction to the special issue on the double-deficit hypothesis. *Journal of learning disabilities*, *33*(4), 322-324.

Wolff, U. (2014). RAN as a predictor of reading skills, and vice versa: Results from a randomised reading intervention. Annals of Dyslexia, 64(2), 151-165.

World Health Organization The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva: World Health Organization, 1993

Wybrow, D. P. (2014). *Subtypes of developmental dyslexia* (Doctoral dissertation, University of Essex).

Wydell, T. N. (1991). Processing in the Reading of Japanese: Comparative Studies between English and Japanese Orthographies. Unpublished PhD thesis, University of London.

Wydell, T. N. (2012). Cross-cultural/linguistic differences in the prevalence of developmental dyslexia and the hypothesis of granularity and transparency. *Dyslexia—A comprehensive and international approach*, 1-14.

Wydell, T. N., & Butterworth, B. (1999). A case study of an English-Japanese bilingual with monolingual dyslexia. *Cognition*, *70*(3), 273-305..

Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for Kanji: Is a rows still a rose in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 491.

Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience*, *15*(7), 1052-1062.

Yang, J., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, *61*(2), 238-257.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502-529.

Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*(4), 992-1003.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971-979.

Yildirim, K., Rasinski, T., Ates, S., Fitzgerald, S., Zimmerman, B., & Yildiz, M. (2014). The relationship between reading fluency and vocabulary in fifth grade Turkish students. Literacy Research and Instruction, 53(1), 72-89.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. Reading Research Quarterly.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). The educator's word frequency guide. New York: Touchstone Applied Science Associates. *Inc. My Book*.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. Journal of Memory and language, 47(1), 1-29.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*(2), 145-160.

Ziegler, J. C. (2011). *Understanding developmental dyslexia through computational modeling: An individual deficit-based simulation approach.* In P. McCardle, B. Miller, J. R.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. Psychological bulletin, 131(1), 3.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*(1), 151-178.

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 504-515.

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*(3), 413-430.

Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin & Review*, *10*(4), 947-953.

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120397.

Ziegler, J. C., Perry, C., & Zorzi, M. (2019). 16 Modeling the Variability of Developmental Dyslexia. *Developmental Dyslexia across Languages and Writing Systems*, 350.

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379-384.

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal?. Journal of experimental child psychology, 86(3), 169-193.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for-ough and the spelling for/u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods*, *29*(4), 600-618.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H.,

Zipf, G. (1950). The principle of least effort. *Reading, MA7 Addison Wesley*.

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.

Zoccolotti, P., De Luca, M., Di Pace, E., Judica, A., Orlandi, M., & Spinelli, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with high grapheme–phoneme correspondence. *Applied Psycholinguistics*, *20*(2), 191-216.

Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology*, *22*(5), 836-860.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(4), 1131.

Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New insights on developmental dyslexia subtypes: heterogeneity of mixed reading profiles. *PloS one*, *9*(6), e99337.

APPENDICES

Appendix 1: Modified version of a syllabification algorithm for Turkish based on Altınok (2016).

```
#! /usr/bin/python
# -*- coding: utf-8 -*-

import sys
import codecs

def lowercase(ch):
    return {
    'İ':u'i',
    'I':u'ı',
    'Ç':u'ç',
    'Ğ':u'ğ',
    'Ş':u'ş'
    }.get(ch, ch.lower())

def sesli(ch):
    ch = lowercase(ch)
    if ch in [u'a', u'e', u'i', u'ı', u'o', u'ö', u'u', u'ü']:
        return True
    else:
        return False

def hecele(str):
    index=0
    length=len(str)
    while sesli(str[index]) == False and length>index+1:
        index=index+1
    try:
        if sesli(str[index+1]):
            print str[0:index+1],
            hecele(str[index+1:])
        elif length>index+2:
            if sesli(str[index+2]):
                print str[0:index+1],
                hecele(str[index+1:])
            elif length>index+3:
                if sesli(str[index+3]):
                    print str[0:index+2],
                    hecele(str[index+2:])
                else:
                    if str[index+1:index+4] in [u'str', u'ktr', u'mtr', u'nsp']:
                        #print "istisna!.."
                        print str[0:index+2],
                        hecele(str[index+2:])
                    else:
                        #print "üç sessiz, normal kural"
                        print str[0:index+3],
                        hecele(str[index+3:])
            else:
                print unicode(str),
        else:
            print unicode(str),
```

```
    except:
        print unicode(str),
        return

f = codecs.open("subWordFinal2020.txt", encoding= 'utf-8')
for line in f:
    hecele(line)
    print
```

APPENDIX 1: ETHICAL APPROVAL LETTER FOR LEXICAL DECISION STUDY

College of Health and Life Sciences Research Ethics Committee (DLS)
Brunel University London
Kingston Lane
Uxbridge
UB8 3PH
United Kingdom

www.brunel.ac.uk

26 February 2018

LETTER OF APPROVAL

Applicant: Mr Evren Raman

Project Title: SUBTLEX_TR

Reference: 6916-LR-Feb/2018- 11760-2

Dear Mr Evren Raman

The Research Ethics Committee has considered the above application recently submitted by you.

The Chair, acting under delegated authority has agreed that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that the conditions of approval set out below are followed:

- The agreed protocol must be followed. Any changes to the protocol will require prior approval from the Committee by way of an application for an amendment.

Please note that:

- Research Participant Information Sheets and (where relevant) flyers, posters, and consent forms should include a clear statement that research ethics approval has been obtained from the relevant Research Ethics Committee.
- The Research Participant Information Sheets should include a clear statement that queries should be directed, in the first instance, to the Supervisor (where relevant), or the researcher. Complaints, on the other hand, should be directed, in the first instance, to the Chair of the relevant Research Ethics Committee.
- Approval to proceed with the study is granted subject to receipt by the Committee of satisfactory responses to any conditions that may appear above, in addition to any subsequent changes to the protocol.
- The Research Ethics Committee reserves the right to sample and review documentation, including raw data, relevant to the study.
- You may not undertake any research activity if you are not a registered student of Brunel University or if you cease to become registered, including abeyance or temporary withdrawal. As a deregistered student you would not be insured to undertake research activity. Research activity includes the recruitment of participants, undertaking consent procedures and collection of data. Breach of this requirement constitutes research misconduct and is a disciplinary offence.

Professor Christina Victor

Chair

College of Health and Life Sciences Research Ethics Committee (DLS)
Brunel University London

APPENDIX 2: ETHICAL APPROVAL LETTER FOR CHILDREN'S PSYCHOLINGUISTIC STUDY

Eastern Mediterranean University

The Department of Psychology
Eastern Mediterranean University
Research & Ethics Committee
Cigir Kalfaoglu

Famagusta, Turkish Republic of Northern Cyprus
Tel: +(90) 392 630 1389
Fax: +(90) 392 630 2475
e-mail: cigir.kalfaoglu@emu.edu.tr

Web: http://brahms.emu.edu.tr/psychology

Ref Code: 14/03-73

Date: 04.12.2015

Dear Evren Raman,

Thank you for contacting us with regards to an extension to your study entitled *Language universality vs. Specificity of reading processes: Evidence from Turkish-speaking monolingual children* which received approval from the Research & Ethics Committee on 04.03.2014.

Your extension has been *approved* for another year. Once again, if any changes to the study described in the original application or supporting documentation is necessary, please notify the committee as you may be required to make a resubmission of the application.

Good luck with the research.

Yours sincerely,

Assoc. Prof. Dr. Shenel Husnu Raman
On Behalf of the Research & Ethics Committee
Psychology Department
Eastern Mediterranean University

Chair of the University Research Ethics Committee
Dr Derek Healy

Brunel University, Uxbridge
Middlesex, UB8 3PH, UK
Telephone +44 (0)1895 274000
Fax +44 (0) 1895 232806
Web www.brunel.ac.uk/

Applicant: Evren Raman

Study title: Language universality vs. specificity of reading processes: Evidence from Turkish-speaking monolingual children

1 September 2020

The University Research Ethics Committee (UREC) has been informed that a number of records relating to research ethics approval for the above-named study have been lost, by both the applicant and the University, while changing over to a new system.

The UREC has not had sight of relevant approval letters which evidence that approval was in place from a Brunel REC; however, the UREC has been assured by the academic supervisor that approval was issued prior to commencement of the study. The UREC has noted that submission of an application to undertake the study to the relevant REC has been evidenced, and also that approval was in place from the host institution (Eastern Mediterranean University) before the study was undertaken. Records of formal written consent have also been retained by the student.

Although it has not been possible to retrieve records of approval, it is considered that assurance by the academic supervisor should be taken into account along with the evidence supplied, namely records of appropriate ethical conduct in the form of consent documentation and submitted application forms.

Although it is the student's responsibility to retain records of approval, the University acknowledges that it is also at fault for its failure to retain records pertaining to this study.

Sincerely,

Dr Derek Healy
Chair, University Research Ethics Committee

APPENDIX 3: ETHICAL APPROVAL LETTER FOR CHILDREN'S PSYCHOLINGUISTIC PILOT STUDY

Brunel University London

Chair of the University Research Ethics Committee
Dr Derek Healy

Brunel University, Uxbridge
Middlesex, UB8 3PH, UK
Telephone +44 (0)1895 274000
Fax +44 (0) 1895 232806
Web www.brunel.ac.uk/

Applicant: Evren Raman

Study title: Language universality vs. specificity of reading processes: Evidence from Turkish-speaking monolingual children

1 September 2020

The University Research Ethics Committee (UREC) has been informed that a number of records relating to research ethics approval for the above-named study have been lost, by both the applicant and the University, while changing over to a new system.

The UREC has not had sight of relevant approval letters which evidence that approval was in place from a Brunel REC; however, the UREC has been assured by the academic supervisor that approval was issued prior to commencement of the study. The UREC has noted that submission of an application to undertake the study to the relevant REC has been evidenced, and also that approval was in place from the host institution (Eastern Mediterranean University) before the study was undertaken. Records of formal written consent have also been retained by the student.

Although it has not been possible to retrieve records of approval, it is considered that assurance by the academic supervisor should be taken into account along with the evidence supplied, namely records of appropriate ethical conduct in the form of consent documentation and submitted application forms.

Although it is the student's responsibility to retain records of approval, the University acknowledges that it is also at fault for its failure to retain records pertaining to this study.

Sincerely,

Dr Derek Healy
Chair, University Research Ethics Committee

e cs [e]l 8 u 1
m cs e[r] 3 u 1
A cs [l]a 1 u 1
b cs g[a] 4 u 1
b cs e[G] e u 1
e cs [i]k 7 u 1
b cs k[e] 7 u 1
m cs n[g] 9 u 1
m cs l[e] 8 u 1
e cs [o]v w u 1
m cs [e]k 7 u 1
m cs v[u] w u 1
m cs g[l] 4 u 1
A sing b b u 1
A sing i i u 1
e sing r @ u 1
A sing O O u 1
A sing l l u 1
A sing d d u 1
A sing u u u 1
m sing e e u 1
A sing n n u 1
A sing a a u 1
A sing o o u 1
A sing v v u 1
b sing g g u 1
b sing e 3 u 1
A sing t t u 1
A sing C 2 u 1
A sing y y u 1
A sing p p u 1
A sing k k u 1
A sing m m u 1
A sing c c u 1
A sing S S u 1
e sing e e u 1
A sing s s u 1
m sing r r u 1
A sing h h u 1
A sing I 6 u 1
A sing U x u 1
e sing z Z u 1
b sing z z u 1
m sing z z u 1
A sing G 5 u 1
A sing f f u 1
b sing r R u 1
m sing g g u 1
A sing j j u 1
e sing g 4 u 1

APPENDIX 5: AVERAGED DRC MODEL PARAMETERS

General Parameters
ActivationRate 0.2
FrequencyScale 0.05
MinReadingPhonology 0.9

Feature Level Parameters
FeatureLetterExcitation 0.005
FeatureLetterInhibition 0.15

Letter Level Parameters
LetterOrthlexExcitation 0.03
LetterOrthlexInhibition 0.90
LetterLateralInhibition 0

Orthographic Lexicon (Orthlex) Parameters
OrthlexPhonlexExcitation 0.3
OrthlexPhonlexInhibition 0
OrthlexLetterExcitation 0.2
OrthlexLetterInhibition 0
OrthlexLateralInhibition 0.10

Phonological Lexicon (Phonlex) Parameters
PhonlexPhonemeExcitation 0.09
PhonlexPhonemeInhibition 0
PhonlexOrthlexExcitation 0.25
PhonlexOrthlexInhibition 0
PhonlexLateralInhibition 0.07

Phoneme Level Parameters
PhonemePhonlexExcitation 0.02
PhonemePhonlexInhibition 0.16
PhonemeLateralInhibition 0.147
PhonemeUnsupportedDecay 0.05

GPC Route Parameters
GPCPhonemeExcitation 0.045
GPCCriticalPhonology 0.25
GPCOnset 7

APPENDIX 6: GRADE 2 DRC MODEL PARAMETERS

General Parameters
ActivationRate 0.2
FrequencyScale 0.10
MinReadingPhonology 0.4

Feature Level Parameters
FeatureLetterExcitation 0.005
FeatureLetterInhibition 0.15

Letter Level Parameters
LetterOrthlexExcitation 0.03
LetterOrthlexInhibition 0.90
LetterLateralInhibition 0

Orthographic Lexicon (Orthlex) Parameters
OrthlexPhonlexExcitation 0.2
OrthlexPhonlexInhibition 0
OrthlexLetterExcitation 0.2
OrthlexLetterInhibition 0
OrthlexLateralInhibition 0.10

Phonological Lexicon (Phonlex) Parameters
PhonlexPhonemeExcitation 0.09
PhonlexPhonemeInhibition 0
PhonlexOrthlexExcitation 0.25
PhonlexOrthlexInhibition 0
PhonlexLateralInhibition 0.07

Phoneme Level Parameters
PhonemePhonlexExcitation 0.01
PhonemePhonlexInhibition 0.16
PhonemeLateralInhibition 0.147
PhonemeUnsupportedDecay 0.05

GPC Route Parameters
GPCPhonemeExcitation 0.050
GPCCriticalPhonology 0.24
GPCOnset 7

APPENDIX 7: GRADE 3 DRC MODEL PARAMETERS

General Parameters
ActivationRate 0.2
FrequencyScale 0.10
MinReadingPhonology 0.4

Feature Level Parameters
FeatureLetterExcitation 0.005
FeatureLetterInhibition 0.15

Letter Level Parameters
LetterOrthlexExcitation 0.03
LetterOrthlexInhibition 0.90
LetterLateralInhibition 0

Orthographic Lexicon (Orthlex) Parameters
OrthlexPhonlexExcitation 0.2
OrthlexPhonlexInhibition 0
OrthlexLetterExcitation 0.2
OrthlexLetterInhibition 0
OrthlexLateralInhibition 0.10

Phonological Lexicon (Phonlex) Parameters
PhonlexPhonemeExcitation 0.09
PhonlexPhonemeInhibition 0
PhonlexOrthlexExcitation 0.25
PhonlexOrthlexInhibition 0
PhonlexLateralInhibition 0.07

Phoneme Level Parameters
PhonemePhonlexExcitation 0.01
PhonemePhonlexInhibition 0.16
PhonemeLateralInhibition 0.147
PhonemeUnsupportedDecay 0.05

GPC Route Parameters
GPCPhonemeExcitation 0.050
GPCCriticalPhonology 0.24
GPCOnset 7

APPENDIX 8: GRADE 4 DRC MODEL PARAMETERS

```
# General Parameters
ActivationRate 0.2
FrequencyScale 0.00
MinReadingPhonology 0.4

# Feature Level Parameters
FeatureLetterExcitation 0.005
FeatureLetterInhibition 0.15

# Letter Level Parameters
LetterOrthlexExcitation 0.03
LetterOrthlexInhibition 0.90
LetterLateralInhibition 0

# Orthographic Lexicon (Orthlex) Parameters
OrthlexPhonlexExcitation 0.2
OrthlexPhonlexInhibition 0
OrthlexLetterExcitation 0.2
OrthlexLetterInhibition 0
OrthlexLateralInhibition 0.10

# Phonological Lexicon (Phonlex) Parameters
PhonlexPhonemeExcitation 0.09
PhonlexPhonemeInhibition 0
PhonlexOrthlexExcitation 0.25
PhonlexOrthlexInhibition 0
PhonlexLateralInhibition 0.07

# Phoneme Level Parameters
PhonemePhonlexExcitation 0.01
PhonemePhonlexInhibition 0.15
PhonemeLateralInhibition 0.15
PhonemeUnsupportedDecay 0.05

# GPC Route Parameters
GPCPhonemeExcitation 0.050
GPCCriticalPhonology 0.30
GPCOnset 10
```

APPENDIX 9: GRADE 5 DRC MODEL PARAMETERS

```
# General Parameters
ActivationRate 0.2
FrequencyScale 0.10
MinReadingPhonology 0.4

# Feature Level Parameters
FeatureLetterExcitation 0.005
FeatureLetterInhibition 0.15

# Letter Level Parameters
LetterOrthlexExcitation 0.03
LetterOrthlexInhibition 0.90
LetterLateralInhibition 0

# Orthographic Lexicon (Orthlex) Parameters
OrthlexPhonlexExcitation 0.3
OrthlexPhonlexInhibition 0
OrthlexLetterExcitation 0.2
OrthlexLetterInhibition 0
OrthlexLateralInhibition 0.10

# Phonological Lexicon (Phonlex) Parameters
PhonlexPhonemeExcitation 0.09
PhonlexPhonemeInhibition 0
PhonlexOrthlexExcitation 0.25
PhonlexOrthlexInhibition 0
PhonlexLateralInhibition 0.10

# Phoneme Level Parameters
PhonemePhonlexExcitation 0.01
PhonemePhonlexInhibition 0.01
PhonemeLateralInhibition 0.100
PhonemeUnsupportedDecay 0.05

# GPC Route Parameters
GPCPhonemeExcitation 0.05
GPCCriticalPhonology 0.19
GPCOnset 20
```