Clinical Markers of Developmental Language Disorder in Arabic

Juhayna Taha

School of Psychology and Clinical Language Sciences

University of Reading

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

September 2021

اهداء

إلى أمي و أبي

لدعمهم الغير متناهي و تشجيعهم لي لتحقيق طموحاتي

Dedication

To my parents,

Who have endlessly supported me and encouraged me to go after my dreams

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Juhayna Taha

Table of Contents

Publications

Chapter 2

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Expressive verb morphology deficits in Arabic-speaking children with Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *64*(2), 561-578. doi.org/10.1044/2020_JSLHR-19-00292

Chapter 3

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Nonword repetition performance of Arabic-speaking children with and without Developmental Language Disorder: a study on diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *64*(7), 2750-2765. https://doi.org/10.1044/2021_JSLHR-20-00556

Chapter 4

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Sentence repetition as a clinical marker of Developmental Language Disorder: Evidence from Arabic. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4876-4899.https://doi.org/10.1044/2021_JSLHR-21-00244

For all manuscripts, authors' contributions were as follows:

- J.T, V.S and E.P conceived and designed the studies.
- V.S and E.P co-supervised the studies.
- J.T designed and piloted the tasks, recruited and tested participants, transcribed and scored the data, conducted all statistical analyses, wrote all the manuscripts and revised the manuscripts according to the reviewer's comments and responded to their comments.
- V.S and E.P contributed to refining and proofreading subsequent manuscripts drafts.
- V.S supervised manuscript revisions.

Data Availability Statement

The data that support the findings of this thesis are openly available on the Open Science Framework at osf.io/945yg

Acknowledgements

I would like to express my deepest gratitude to my supervisors Professor Vesna Stojanovik and Dr Emma Pagnamenta. Thank you for always believing in me, for strongly recommending my PhD project for funding, for continuously encouraging me and for guiding me throughout this journey. I especially appreciate your patience in reading the many drafts of the papers written for this thesis. It has been a blessing to have mentors like you!

I would like to extend my gratitude to Dr Arpita Bose for welcoming me as an attendee in the ABCD lab meetings and for giving me the opportunity to join/help with many of the many activities she organized. These experiences enhanced my organizational, writing and public-speaking skills.

I am grateful for having shared my PhD journey with my colleague, who is more like a sister, Zakiyah Alsiddiqi. She has been there for me through the ups and downs. I wish I could be as calm and collected as she was when things got tough !

I would not have been able to complete this PhD project without the support of Reading University which offered me an International PhD Studentship that covered my tuition fees. The completion of this project has been financially challenging but was possible thanks to the generous support of the center for Literacy and Multilingualism, the British Federation for women Graduates Foundation, and the Palestinian American Research Center.

I would like acknowledge the efforts of the Palestinian speech and language therapists and teachers, especially, Nairouz Nijm Aldeen, who facilitated the recruitment of the children. A special thank you goes to the parents and children who participated in this study.

I would like to thank my parents Muntaha and Basim Taha for their endless support during my PhD journey. Lastly, I would like to thank my sister Warda who always assured me that everything will be okay at the times when everything seemed to go wrong.

Abstract

Data on the typical and impaired acquisition of Arabic is limited and only a few standardized Arabic language assessments are available. As a result, the identification of developmental language disorder (DLD) in Arabic is notoriously challenging. Developing new diagnostic language tools is thus imperative to facilitate early and accurate identification of DLD in Arabic-speaking children with a view to developing relevant interventions. This thesis addressed this issue by investigating potential clinical markers of DLD in Arabic through three theoretically grounded studies focusing on the linguistic and processing deficits that characterize Arabic speaking children with DLD and could be used as indicators of the presence of the disorder. Study 1 showed that the production of verb tense and subject-verb agreement is generally impaired in 5-year-old Arabic-speaking children with DLD relative to same-age peers. Study 1 showed that poor use of present tense and subject-verb feminine agreement could be potential grammatical markers of DLD in Arabic. Study 2 revealed that nonword repetition is an area of difficulty for 4 to 6-year-old Arabic-speaking children with DLD. Importantly, Study 2 found that poor nonword repetition accurately identified 93% of children with DLD and 93% of age-matched TD children, suggesting that poor nonword repetition could also be a possible clinical marker of DLD in Arabic. Study 3 reported poor sentence repetition abilities in 4 to 6-year-old Arabic-speaking children; the sentence repetition task correctly identified more than 90% of children with DLD and more than 90% of age-matched TD children. Study 3 thus suggests that poor sentence repetition may also hold promise as a potential clinical marker for the presence or absence of DLD in Arabic. The findings of this thesis could help enhance the diagnostic practices of DLD in Arabic-speaking children by focusing clinicians' attention on relevant tasks which could aid diagnosis. The findings extend our understanding of the underlying mechanisms of DLD. Specifically, the language difficulties of Arabic-speaking children with DLD seem to reflect a combination of deficits in linguistic knowledge and processing capacity. This thesis is the first study to my knowledge to address the issue of clinical markers of DLD in Arabic and as such it paves the way and highlights the need for further research to better characterize the linguistic and non-linguistic, as well as the functional limitations in Arabic-speaking children with DLD.

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

1.1 Motivation behind this thesis

About two in 30 children experience unexplained difficulties with using and/or understanding language that hinder their everyday social functioning and academic progress due to developmental language disorder (DLD; Norbury et al., 2016; Tomblin et al., 1997). DLD is typically diagnosed during early childhood, but for many, it is likely to persist into adolescence (known as Language Disorder; Johnson et al., 1999) and adulthood (Botting, 2020; Clegg et al., 2005). DLD is associated with limitations in areas of functioning that go beyond language itself. For instance, children with DLD are vulnerable to social (for a review, see Lloyd-Esenkaya et al., 2020), emotional and behavioural difficulties (Conti-Ramsden et al., 2013; St Clair et al., 2011; Yew & O'Kearney, 2013). These problems are also evident in adolescents and young adults with a history of DLD (Conti-Ramsden & Durkin, 2008; Durkin et al., 2017; Durkin & Conti-Ramsden, 2007; St Clair et al., 2011; Whitehouse et al., 2009; Yew & O'Kearney, 2013).

The long-term language deficits and functional limitations associated with DLD can negatively impact the quality of life of the affected individuals (Eadie et al., 2018; Nicola & Watter, 2015). Hence, early identification of DLD and the provision of timely intervention are necessary to attenuate the adverse consequences associated with the disorder. Unfortunately, DLD is under-detected (McGregor, 2020; Norbury et al., 2016; Tomblin et al., 1997; Wittke & Spaulding, 2018). Two epidemiological studies (Norbury et al., 2016; Tomblin et al., 1997) revealed that only a small proportion of children who met the criteria of DLD were receiving intervention. The situation is likely to be worse in developing countries.

In Palestine, the diagnosis of Arabic-speaking children with DLD – which are the focus of this thesis- is notoriously challenging. There is a shortage of research on typical and impaired Arabic language acquisition. As a result, few Arabic language assessment tools are available (see ELO-L for Lebanese Arabic, Zebib et al., 2019 and ALEF for Gulf-Arabic, Rakhlin et al., 2021). To assess the language abilities of Arabic-speaking children, speech and language therapists (SLTs) tend to rely on informal assessments (Khoja, 2017). The interpretation of the findings of these assessments is highly dependent on the subjective judgment and clinical experience of the SLTs. Consequently, Arabic-speaking children with DLD are at a high risk of being misdiagnosed or not diagnosed, depriving them of their chance of receiving the early intervention services they need.

The issue of DLD under-identification has motivated the concept of clinical markers (Conti-Ramsden et al., 2001a; Conti-Ramsden, 2003; Poll et al., 2010; Rice & Wexler, 1996). A clinical marker refers to behavioural characteristics that are indicative of DLD (Rice & Wexler, 1996). A clinical marker often shows good diagnostic accuracy. That is, it can reliably identify children with DLD (sensitivity) and exclude those without DLD (specificity; Rice & Wexler, 1996). Good diagnostic accuracy is also manifested by the lack of overlap in the performance of children with and without DLD on a clinical marker task (Stokes et al., 2006). To date, tense marking, nonword repetition and sentence repetition have been identified as promising clinical markers of DLD in English (e.g., Archibald & Joanisse, 2009; Ash & Redmond, 2014; Conti-Ramsden, 2003; Conti-Ramsden et al., 2001b; Dollaghan & Campbell, 1998; Redmond et al., 2019; Rice & Wexler, 1996). Importantly, cross-linguistic investigations have revealed that the type and severity of DLD symptoms depend on the structural characteristics of the language being acquired (for a review, see Leonard, 2014). Therefore, clinical markers may vary from one language to another. Therefore, there is a need to determine the language-specific clinical markers in children with DLD and how we can best identify DLD across languages. This thesis extends this strand of research to Arabic.

Little is known about how DLD manifests in Arabic (e.g., Abdalla et al., 2013; Abdallah & Crago, 2008; Balilah, 2017; Saiegh-Haddad & Ghawi-Dakwar, 2017; Shaalan, 2010). To address this gap, this thesis aims to define the language profiles of Arabic-speaking children with DLD. Specifically, the potential of verb morphology production, nonword repetition and sentence repetition as possible clinical markers of DLD in Arabic is investigated.

The remainder of this chapter briefly reviews the recent changes in terminology and criteria used for DLD. It also summarises the difficulties that children with DLD demonstrate across language domains and beyond language itself. This is followed by an overview of what is known about DLD in Arabic and an overview of some of the most prominent theoretical accounts of DLD. Lastly, the specific aims of this thesis are stated.

1.2 Developmental Language Disorder: A debate over terminology and criteria

Different labels have been used to describe children with unexplained language problems, including Specific Language Impairment (SLI), Primary Language Disorder (PLI), Language Learning Impairment (LLI) and Developmental Language Disorder (DLD). Since the early 1980s, researchers have widely adopted the term SLI (e.g., Leonard, 1981). SLI refers to children who demonstrate difficulties with understanding and/or using language that occur without an identified cause. These children have within-normal hearing and age-appropriate non-verbal abilities and do not show neurological/physical deficits (Leonard, 2000).

Several issues have been raised regarding the terminology and criteria of SLI (for a discussion, see Bishop, 2014; Ebbels, 2014; Reilly et al., 2014). Reilly et al. (2014) argued that the term SLI denoted "pure" language difficulties. However, there is growing evidence showing that many (though not all) children with SLI present with co-morbid deficits in non-linguistic domains such as executive functioning (e.g., Im-Bolter et al., 2006; Pauls & Archibald, 2016), motor control (e.g., Flapper & Schoemaker, 2013), memory (e.g., Archibald & Gathercole, 2006; Montgomery et al., 2010; Vugs et al., 2013) and processing of music and speech (e.g., Ladányi et al., 2020) among others. Another primary concern was the inconsistency of thresholds used for SLI exclusionary criteria. In research and clinical practice, the cut-off scores indicating low language ability varied widely (e.g. the cut-off scores used were at or below -1.25, -1.5 or -2 standard deviations on a standardized language test; for a review, see Nitido & Plante, 2020) and were described as being arbitrary (Spaulding et al., 2006). In terms of cognitive referencing, while most SLI studies excluded children whose nonverbal IQ scores were less than 85, in some studies, children with SLI had nonverbal IQ scores between 70 and 90 (for a meta-analysis, see Gallinat & Spaulding, 2014). Overall, the interpretation of SLI exclusion criteria was not universal (Reilly et al., 2014).

To tackle these issues, a panel of 59 experts from different disciplines (the CATALISE consortium) took part in two Delphi exercises to reach a consensus on the criteria of identifying language disorders in children (Bishop et al., 2016, 2017). The CATALISE consortium endorsed the use of the term "Developmental Language Disorder (DLD)". DLD refers to receptive and/or expressive language difficulties of unknown etiology; these difficulties are likely to persist into school age and beyond and impede everyday social interaction and educational progress

(Bishop et al., 2017). The DLD definition highlights the long-term nature of the language difficulties (poor prognosis) and their negative impact on daily functioning, which are unlikely to resolve without intervention (Bishop et al., 2017). Like SLI, DLD does not apply to children whose language difficulties are associated with differentiating conditions (e.g., hearing loss, autism spectrum disorder). For these children, the suggested label is "Language Disorder associated with X". DLD is inclusive of children whose language difficulties are associated with impairments in attention, motor coordination, social-emotional and behavioural functioning, literacy and speech. These impairments often co-occur (though not necessarily cause) DLD and could impact the pattern of language difficulties and response to treatment (see Figure 1.1; Bishop et al., 2017).

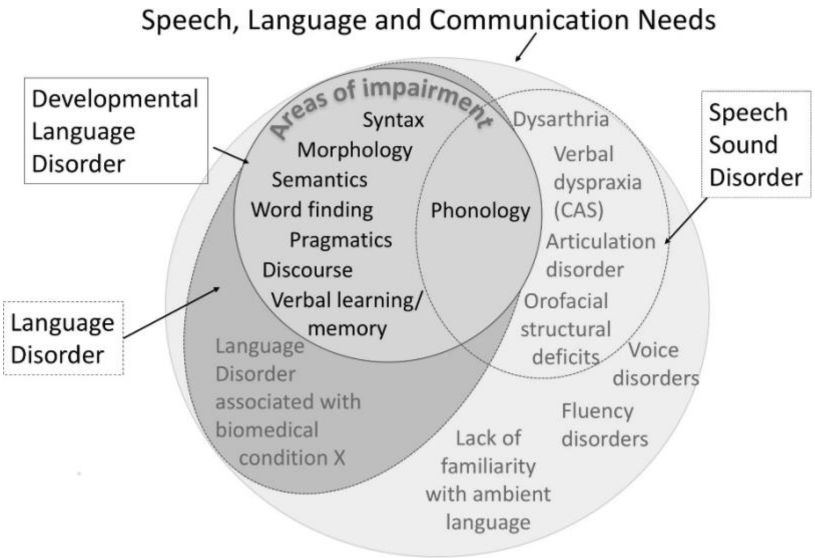

Figure 1.1. *Venn diagram showing the relationship between DLD and other communication disorders (from Bishop et al., 2017)*

DLD is less restrictive than SLI in that it includes children whose nonverbal cognitive abilities are below average but do not meet the criteria of intellectual disability (i.e., non-verbal IQ scores between 70 and 85). According to the CATALISE consortium, cognitive referencing is not required for diagnosing DLD. This decision was based on research findings showing that children with high and low non-verbal IQs did not differ significantly in their language characteristics (e.g., Norbury et al., 2016) or in their response to intervention (e.g., Bowyer-Crane et al., 2011; Ebbels, 2014a). Furthermore, the IQ scores of children who initially met the criteria for SLI were found to change (i.e., drop) over time (Cole et al., 1995). Many children with low non-verbal IQ have sufficient language functioning which challenges the view that low nonverbal ability limits the rate of language acquisition. Hence, non-verbal IQ scores cannot provide a reliable basis for DLD classification/diagnosis (Bishop et al., 2016). Lastly, the procedures of quantifying the significance of language difficulties differ across the SLI and DLD approaches. SLI diagnosis relies heavily on low scores on standardized language tests. To diagnose DLD, the CATALISE consortium emphasizes that, in addition to low language scores, clinicians should consider the day-to-day language functioning of the children in educational and social contexts (Bishop et al., 2016, 2017; McGregor et al., 2020).

This thesis will follow the criteria and terminology of DLD as recommended by the CATALISE consortium. It is acknowledged that previous studies have used different diagnostic labels and criteria. The criteria used in these studies might be more restrictive yet consistent with the DLD label. Therefore, for the sake of consistency, we use the term DLD when referring to these studies.

1.3 A look at the profiles of children with DLD

1.3.1 Syntactic and morphological deficits

Morpho-syntax is a core area of weakness in children with DLD (Leonard, 2014). In English, children with DLD have difficulties with using tense and agreement morphemes in obligatory contexts. Compared to age-matched and younger language-matched typically developing (TD) children, English-speaking children with DLD are significantly more likely to omit past tense –*ed*, present third-person singular –*s*, auxiliary and copula *be* forms and auxiliary *do* forms (e.g., Bedore & Leonard, 1998; Hadley & Rice, 1996; Rice & Wexler, 1996; Rice et al., 1995; Rice & Blossom, 2013). This difficulty with verb morphology is longstanding and continues throughout the school years (Conti-Ramsden et al., 2001; Rice et al., 1998; Rice et al., 2009). Moreover, measures of finite verb morphology are shown to have good accuracy in differentiating children with DLD from TD children (e.g., Bedore & Leonard, 1998; Conti-Ramsden, 2003; Conti-Ramsden et al., 2001; Gladfelter & Leonard, 2012; Rice & Wexler, 1996). Accordingly, poor verb finiteness marking has been established as a clinical marker of DLD in English. English-speaking children with DLD appear to be similar to younger language-matched TD peers in using articles (Leonard et al., 1992; Rice & Wexler, 1996) and noun plurals (Oetting & Rice, 1993), suggesting that these forms are not as problematic as finite verb morphology for these children.

Empirical evidence suggests that the grammatical difficulties of children with DLD vary from one language to another (for a review, see Leonard, 2014a). Unlike the findings from English, verb morphology is not severely impaired in Italian-speaking children with DLD, at least in comparison to their language-matched TD peers (Leonard et al., 1992). On the other hand, Italian-speaking children with DLD are significantly less accurate than age and language-matched TD peers in using articles and clitics (Bortolini et al., 1997; Leonard et al., 1992), and are more likely to omit them (Bortolini et al., 2006; Guasti et al., 2016; Leonard & Dispaldro, 2013). The omission of third-person direct object clitic pronouns is a clinical marker of Italian-speaking children with DLD during pre-school and school-age years (Arosio et al., 2014; Bortolini et al., 2006; Guasti et al., 2016). The third-person plural morpheme seems to be the only present tense inflection used by Italian-speaking children with DLD with a lower accuracy than language-matched children (Bortolini et al., 1997, 2006; Leonard et al., 1992; Leonard & Dispaldro, 2013). Finally, Italian-speaking children with DLD mark noun plurals, noun-adjective agreement and third-person singular verb inflections with similar accuracy to their language-matched peers (Leonard et al., 1992).

Parallel to the findings in Italian, Spanish-speaking children with DLD seem to have pronounced difficulties in producing direct object clitics (Bedore & Leonard, 2001; Castilla-Earls et al., 2020) and articles (Bedore & Leonard, 2005). They also have some difficulty with using verb morphology, especially with marking subject-verb agreement (Castilla-Earls et al., 2020). Unlike Italian-speaking children with DLD, Spanish-speaking children with DLD show difficulties with marking adjective-noun agreement (Bedore & Leonard, 2001). The lower level of difficulty with verb morphology observed in Italian and Spanish (compared to English) could be because the verb paradigms in these languages are relatively transparent and phonologically simple (For a review, see Leonard, 2014). Furthermore, in Spanish and Italian, verb inflections are syllabic, and most of the tense inflections are word-final syllables that follow strong syllables (Bedore & Leonard, 2001a; Gerken, 1996).

On the other hand, object clitics are final-word consonants or weak syllables that usually initiate a phrase or follow a weak rather than a strong syllable (Bedore & Leonard, 2001a). Accordingly, the lower perceptual saliency of object clitics and articles may pose a challenge to Spanish and Italian-speaking children with DLD which could explain their pronounced difficulties with using these structures (see the Surface Account in section 1.5.2). Verb morphology does not appear to be greatly affected in children DLD acquiring Semitic languages. For example, in Hebrew, the difference between children with DLD and younger language-matched TD peers in using verb inflections appears to be structure-specific (Dromi et al., 1993, 1999; Leonard & Dromi, 1994; Leonard et al., 2000). That is, Hebrew-speaking children with DLD are less accurate than language-matched controls in using past tense agreement inflections, but they are as proficient in marking present tense agreement inflections. In Hebrew, the past tense agreement paradigm requires marking three categories (number, gender, and agreement) whereas the present tense agreement paradigm only involves marking two categories (gender and number; Dromi et al., 1999).

Additionally, Hebrew-speaking children with DLD seem to have difficulty with producing noun plural inflections and noun-adjective agreement inflections compared to age-matched but not language-matched peers (Dromi et al., 1993). Overall, it is clear that morphological deficits in children with DLD are not uniform across languages. Rather, the type and severity of morphological difficulties caused by DLD depend on the features of the grammatical system of a given language (Leonard, 2014a).

Research has identified several aspects of grammar that appear to be cross-linguistically impaired in children with DLD. Such grammatical structures may be described as "language-universal" or "language-independent". Typically, these are syntactic constructions with linguistic operations involving long-distance dependencies such as syntactic movement (e.g., object Wh-questions) and embedding (e.g., relative clauses). For instance, the production and comprehension of Wh- object questions pose challenges for children with DLD across several languages such as English, Greek, French, Hebrew and Swedish (Deevy & Leonard, 2004; Ebbels & Van Der Lely, 2001; Fleckstein et al., 2018; Friedmann & Novogrodsky, 2011; Hansson & Nettelbladt, 2006; Jakubowicz, 2011a; Prévost et al., 2014; Stavrakaki, 2006). Similarly, difficulties with clausal complements have been previously noted in Czech, English, French, Greek-speaking children with DLD (Eisenberg, 2004; Fleckstein et al., 2018; Mastropavlou & Tsimpli, 2011; Owen & Leonard, 2006a; Smolík & Vávrů, 2014). Furthermore, relative clauses pose special difficulties for children with DLD in many languages such as English, Hebrew, Italian, French, German, Danish and Arabic (Adani et al., 2016; Arosio et al., 2009; De López et al., 2014; Fleckstein et al., 2018; Friedmann & Novogrodsky, 2004; Frizelle et al., 2017; Frizelle & Fletcher, 2014; Hestvik et al., 2010; Novogrodsky & Friedmann, 2006; Riches, 2017; Shaalan, 2010). Though under-investigated, passive constructions are also problematic for children with

DLD acquiring English, Cantonese and Catalan (Gavarró, 2017; Leonard et al., 2006; Marinis & Saddy, 2013).

Sentence repetition is a task that has often been used to investigate the morpho-syntactic abilities in children (Devescovi & Caselli, 2007; Komeili & Marshall, 2013; Seeff-Gabriel et al., 2008). There have been divergent views regarding what the task exactly measures, but it is now generally agreed that sentence repetition is a complex task that taps into underlying linguistic representations and processing abilities (see section 4.1.1; Marinis & Armon-Lotem, 2015; Moll et al., 2015; Potter & Lombardi, 1998; Riches, 2012). Research has consistently shown that children with DLD score below age-matched TD children on sentence repetition tasks, this finding has been documented in different languages such as English (Conti-Ramsden et al., 2001; Redmond et al., 2011; Seeff-Gabriel et al., 2010), Cantonese (Stokes et al., 2006), Italian (Devescovi & Caselli, 2007), French (Fleckstein et al., 2018; Leclercq et al., 2014), Catalan (Gavarró, 2017), Vietnamese (Pham & Ebert, 2020) as well as Hebrew and Russian (Armon-Lotem & Meir, 2016) among others. In these languages, sentence repetition has demonstrated good levels of accuracy in differentiating children with DLD from TD children (see section 4.1.2), thus revealing that sentence repetition could be an effective clinical marker of DLD (for a recent review, see Rujas et al., 2021). The fact that sentence repetition has been identified as an area of vulnerability in children with DLD across typologically diverse languages points out its potential as a universal clinical marker of DLD (Pham & Ebert, 2020).

1.3.2 Phonological deficits

In this section, phonological deficits refer to speech production difficulties that are linguistic in origin (e.g., the child fails to make distinctions between speech sounds that convey different meanings). According to the CATALISE consortium, children who only demonstrate phonological problems that are not accompanied by language difficulties do not meet the criteria of DLD. Rather, these children may be diagnosed with a phonological disorder, a sub-category of the more generic label Speech Sound Disorder (SSD; Bishop et al., 2017). SSD is also used for children with speech production errors due to motor or physical abnormalities (e.g., dysarthria or orofacial structural deficits; see Figure 1.1). DLD and SSD may co-occur and children who present with language deficits alongside speech difficulties of a motor/structural origin qualify for both labels (Bishop et al., 2017).

In the initial descriptions of DLD, a deficit affecting phonology and morpho-syntax was identified as the most prevalent subcategory of the disorder (Rapin & Allen, 1988; Rapin & Allen, 1983). Many children with DLD under this category exhibited omissions, substitutions and distortions of consonants and consonant clusters, produced unrecognizable phonemes and had unintelligible speech. These deficits were more severe than the phonological processes observed in young TD children (Rapin & Allen, 1983). Children with DLD were reported to make atypical phonological errors, i.e., unusual sound changes (e.g., initial consonant deletion) as well as typical phonological errors (e.g., weak syllable deletion) that remained beyond what is considered to be age-appropriate (Menyuk, 1993). In comparison to age-matched peers, children with DLD have been found to vocalize less frequently, have restricted phonetic inventories, low speech intelligibly and use a limited range of syllable shapes (Pharr et al., 2000; Mirak & Rescorla, 1998; Paul & Jennings, 1992; Rescorla et al., 1996; Roberts et al., 1998). Toddlers at risk of DLD have been reported to be similar to age-matched controls in their production frequency of simple syllabic structures (e.g. CV), but were less proficient in producing complex syllabic structures such as those containing two or more different consonants (e.g., [dogi] "doggy"), final consonants (e.g., CVC, [kʌp] "cup") or consonant clusters (e.g., [dont] "don't"; Pharr et al., 2000).

In several languages, the phonological production abilities of children with DLD were found to be impaired relative to younger language-matched TD children (Aguilar-Mediavilla et al.,2002; Aguilar-Mediavilla, 2013; Beers, 1995; Bortolini & Leonard, 2000; Maillart & Parisse, 2006; Owen et al., 2001). In comparison to TD children matched on phonetic inventory and Mean Length of Utterance (MLU), English-speaking children with DLD have been reported to be less accurate than younger TD controls in the production of word-initial and word-final consonants, non-final weak syllables, and word-final consonant clusters. A similar pattern was observed in Italian-speaking children with DLD who also exhibited an additional difficulty with producing non-final consonant clusters (Bortolini & Leonard, 2000). Interestingly, English and Italian-speaking children presented with phonological difficulties even when MLU (grammatical abilities) was considered. This suggested that the phonological limitations in children with DLD are above and beyond their morpho-syntactic difficulties (Bortolini & Leonard, 2000). This notion also comes from Hebrew, a language where phonology and morphology closely interact (Owen et al., 2001). Hebrew-speaking children with DLD have been found to have a higher rate of phonological errors relative to younger TD controls matched on

Mean Morpheme per Utterance (MPU; an index of morpho-syntactic development). The phonological errors could not be accounted for by the impaired morpho-syntactic abilities as they were evident in consonants and constant clusters that did not carry any grammatical function. However, it was acknowledged that phonological errors could exacerbate morphological deficits in children with DLD (Owen et al., 2001).

Children with DLD appear to have weaker phonological processing skills relative to TD peers. This is evidenced by their poor performance on measures of phonological awareness, rapid automated naming and verbal short-term memory (e.g., Brewer et al., 2016; Briscoe et al., 2001; Catts et al., 2005; Claessen et al., 2013; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Kelso et al., 2007; Leitãto et al., 1997; Ramus et al., 2013; Vandewalle et al., 2012, 2010). In a 3-year-longitudinal study, Vandewalle et al (2012) found that Dutch-speaking children with DLD had a lower composite phonological awareness score than that of the age-matched TD group in Grades 1, 2 and 3. Additionally, the DLD group obtained significantly lower scores than TD children on digit span and nonword repetition tasks, suggesting an impairment in verbal short-term memory (Vandewalle et al., 2012). In another study, 11-year-old, English-speaking children with DLD scored significantly below TD children on several phonological skills including phonological awareness (Rhymes and Spoonerisms), rapid automated naming (Digit naming), production and perception of melody and prosody as well as measures of phonological representations (nonword repetition, nonword discrimination, word-picture matching of minimal pairs; Ramus et al., 2013). It is important to point out that while many children with DLD show phonological impairments, many only show mild or no deficits (Catts et al., 2005; Gardner et al., 2006; Kelso et al., 2007; Ramus et al., 2013).

Nonword repetition is of a particular focus in this thesis. In this task, children are asked to repeat meaningless (nonsense) stimuli. Research has found that performance on nonword repetition tasks is correlated with several language indices such as vocabulary (e.g., Gathercole, 2006), sentence repetition (Almeida et al., 2017; Thordardottir & Brandeker, 2013), word repetition (e.g., Dispaldro et al., 2013), and standardized language tests (e.g., Gray, 2003). Given the link between nonword repetition and language abilities, it is unsurprising that children with DLD have difficulties with nonword repetition relative to their TD peers (Chiat, 2015; Graf Estes et al., 2007). Poor performance on nonword repetition has often been interpreted as evidence of a deficit in verbal short-term memory (Gathercole & Baddeley, 1990). However, besides the storage of phonological information, accurate nonword repetition involves auditory perception, encoding of phonological information, motor planning and articulation. A deficit in any of these components may lead to poor nonword repetition (Archibald et al., 2013; Coady & Evans, 2008; Pigdon et al., 2019). Generally, poor nonword repetition performance has been proposed as a phenotypic marker of DLD (Bishop et al., 1996; Conti-Ramsden et al., 2001; Conti-Ramsden, 2003). Moreover, nonword repetition tasks have been found to have good discriminative power in detecting DLD in monolingual children (see section 3.1.3) and bilingual children across many languages (for a recent review and meta-analysis, see Schwob et al., 2021). These findings suggest that nonword repetition not only could be a robust and recommendable cross-linguistic task for detecting DLD (Pham & Ebert, 2020) but also deserves to have a fixed place (in combination with other tools) in language assessment protocols of monolingual and bilingual children with DLD (Schwob et al., 2021).

1.3.3 Lexical-semantic deficits

Vocabulary deficits are well-attested in children with DLD. There is evidence of a persisting gap in receptive vocabulary between individuals with DLD and TD controls from 2 to 21 years of age (Rice & Hoffman, 2015). The lexicons of children with DLD are often described as being limited in terms of breadth (i.e., number of words known) and depth (richness of word knowledge; Dollaghan, 1998; Kail & Leonard, 1986; Leonard et al., 1983; McGregor & Waxman, 1998; McGregor et al., 2002, 2013; Sandgren et al., 2020). For instance, relative to TD classmates, children with DLD have been reported to know fewer words and have shallower knowledge of word meanings across all grades (McGregor et al., 2013). In a recent study, six to nine-year-old Swedish-speaking children with DLD demonstrated significantly lower scores than TD peers in tasks of picture recognition, paradigmatic and syntagmatic associations and verbal fluency, suggesting that lexical organization is an area of weakness in these children (Sandgren et al., 2020).

Novel word learning is also compromised by DLD in preschool-age children (e.g., Gray, 2004; Rice et al., 1990; Storkel et al., 2017; Windfuhr et al., 2002), school-age children (e.g., Alt & Spaulding, 2011; Bishop & Hsu, 2015; Gray et al., 2020; Jackson et al., 2016, 2021; Nash & Donaldson, 2005; Oetting et al., 1995) and even in adults (e.g., McGregor et al., 2020; McGregor et al., 2013). Experimental word-learning studies revealed that children with DLD, as a group, often require greater exposure relative to TD peers to learn a similar number of new words (e.g., Gray, 2004; McGregor et al., 2013; Nash & Donaldson, 2005), though this is not always the case (see Gray et al., 2012; Gray & Brinkley, 2011). Learning a novel word comprises several processes including encoding, re-encoding and retention (for a review of word learning models, see Gray et al., 2020). Research suggests that the word-learning difficulties in children and adults with DLD may be attributed to a deficit in the encoding stage (Bishop & Hsu, 2015; Haebig et al., 2019; Jackson et al., 2021; Leonard et al., 2019; McGregor et al., 2013).

Another common feature of children with DLD is their word-finding difficulties. This difficulty may manifest as "long pauses in speech, circumlocution i.e., speaking around the topic, and/or the frequent use of non-specific words such as *it* or *stuff* " (Leonard, 2014, p.57). The storage-elaboration hypothesis postulates that the word-finding difficulties in children with DLD are not due to an impairment in the word retrieval mechanism itself, rather, due to underdeveloped semantic representations in the lexicons of these children (Kail & Leonard, 1986). As mentioned earlier, not only do children with DLD know fewer words than age-matched TD peers, but they also have less information about the words they know (limited vocabulary breadth). That is, they have a weaker representation of the words in the mental lexicon. This means that less semantic information is available to guide the retrieval of words (Kail & Leonard, 1986). In turn, this may hinder accurate and efficient word retrieval.

1.3.4 Pragmatic deficits

Children with DLD may face challenges with pragmatics i.e., using language for communication and social interaction. Their linguistic pragmatic deficits entail: a literal interpretation of figurative language such as idioms, metaphors, irony and implicatures (e.g., (Andrés-Roqueta & Katsos, 2020; Bishop & Adams, 1992; Bühler et al., 2018; Katsos et al., 2011; Norbury et al., 2004; Norbury, 2005; Rinaldi, 2000), inefficient processing of contextual information to resolve lexical ambiguities (e.g., Norbury, 2005a), poor inferencing skills (e.g., Adams et al., 2009; Bishop & Adams, 1992; Botting & Adams, 2005; Ford & Milosky, 2008; Lucas & Norbury, 2015; Newton et al., 2010; Weismer, 1985) as well as difficulties with formulating narratives as indicated by lack of cohesion (Swanson et al., 2005), poor story quality and limited use of story grammar and components (Fey et al., 2004; Norbury et al., 2014).

Children with DLD may exhibit inadequate conversational skills. In comparison to age-matched TD peers, children with DLD are less likely to respond verbally or non-verbally to adult utterances and, when they respond verbally, they are more likely to give inadequate and/or pragmatically odd responses (e.g., Bishop et al., 2000). Children with DLD use a lower rate of multi-word responses and a higher rate of nonverbal responses. They are also less likely to initiate verbal interactions with other children, rather, they show a tendency of initiating verbal interactions with adults (Rice et al., 1991). Children with DLD may experience difficulties with topic maintenance and show violations of conversational turn-taking rules by excessively interrupting their conversational partner (Adams & Bishop, 1989). Children with DLD seem to be as sensitive as TD peers to violations of the conversational Grecian maxims of truthfulness, relevance and politeness but are less sensitive to violations to the maxim of quantity (Surian et al., 1996). In conversational exchanges, children with DLD tend to produce over-informative or under-informative utterances (Adams & Bishop, 1989). More recently, Davies and colleagues (2016) suggested that children with DLD are sensitive to pragmatic violations of quantity maxims but display a more tolerant/accepting attitude towards these violations relative to TD peers.

Some argue that the pragmatic abilities of children with DLD are in proportion with their structural language abilities (Andrés-Roqueta & Katsos, 2020; Norbury et al., 2004). This led to the suggestion that the pragmatic deficits in children with DLD may stem from or be secondary to limited competence in language production and/or comprehension i.e., structural language deficits (Bishop, 2000; Davies et al., 2016; Hadley & Rice, 1991). However, this view has been challenged. Researchers have identified a subgroup[1] of children with DLD who present with substantial pragmatic language impairments that are not proportionate with their structural language skills (Bishop et al., 2000b; Bishop, 2000; Brinton et al., 1997; Friedmann & Novogrodsky, 2008). Accordingly, studies have suggested that factors beyond structural language deficits could contribute to the pragmatic difficulties in children with DLD such as limitations in processing capacity (Bishop & Adams, 1991), social cognition (Andrés-Roqueta et al., 2016; Bakopoulou & Dockrell, 2016), emotional regulation (Fujiki et al., 2002).

1 "Pragmatic Language Impairment (PLI)" (Bishop, 2000), "semantic pragmatic deficit syndrome" (Rapin & Allen, 1983), or more recently, "social (Pragmatic) Communication Disorder" (DSM-5, American Psychiatric Association, 2013) are examples of taxonomies that have been used to describe children with significant pragmatic difficulties that are disproportionate with their overall structural language abilities. There are many concerns about use of these terms which are beyond the scope of this thesis (see Norbury, 2014). In line with Bishop et al (2017), poor pragmatic abilities is considered an area of impairment within DLD.

1.3.5 Beyond the language difficulties

Children with DLD often exhibit reading difficulties (Adlof, 2017; Botting et al., 2006; Kelso et al., 2007; Snowling et al., 2020). It is estimated that 20 to 80% of children who are diagnosed with DLD also meet the criteria for Dyslexia (Catts et al., 2005; McArthur et al., 2000; Snowling et al., 2019, 2020). Poor literacy skills have also been observed in adolescents (Snowling et al., 2001) and young adults with a history of DLD (Conti-Ramsden et al., 2018; Law et al., 2009, 2013). Longitudinal investigations generally reveal that young adults with DLD, as a group, display lower educational and employment achievements than their peers, though individual differences exist (for a review, see Dubois et al., 2020). For some but not all children with DLD, social skills and peer relationships is an area of vulnerability (for a review, see Lloyd-Esenkaya et al., 2020; Toseeb & St Clair, 2020). According to teacher reports, children with DLD tend to be more socially withdrawn (Fujiki et al., 1999; Hart et al., 2004), more likely to be left out of play (Fujiki et al., 1996) and are at a higher risk of being bullied and/or victimized relative to TD peers (Conti-Ramsden & Botting, 2004; Redmond, 2011). Moreover, evidence suggests that children with DLD face difficulties with initiating or joining ongoing social interactions (Gibson et al., 2013; Hadley & Rice, 1991; Liiva & Cleave, 2005) and with managing and solving conflicts (Bakopoulou & Dockrell, 2016; Campbell & Skarakis-Doyle, 2011; Timler, 2008). Children with DLD have smaller peer social networks (Chen et al., 2020), have a higher risk of being rejected by peers (Fujiki et al., 1999) may be less liked by their peers (Fujiki et al., 2013) and, according to parental reports, have significantly fewer close friends than TD peers (Redmond, 2011). Furthermore, adolescents with a history of DLD appear to have difficulties with peer interactions (St Clair et al., 2011) and tend to have poorer quality of friendships (Durkin & Conti-Ramsden, 2007). Moreover, young adults with a history of DLD are less socially confident, shyer and demonstrate lower levels of social self-efficacy in comparison to their peers (Durkin et al., 2017).

DLD has been also linked to emotional and behavioural difficulties (Conti-Ramsden et al., 2013; St Clair et al., 2011; Yew & O'Kearney, 2013). Relative to TD peers, children with DLD have been reported to show higher rates of withdrawal (Hart et al., 2004; Maggio et al., 2014). In addition, a meta-analysis revealed that children and adolescents with DLD were approximately two times more likely to show internalizing difficulties, externalizing difficulties and attention/hyperactivity behaviours relative to TD peers (Yew & O'Kearney, 2013).

Adolescents with a history of DLD have been reported to have higher rates of anxiety and depression (Conti-Ramsden & Botting, 2008). It is also documented that adolescents with DLD display lower global self-esteem (Wadman et al., 2008). These individuals are also at a higher risk of social phobia in late adolescence (Voci et al., 2006).

To summarize, children with DLD present with a wide range of difficulties across the different language domains. Morpho-syntax is a core area of deficit in children with DLD. While some aspects of grammar (e.g., verb and noun morphology) appear to be impaired in ways that depend on the typology of language that the child is acquiring, other aspects (e.g., comprehension and production of complex syntactic structures and sentence repetition) seem to be adversely affected by DLD cross-linguistically. In regards to the grammatical markers of DLD, the potential of verb morphology as a clinical marker varies from one language to another whereas sentence repetition appears to be a potential cross-linguistic marker of DLD. In terms of phonology, children with DLD show difficulties with phonological production, processing and perception as well as nonword repetition. It is important to point out that nonword repetition has been consistently found to be impaired in children with DLD in many languages, and has the potential to be a cross-linguistic marker of DLD. The semantic deficits in children with DLD are characterized by a delay in vocabulary acquisition, a limitation in vocabulary depth and breadth in addition to difficulties with word-learning and word-retrieval. Children with DLD may also display pragmatic deficits that could be evident across linguistic and social contexts. Lastly, the impact of DLD extends beyond language and appears to affect the academic performance, social-functioning, emotional-being as well as quality of life of the affected individuals.

1.4 DLD in Arabic: What do we know so far?

Research investigations of DLD in Arabic are relatively limited but emerging. A systematic search of published literature (up until June 2021) on DLD in Arabic was conducted through the following electronic databases: PubMed, PsycInfo and EBSCO. Furthermore, EThOS and ProQuest were used to search for relevant unpublished PhD theses. A summary of the samples, methods and primary findings of these studies is provided in Table 1.1. In terms of **morpho-syntactic abilities,** Arabic-speaking children with DLD display lower Mean Morpheme per Utterance (MPU) scores relative to same-age peers (Abdallah, 2002; Alsiddiqi et al., 2021; Rakhlin et al., 2020), suggesting an overall reduced proficiency in using grammatical inflections.

Several studies have identified verb morphology production as an area of vulnerability in Arabic-speaking children with DLD (Abdallah, 2002; Abdallah & Crago, 2008; Fahim, 2017; see section 2.1.2). For example, Abdallah and Crago (2008) reported that Hijazi Arabic-speaking children with DLD were significantly less accurate than age and language-matched TD children in using verb tense and agreement forms. However, they noted that not all verb forms were challenging for Arabic-speaking children with DLD. That is, structurally complex verb forms (e.g., marked with inflections) such as present tense, third person and feminine verbs posed more difficulty for children with DLD than past tense, first-person and masculine verbs, respectively. The latter verbs are structurally less complex (Abdallah & Crago, 2008; see section 2.1.1). Overall, similar to English, Spanish and Hebrew-speaking children with DLD, Arabic-speaking children with DLD have difficulties with verb morphology. Yet, the nature of these difficulties is different. For example, past tense is greatly problematic for English (Rice et al., 1996) and Hebrew-speaking children with DLD (Dromi et al., 1999), but it is less severely affected in Arabic. Rather, Arabic-speaking children with DLD seem to have more pronounced deficits with the use of present tense forms. Generally, verb morphology appears to be an area of impairment in Arabic-speaking children with DLD. However, it is important to identify which verb forms could serve as clinical markers of DLD in Arabic.

Arabic-speaking children with DLD have considerable difficulties with the use of determiners (definite articles) compared to age and language-matched peers (Abdallah, 2002). These children also have obvious problems with using prepositions (Abdallah, 2002), noun plurals (Abdallah et al., 2013) and bound pronouns (Faquih, 2014) relative to same-age TD children.

Shaalan (2010) found that Gulf Arabic-speaking children with DLD scored below same-age TD peers on sentence comprehension and that the performance of the DLD group on this task was similar to that of TD children who are two years younger. Shaalan (2010) noted that Arabic-speaking children with DLD had particular difficulties with the comprehension of sentences with non-canonical word order (e.g., sentences with clitic left dislocation). Furthermore, Arabic-speaking children with DLD perform poorly on sentence repetition tasks (Alsiddiqi et al., 2021; Rakhlin et al., 2020; Shaalan, 2010). According to Shaalan (2010), the repetition of subject and object relative clauses were especially challenging for Arabic-speaking children with DLD. Overall, sentence repetition and the comprehension of complex syntactic structures are impaired in Arabic-speaking children with DLD, just as is the case in other languages (see section 1.3.1). The observed deficits in sentence repetition tasks in Arabic-speaking children with DLD reinforce the potential of sentence repetition as a cross-linguistic clinical marker of DLD, especially in Arabic (for a review see Rujas et al., 2021)

Table 1.1 *Summary of Arabic studies on DLD*

Study Arabic Dialect	Sample	Tests/material	Main findings
Abdallah (2002) Hijazi Arabic	10 DLD (M_{age} = 57.3 months) 10 TDAM (M_{age} = 56.7months) 10 TDLM (M_{age} = 28.7 months ; matched on MLU)	Analysis of spontaneous language samples	• DLD < TDAM & TDLM in using tense, subject-verb agreement and determiners (definite articles). DD < TDAM in using prepositions.
Abdallah et al (2013) Kuwaiti Arabic	12 DLD (M_{age} = 55.7months) 12 TDAM (M_{age} = 55.6 months)	A picture description task assessing the production of plurals of real and nonsense words.	• DLD < TDAM in producing plural types including feminine sound plurals, masculine sound plurals, and broken plurals across real and nonsense words.
Alsiddiqi et al. (2021) Saudi Arabic	24 DLD (M_{age} = 62.96 months) 40 TD (M_{age} = 65.45 months)	A battery of Arabic language and emergent literacy measures	• DLD < TDAM on phonological awareness skills (syllable segmentation and phoneme awareness), language measures (MPU, vocabulary knowledge, listening comprehension and SR) and verbal STM measures (digit recall and NWR). • Language measures (except for listening comprehension and MPU in the TD group) were positively correlated with emergent literacy skills in both groups.
Balilah (2017) Hijazi Arabic	52 DLD (M_{age} = 8;4 years) 369 TD (M_{age} = 7;11 years)	A battery of language measures and cognitive measures of verbal and visual short-term and working memory.	• DLD < TD on all language measures: expressive and receptive vocabulary, expressive language test and single-word reading. • DLD < TD on all verbal STM measures (digit recall, word recall and NWR) and verbal WM measures (backwards digit recall). • DLD = TD in visual STM and WM measures: Dot Matrix, Block recall, Odd One Out and Spatial Span.
Fahim (2017) Egyptian Arabic	3 DLD (M_{age} = 54 months) 6 TD (M_{age} = 36 months)	Spontaneous language samples	• DLD < TD in the production of correct tense and subject-verb agreement. Production of the subject-verb agreement was more impaired than tense in the DLD group. DLD showed no difficulty with the use of past tense.

Khater (2016) Qatari Arabic	15 DLD (M_{age} = 43 months)[1] 44 TD (M_{age} = 38 months)	WR, NWR, Arabic expressive and receptive vocabulary Tests.	• DLD < TD in WR and NWR across one, two and three-syllable words and nonwords. Length affected WR and NWR accuracy in both groups.
Rakhlin et al (2020) Gulf-Arabic	150 children (M_{age} = 8;4 years)[2] including 27 with suspected DLD [3] and 150 TD children	A battery of language measures and indices of narrative samples	• DLD < TD on language measures: receptive and expressive vocabulary, SR and NWR. DLD = TD on most narrative microstructure indices. However, DLD > TD in making subject omission errors and DLD < TD in using subject-verb agreement.
Saiegh-Haddad & Ghawi-Dakwar (2017) Palestinian Arabic	25 DLD (M_{age} = 66 months)[3] 25 TDAM (M_{age} = 70 months) 25 DLD (M_{age} = 83 months) 25 TDAM (M_{age} = 83 months)	WR and NWR tasks with 1-4 syllable-long items of simple syllabic structures. Items varied in phonological and lexical distance to standard and spoken Arabic.	• DLD < TD on WR and NWR tasks, even when the forms conformed to the phonology of spoken Arabic. • The phonological novelty effect on WR and NWR was larger in the DLD relative to the TD group. Lexical novelty effects on WR and NWR in both groups were limited.
Shaalan (2010) Gulf Arabic	26 DLD (M_{age} = 85.1 months) 112 TD (M_{age} = 82.2 months, range = 54-112)	Tests of Sentence Comprehension, Expressive Language, SR, Arabic Picture Vocabulary.	• DLD < TD on all tests suggesting deficits in receptive vocabulary, expressive language, the production and comprehension of syntactic structures.
	13 DLD (M_{age} = 95.5 months) 13 TDAM (M_{age} = 95.4 months) 13 TDLM (M_{age} = 68.3 months; matched on sentence comprehension scores).	Sentence Comprehension	• DLD < TDAM & TDLM in the comprehension of complex sentences with non-canonical word order. DLD = TDAM & TDLM in the comprehension of sentences with canonical word order.
Shaalan (2020) Gulf Arabic	11 DLD (M_{age} = 93.9 months), 11 TDAM (M_{age} = 68.3 months) 11 TDLM (M_{age} = 72.3 months, matched on sentence comprehension scores).	NWR test of two and three-syllable nonwords with 0 to 2 CCs	• DLD < TDAM &TDLM in NWR. No length effect was detected in either group. DLD group was sensitive to syllabic complexity (# of CC) of the nonwords. Unlike TDLM, DLD and TDAM groups were not sensitive to the wordlikeness of the nonwords.

Note. **DLD:** Developmental Language Disorder. **TDAM:** Typically Developing Age-Matched. **TDLM:** Typically Developing Language-Matched. **MLU:** Mean Length of Utterance. **MPU:** Mean Morpheme Per Utterance. **SR:** Sentence Repetition. **STM:** Short-Term Memory. **NWR:** Nonword Repetition.**TD:** Typically Developing. **WM:** Working Memory. **WR:** Word Repetition. **CC:** Consonant Clusters.

[1] Though Khater (2016) refers to these children as "Clinical group" due to the absence of non-verbal scores, they meet the criteria for DLD.

[2] Children were allocated across age groups and no overall age statistics were available for TD and DLD children.

[3] Age range and standard deviations were not reported.

Phonological deficits are also evident in Arabic-speaking children with DLD. Recently, Alsiddiqi et al (2021) reported that Saudi Arabic-speaking children with DLD performed significantly lower than TD children in tasks assessing syllable segmentation and phoneme awareness (e.g., initial and final phoneme isolation/deletion), indicating poor phonological awareness skills in Arabic-speaking children with DLD. Furthermore, nonword repetition has been consistently reported as an area of difficulty for monolingual (Alsiddiqi et al., 2021; Balilah, 2017; Khater, 2016; Rakhlin et al., 2020; Saiegh-Haddad & Ghawi-Dakwar, 2017; Shaalan, 2010, 2020) and bilingual children with DLD whose first or second language is Arabic (Abi-Aad & Atallah, 2020). Importantly, when the number of consonant clusters of the nonwords increased from zero consonant clusters to one and two consonant clusters, the repetition accuracy decreased significantly for the DLD group but less so for the age-matched and language-matched TD children (Shaalan, 2010). This suggests that the production of phonologically complex forms (i.e., syllables containing consonant clusters) is impaired in Arabic-speaking children with DLD (Shaalan, 2010, 2020). Given that nonword repetition is impaired in Arabic-speaking children with DLD, and that the task is a clinical marker of DLD across many languages (Schwob et al., 2021), nonword repetition could potentially be a clinical marker of DLD in Arabic.

Much less is known about the **lexical-semantic difficulties** of Arabic-speaking children with DLD. What is known so far is that Arabic-speaking children exhibit limited expressive and receptive vocabularies compared to age-matched peers (Alsiddiqi et al., 2021; Balilah, 2017; Khater, 2016; Rakhlin et al., 2020; Shaalan, 2010). Some studies investigated the **memory abilities** of Arabic-speaking children with DLD. Children with DLD scored lower than same-age TD peers in tasks that assess verbal short-term memory including word repetition (Saiegh-Haddad & Ghawi-Dakwar, 2017), digit recall (Alsiddiqi et al., 2021; Balilah, 2017) and wordlist recall tasks (Balilah, 2017). These findings suggest a reduced capacity of the verbal short-term memory in Arabic-speaking children with DLD. Lastly, the processing of verbal information (verbal working memory) is impaired in Arabic-speaking children with DLD as evident by their poor performance on backward digit recall (Balilah, 2017).On the other hand, the processing of visual and spatial information (non-verbal working memory) appears to be preserved in this population (Balilah, 2017).

In summary, relative to TD peers, Arabic-speaking children with DLD have linguistic deficits characterized by: a) a lower proficiency in using morphological structures as evident by their lower Mean Morpheme per Utterance scores, b) difficulties with using verb tense and subject-verb agreement forms, c) poor use of noun plurals, d) difficulties with producing bound pronouns attached to verbs or nouns, e) poor sentence repetition and comprehension, f) poor phonological awareness skills and g) difficulties with nonword repetition, and h) reduced expressive and receptive vocabulary. There's also evidence of limitations in verbal short-term and working memory in Arabic-speaking children with DLD.

1.5 Theoretical frameworks of DLD

Different theoretical frameworks have been put forward to explain the language deficits of children with DLD and to capture the nature of the causal mechanisms that underlie the disorder. Theories of DLD broadly fall into three categories:

(1) **Theories which postulate that DLD stems from a deficit in linguistic knowledge** such as the Extended Optional Infinitive account (Rice & Wexler, 1996; Rice et al., 1995; Rice et al., 1998) and its updated versions the Agreement/Tense Omission Model (Schütze & Wexler, 1996; Wexler et al., 1998) and the Extended Unique Checking Constraint (Wexler, 1998, 2003); the Grammatical Agreement Deficit account (Clashen & Hansen, 1997; Clashen, 1989, 1991); as well as the Representation Deficit for Dependent Relationships account (van der Lely, 1998; van der Lely & Stollwerck, 1997) and it's extended version the Computational Grammatical Complexity account (Marshall & van der Lely, 2006, 2007; Marshall, 2006; Van Der Lely, 2005).

(2) **Theories that view DLD as a limitation in information processing capacity or speed** such as the Generalized Slowing Hypothesis (Kail, 1994), the phonological short term memory (Baddeley, 2003; Gathercole & Baddeley, 1990, 1993; Gathercole, 2006) and working memory accounts (Archibald & Gathercole, 2006; Jakubowicz, 2011; Montgomery, 1995b, 2000), the Surface account (Leonard, 1989; Leonard et al., 1992, 1997) and the Morphological Richness Account (Dromi et al., 1999; Leonard et al., 1987; Lukács et al., 2009).

(3) **Theories that attribute DLD to a deficit in specific mechanisms** such as deficits in auditory temporal processing (Tallal & Piercy, 1973, 1974), rhythmic processing and production (Ladányi et al., 2020), or procedural learning (Ullman & Pierpont, 2005). In the following section, some of the most prominent theories of DLD are reviewed. The review is restricted to the theoretical

accounts that are most pertinent to the studies reported in this thesis (for an exhaustive review of the different DLD theories, see Leonard, 2014).

1.5.1 DLD as a deficit in linguistic knowledge

Knowledge-based (alternatively known as representational) accounts propose that DLD is caused by inadequate knowledge of the language rules, principles or constraints. These accounts, however, vary in terms of the nature/type of incomplete linguistic knowledge in children with DLD (Leonard, 2014a). **The Extended Optional Infinitive** (EOI; Rice & Wexler, 1996; Rice et al., 1995; Rice et al., 1998) proposes that the deficits with verb finiteness marking observed in children with DLD could be attributed to a delay in the emergence of a biologically-based maturational principle. This theory is based on a developmental framework suggesting that English-speaking TD children go through a stage in which they treat finiteness marking as being optional in obligatory contexts (Wexler, 1994). Accordingly, they omit inflections that mark tense and agreement features including past tense *–ed*, present tense third-person singular *–s* and auxiliary *do* and *be* forms (Rice & Wexler, 1996).

When children fail to mark finiteness features, they produce a bare stem i.e., the infinitive (e.g., eat for eats, chase for chased). Hence, this stage is referred to as the optional infinitive stage. The EOI further proposes that when children at the optional infinitive stage produce tense and agreement markers, they do so correctly. It is argued that children with DLD remain at the optional infinitive stage for a protracted period due to a slower emergence of a maturational principle. This principle is necessary for the linguistic representations of finiteness to become fully specified so that children move out of the optional infinitive stage. This means that the omission of tense and agreement inflections lasts longer than usual in children with DLD (Rice & Wexler, 1996).

Evidence suggests cross-linguistic variations in the characteristics of the optional infinitive stage. For instance, the use of infinitive forms in place of correct tense and agreement is infrequent in the expressive language of Italian (Bortolini et al., 1997) and Spanish-speaking children with DLD (Bedore & Leonard, 2001). In these languages, when children with DLD fail to produce the target tense or agreement forms, it is usually a case of substituting one finite form with another (e.g., 1st person present tense singular verb is used in place of 1st person present tense plural verb; Bortolini et al., 1997). In languages such as German (Rice et al., 1997), Dutch (Wexler et al., 2004), Swedish (Hansson & Leonard, 2003), and French (Paradis &

Crago, 2001), children with DLD tend to produce infinitives when they are unsuccessful in marking tense or agreement. Interestingly, the infinitive in these languages are verbs with overt inflections rather than a bare stem (as in English). In languages such as Arabic, where infinitives do not exist, children with DLD have been observed to use the imperative - a non-finite form- in place of tensed verb forms (Abdallah & Crago, 2008). Therefore, to capture children's use of finite and non-finite verb forms in the optional infinitive stage across different languages, Paradis and Crago (2001) proposed renaming the Extended Optional Infinitive stage to the Extended Optional Default stage.

The key notion of the **Grammatical Agreement Deficit (GAD)** account (Clashen & Hansen, 1997; Clashen, 1989, 1991) is that features that enter into agreement relations and are controlled by other elements in a phase or a clause are adversely affected by DLD (Clahsen, 1989; Clahsen & Hansen, 1997). A narrower version of GAD proposes that the optional phi-features of subject-verb agreement which do not have a semantic interpretation (e.g., person, number and gender), are specifically affected by DLD (Clahsen et al., 1997; Dalalakis & Clahsen, 1999). On the other hand, the functional categories of case and tense are suggested to be intact (Clahsen et al., 1997). This hypothesis has been supported by several studies showing that subject-verb agreement is impaired in children with DLD across several languages such as German, Greek and Dutch (Clahsen et al., 1997; Clahsen & Dalalakis, 1999; Eisenbeiss et al., 2006; Rothweiler et al., 2012; Tsimpli, 2001; Verhagen & Blom, 2014). For instance, German-speaking children with DLD exhibit a substantial difficulty with producing subject-verb agreement but show comparable performance to TD peers in case marking on direct and indirect objects (Eisenbeiss et al., 2006). Greek-speaking children with DLD experience difficulties with expressing subject-verb agreement but have no difficulties with marking tense (Clahsen & Dalalakis, 1999). On the other hand, English-speaking children with DLD show difficulties with using agreement-related morphemes such as copula *be* and *do* forms and the present person singular *–s* (Rice & Wexler, 1996). However, they also show poor use of the regular past tense *–ed*. Similarly, Arabic-speaking children with DLD show difficulties with marking subject-verb agreement and show difficulties with marking tense. Arabic-speaking children with DLD seem to be better at marking subject-verb agreement than marking tense (Abdallah & Crago, 2008). Therefore, findings from English and Arabic contrast the GAD's suggestion that tense is intact in children with DLD. As opposed to the GAD's theory, few errors in subject-verb agreement are observed in children with DLD acquiring Spanish (Bedore & Leonard, 2001, 2005; note that difficulties with tense and subject-verb agreement were reported by Grinstead et al., 2013), Italian (Bortolini et al., 1997; Leonard et al., 1992) or Hebrew (Dromi et al., 1993, 1999). Generally, the GAD account cannot explain grammatical difficulties beyond agreement difficulties in children with DLD such as those related to tense production. It cannot adequately explain the less serious agreement deficits in children with DLD acquiring richly inflected languages.

Whilst the EOI and GAD theories posit that the computational system of children with DLD is intact (Marinis, 2011), the **Representation Deficit for Dependent Relationships (RDDR)** account (van der Lely, 1998; van der Lely & Stollwerck, 1997) suggests that the computational system in children with DLD is compromised. The RDDR hypothesis was originally devised to account for a group of children with DLD who have discrete difficulties with expressive and receptive grammar (referred to as Grammatical Specific Language Impairment (G-SLI) by van der Lely and colleagues). The RDDR holds that the difficulties of children with DLD are caused by a deficit in the computational grammatical system that involves the movement of constituents or features (van der Lely, 1998). Specifically, basic grammatical movement that is obligatory in neuro-typical grammar, is treated as optional in DLD grammar. Importantly, RDDR contends that the operation/rule "MOVE" is available in the grammar of children with DLD. However, its implementation is impaired i.e., optional (van der Lely, 1998). Van der Lely (1998) described an "Economy 2" principle which forces checking of unchecked features, if the target has not had its features checked, ensuring that movement is compulsory. This principle is missing from the grammar of children with DLD. Findings from English have concurred with the RDDR account. Grammatical structures involving syntactic dependencies have been reported to be problematic for English-speaking children with DLD such as agreement morphemes (Rice & Wexler, 1996), Wh questions (Marinis & Van Der Lely, 2007) and passives (Leonard et al., 2003; Marinis & Saddy, 2013). Furthermore, movement-related difficulties have also been documented in children with DLD across several languages: Hebrew (N. Friedmann & Novogrodsky, 2004; Naama Friedmann & Novogrodsky, 2011), Italian (Arosio & Guasti, 2019), Arabic (Shaalan, 2010) and French (Hamann, 2006) among others.

The RDDR was expanded into the **Computational Grammatical Complexity (CGC)** account (Marshall & van der Lely, 2006, 2007; Marshall, 2006; van der Lely, 2005). The CGC posits that children with DLD have a deficit in representing one or more of the core components of grammar: syntax, morphology and phonology. The deficit is evident in the formulation of representations of complex structures. The deficit compromises the computation of hierarchically-organized syntactic structures with long-distance dependencies (e.g., Wh object questions), the construction of inflected forms and the creation of complex syllabic structures (van der Lely, 2005). Findings from several studies have concurred with the assumptions of the CGC account. A study by van der Lely and Ullman (2001) revealed that, in contrast to younger language-matched controls, children with DLD produced regular and irregular past tense verbs at a similar rate, and their production accuracy of these forms was correlated with their frequency. It was concluded that, in TD children, irregular forms are stored in and recalled from the lexicon. In contrast, regular forms are computed via a grammatical rule, which governs attaching the –*ed* suffix to the verb stem. However, the computation of the regular past tense –*ed* rule is impaired in children with DLD. Therefore, they tend to store and retrieve both regular as well as regular verbs from the lexicon. Moreover, Marshall and van der Lely (2006) found that regular past tense verbs ending in monomorphemically legal clusters (clusters which also occur in monomorphemic words e.g., *scowled*/*cold*) were harder for children with DLD to inflect compared to verbs ending in monomorphemically illegal clusters (clusters that do not occur in monomorphemic words e.g., rushed). This effect of verb ending's phonotactics was not observed in younger TD children. The better performance of children with DLD on verbs ending with monomorphemically legal clusters, which are also more frequent, further confirmed that past tense productions in these children were not morphological rule-products. Rather, they were retrieved from the lexicon. Together, these two studies showed that the formation of morphological rules is impaired in children with DLD (Marshall & van der Lely, 2006; der Lely & Ullman, 2001). In a subsequent study, Marshall and van der Lely (2007) found that children with DLD were less successful in suffixing verb stems when the inflected verb required the production of two-consonant clusters (e.g., hugged) or three-consonant clusters (e.g., danced) compared to when no consonant clusters were needed (e.g., sewed). They were also less accurate in producing past tense verbs ending with two-consonant clusters than verbs ending with one-consonant cluster. The effects of increasing phonological complexity on regular past tense inflection accuracy were not observed in the younger language-matched TD children. Further studies have also shown that, in nonword repetition tasks, children with DLD are more adversely affected by the presence of consonant clusters relative to TD peers (Briscoe et al., 2001; Gallon et al., 2007; Leclercq et al., 2013; Munson et al., 2005). These outcomes suggest that phonological complexity poses a challenge for children with DLD.

The CGC framework has successfully accounted for the phonological, morphological and syntactic movement-related deficits in children with DLD. However, the idea of "complexity" has been described as being general (Leonard, 2014, p.263). Furthermore, the CGC does not address the mechanisms that underlie the deficits across the three domains of computational grammar: are there separate mechanisms that underlie deficits within each domain and these mechanisms are somehow connected? Or is there one mechanism that commonly underlies these deficits? The latter is less likely to be the case given the extensive amount of research emphasizing that DLD is a multifactorial, complex disorder (e.g., Bishop et al., 2006; Bishop et al., 2017). The CGC does not take into account deficits beyond computational grammar that have been observed in children with DLD such as lexical-semantic deficits (e.g., word learning) or non-linguistic deficits.

1.5.2 DLD as a processing limitation

Processing-based accounts attribute the language problems of children with DLD to a deficit in information-processing capacity or speed. These accounts differ in whether specific or general processing mechanisms are limited and whether they account for the broader language profiles of children with DLD or only focus on narrow yet seriously impaired language domains.

Some researchers have linked DLD to **limitations in working memory.** The nature of this deficit has been a subject of debate (for a review, see Archibald, 2016, 2018; Henry & Botting, 2016; Montgomery et al., 2018). Working memory is a domain-general, capacity-limited resource responsible for the short-term storage and processing of verbal and visual-spatial information or other inputs (Archibald, 2016). Baddeley's working memory model (Baddeley & Hitch, 1974; Baddeley, 2000, 2012) has been extensively applied in the DLD literature. According to this framework, working memory is a multi-component system that encompasses: 1) the central executive, a construct responsible for controlling and coordinating information processing within working memory. It has a key role in controlling its finite-capacity attentional resources (mental energy) and in allocating, sustaining, dividing and switching attention to achieve concurrent cognitive tasks; 2) the phonological loop and 3) visuospatial sketchpad which are passive storage systems responsible for temporarily storing verbal and visuospatial information, respectively; and 4) the episodic buffer which integrates and binds information across the working subsystems (e.g., the phonological loop and visuospatial sketchpad) with information in long term memory (Baddeley, 2000).

Within the DLD literature, the term verbal short-term memory (also known as phonological short-term memory) has been used to refer to the function of the phonological loop. That is the temporary storage of verbal information. Gathercole and Baddeley (1990) documented impairments in tasks of simple phonological storage (e.g., word recall and nonword repetition) in children with DLD, leading to the proposal that these children have an underlying deficit in their verbal short-term memory. This deficit may entail: 1) a noisy segmental analysis of phonological input leading to imprecise phonological representations, or 2) a limited capacity of the verbal short-term memory resulting in inadequate memory traces of stored items or 3) a quicker decay of memory traces. Gathercole and Baddeley (1990) argued that there is a causal link such that limitations in verbal short-term memory underpin the language difficulties of children with DLD. It is now clear that, in addition to reduce verbal short-term memory capacity, impaired nonword repetition in children with DLD may reflect deficits in "phonological discrimination, encoding/processing or motor production" (Leonard, 2014, p.279; see section 1.3.2).

Verbal working memory refers to the concurrent storage and processing of verbal information (Archibald & Gathercole, 2006). Verbal working memory is often assessed using complex verbal span tasks (e.g., listening span, backward digit recall) that draw on the central executive and the verbal short-term memory. An example of such tasks is the listening span (also known as competing language processing task). In this task, sets of sentences are presented and the child must judge whether the sentence is true or not; and at the same time, remember the final word of each sentence. At the end of each set, the child is asked to recall as many final-sentence words as possible. It is generally reported that children with DLD do not differ in their sentence comprehension accuracy scores from age-matched TD peers, however, they are consistently reported to recall fewer words than their age-matched peers (Archibald & Gathercole, 2006; Mainela-Arnold & Evans, 2005; Montgomery & Evans, 2009; Weismer et al., 1999). These findings support the notion that children with DLD have fewer overall attentional resources to support simultaneous verbal processing and verbal storage. That is, they present with verbal working memory deficits. Extensive evidence has shown poor performance of children with DLD in verbal working memory tasks (Gray et al., 2019; Henry & Botting, 2016; Marini et al., 2014; Montgomery et al., 2010). Not all children with DLD have verbal working memory deficits (e.g., Archibald & Griebeling, 2016; Freed et al., 2012). Nevertheless, some researchers suggest that the language impairments in children with DLD stem from a dual deficit in the processing and storage of verbal information (e.g., Archibald & Gathercole, 2006; Jackson et al., 2020; Weismer et al., 1999).

How do deficits in working memory account for the language profiles of children with DLD? It is thought that verbal short-term memory is fundamental for vocabulary acquisition in the early years (Adams & Gathercole, 1995; Adams et al., 1999; Baddeley et al., 1998; Gathercole & Baddeley, 1989; Gathercole, 2006b; Ramachandra et al., 2011). Research findings suggest that word-learning difficulties in children and adults with DLD could be attributed to a deficit in the verbal short-term memory, specifically, in the encoding stage of word learning (Bishop & Hsu, 2015; Haebig et al., 2019; Jackson et al., 2021; Leonard et al., 2019; McGregor, Licandro, et al., 2013). Having a reduced verbal short-term memory capacity will comprise the storage of the phonological information of new words. Therefore, children with DLD may require more encounters with the new words relative to their TD peers before the phonological form of the word is established. This deficit may account for the limited vocabulary of children with DLD (Baddeley et al., 1998).

A deficit in verbal short-term memory may also account for the receptive language difficulties of children with DLD especially their poor sentence comprehension. For instance, nonword repetition accuracy scores of children with and without DLD were positively correlated with their overall comprehension accuracy of sentences varying in length and grammatical complexity (Montgomery, 1995; Robertson & Joanisse, 2010). These outcomes suggested that a reduced capacity of verbal short-term memory may limit the amount of verbal information retained. As a result, lexical or grammatical details necessary for accurate comprehension will be lost, leading to compromised sentence comprehension (Montgomery et al., 2016). Montgomery and Evans (2009) found that for children with DLD, verbal short-term memory (indexed by nonword repetition) and verbal working memory (indexed by complex listening span task i.e., resource capacity/allocation) were positively correlated with their comprehension accuracy of simple sentence and complex sentence comprehension accuracy, respectively. On the other hand, neither verbal short-term memory nor verbal working memory capacity were correlated with the comprehension of simple or complex sentences. Notably, comprehending complex, non-canonical sentences requires establishing non-local dependencies which entails temporary storage and re-activation of earlier verbal input while concurrently processing incoming input. A limitation in the central executive's attention resource capacity/allocation components may hinder these processes, resulting in poor comprehension in children with DLD. The result also suggests that children with DLD require more mental effort than TD children to understand simple and complex sentences as evidenced by the correlation between verbal short-term and working memory and complex sentence comprehension in the DLD group (Montgomery & Evans, 2009).

Several studies found correlations between working memory and grammatical production (Duinmeijer et al., 2012; Marini et al., 2014; Montgomery, 2002; Shahmahmood et al., 2020), grammatical processing (Frizelle & Fletcher, 2015) and morphological learning in children with DLD (Weismer, 1996). These findings indicate that limitations in information storage or processing in working memory may underpin the impaired morpho-syntactic abilities of children with DLD. This notion is further supported by recent findings showing that working memory training is linked to improvements in the performance of French-speaking children with DLD on morpho-syntactic markers of DLD including the production of 3rd person accusative clitics (Stanford et al., 2019) and sentence repetition, with the most pronounced benefit being evident for the repetition of complex syntactic structures (Delage et al., 2021).

Much less research has examined non-verbal working memory mechanisms: visuospatial short-term memory or visuospatial working memory. The results so far have been variable: while some researchers identified substantial impairments in the visual-spatial domains in children with DLD, others reported intact abilities (for a meta-analysis see Vugs et al., 2013). Though inconsistent, evidence of impaired visuospatial working memory mechanisms may suggest a domain-general working memory deficit in children with DLD. However, future research is necessary to confirm this notion.

The working memory accounts describe how processing limitations cause broader language deficits in children with DLD. On the other hand, the **Surface Account** (**SA**, Leonard, 1989; Leonard et al., 1992, 1997) employs processing limitations to explain these children's specific and extraordinary difficulties with grammatical morphology. The SA posits that children with

DLD have general processing capacity limitations which profoundly impact their ability to perceive and hypothesize the function of grammatical morphemes of low perceptual (surface) salience (Leonard, 1989). These are grammatical inflections that occur as word-final consonants or unstressed (weak) non-final syllables with non-significant vowel lengthening. Hence, these morphemes have low phonetic substance, are brief in duration and are less perceptually salient. According to the SA, children with DLD are capable of perceiving grammatical morphemes of low acoustic salience. However, when these forms have a morphological function, their processing is challenging for children with DLD as it entails detecting the morphemes, discovering their morphological role and placing them in a morphological paradigm (Leonard et al., 1997a). These operations occur in real-time while the child is also hearing the rest of the unfolding sentence. The demands of these operations and the low salience of the grammatical morphemes will tax the already-limited processing capacity of children with DLD resulting in the inadequate/incomplete processing of these morphemes (Leonard et al., 1997a). As a result, children with DLD require a greater number of exposures for these grammatical morphemes to become incorporated into their grammar.

The SA account can explain why English-speaking children with DLD produce regular past tense –*ed* and the third person singular-*s* (low acoustically salient inflections) with significantly lower accuracy than present tense –*ing* (Rice & Wexler, 1996). The SA further proposes that children with DLD hypothesize the grammatical roles of morphemes following the same developmental order as TD children and that morphemes with similar phonetic features may differ in their acquisition age.

The SA can also explain why children with DLD acquiring richly inflected languages, where tense and agreement inflections are more acoustically salient (e.g., syllabic), have fewer difficulties with verb morphology relative to their English-speaking counterparts (Bortolini et al., 1997; Dromi et al., 1993; Leonard, 2014b; Leonard, McGregor, et al., 1992). For instance, articles and object clitics (which have low acoustic salience) are profoundly impaired in Italian and Spanish -speaking children with DLD (Bedore & Leonard, 2001; Bortolini et al., 1997). Although the SA can account for some distinctive morphological profiles of children with DLD within the same language or between languages, it is not comprehensive and does not explain difficulties that children with DLD show in areas other than grammar (e.g., semantics, pragmatics, non-linguistic deficits).

Another processing-based account is the **Morphological Richness Account** (**MRA**; Bishop et al., 2000; Dromi et al., 1999; Leonard et al., 1987; Lukács et al., 2009). According to this framework, the morphological deficits observed in children with DLD result from an interaction between limitations in processing capacity and the characteristics of the particular grammatical system of the language to be learned (Lukács et al., 2009). The MRA presupposes that children with DLD dedicate their limited processing resources to the most dominant grammatical cues of the linguistic input, leaving few resources for processing information conveyed by other grammatical means. English has a sparse morphological system and bare-stem forms are frequent. Therefore, English-speaking children with DLD will devote their resources to information carried by word order. The remaining resources will be insufficient to process inflectional morphology (e.g., tense and agreement).

On the other hand, in languages with rich morphology (e.g., where nouns, verbs and adjectives are inflected) such as Arabic, Italian, Hebrew or Hungarian, children with DLD will dedicate their restricted resources to the processing of grammatical inflections. Hence, children with DLD acquiring richly inflected languages are expected to have fewer grammatical morphology difficulties than those acquiring languages with sparse morphology such as English (Leonard, 2014, p.295). The MRA also posits that the rich morphology advantage is conditional on the number of grammatical functions encoded by the inflections. For instance, processing inflections that encode several dimensions (e.g., verb tense, number, gender and person agreement) entails the simultaneous retention, hypothesis and retrieval of each dimension. Morphemes reflecting 4 or more grammatical dimensions will tax the resources of children with DLD. Accordingly, such forms will be incompletely processed and require additional encounters before they are established within the children's grammar (Dromi et al., 1999).

When children with DLD make errors, it is suggested that substitute inflection will share most of the dimensions with the target inflection. This would be resembled by "near misses" errors where the substitute and the target inflections differ in one dimension (e.g., correct tense, gender and number agreement but wrong person agreement). These types of errors indicate that children with DLD have considerable knowledge of the grammatical functions of the inflections. Lastly, the MRA posits that when near-miss errors do not occur, the substitute inflections tend to be highly frequent forms in the language. Evidence supporting the MRA comes from cross-linguistic investigations of verb tense and agreement morphology use in

Hebrew (Dromi et al., 1993, 1999), Spanish (Bedore & Leonard, 2001), Hungarian (Lukács et al., 2009), Arabic (Abdallah & Cargo, 2008), and French-speaking children with DLD (Thordardottir & Namazi, 2007). The gap between children with DLD acquiring these languages and their TD peers was not as large as reported for English.

To summarize, this section reviewed some of the most influential accounts of DLD. Of the linguistic accounts, the Extended Optional Infinitive theory explained the verb morphology deficits in children with DLD in light of a maturational delay (Rich & Wexler., 1996), whereas the Grammatical Agreement Deficit model suggested that these difficulties stem from difficulties in establishing in subject-verb agreement relations (Clahsen et al., 1997; Dalalakis & Clahsen, 1999). The Representational Deficit for Dependent Relationships and Computational Grammatical Complexity accounts addressed the morphological, syntactic and phonological deficits in children with DLD. These accounts proposed a deficit in the linguistic computational system of children with DLD which compromises the computation and construction of hierarchically organized syntactic structures with long-distance dependencies, the use of morphological rules and the creation of phonologically complex structures (van der Lely, 2005). As for the processing-based accounts, the Surface account proposed that the morphological deficits in children with DLD result from an interaction between reduced phonetic saliency of grammatical morphemes and the limited processing capacity in these children (Leonard et al., 1997a). On the other hand, the morphological richness account attributed these morphological difficulties to an interaction between the richness of the grammatical system of a given language and the processing limitations in children with DLD (Lukács et al., 2009). Whilst grammar is a core area of deficit in children with DLD, these children also show a broad range of deficits affecting different language domains (semantics, pragmatics, phonology). Working memory accounts offered a link between working memory limitations and the broader language deficits in children with DLD (e.g., word-learning, vocabulary, nonword repetition, and the processing, comprehension, and production of grammatical structures).

One way to enhance our understanding of DLD is to use the findings of cross-linguistic investigations to evaluate the different theories of DLD. Identifying areas that are consistently impaired in children with DLD speaking different languages may pinpoint the nature of mechanisms that are commonly impaired in children with DLD across languages. Moreover, defining the language-specific manifestations of DLD may help refine current DLD theories and re-direct efforts to exploring new theoretical accounts to cover the cross-linguistic variations of the disorder. In this thesis, the extent to which the current DLD theories can account for the language profiles of children with DLD acquiring Arabic will be explored (see section 5.2).

1.6 Aims of this thesis

DLD is associated with long-standing functional limitations that extend beyond language itself (e.g., Conti-Ramsden et al., 2019; Durkin & Conti-Ramsden, 2007; Lindsay & Dockrell, 2012; St Clair et al., 2011). Consequently, DLD may negatively affect the quality of life of the affected individuals (Eadie et al., 2018; Nicola & Watter, 2015). Therefore, the importance of accurate and early identification and access to intervention for children with DLD cannot be overemphasized. Indeed, research shows that early identification and access to targeted intervention for DLD during childhood are associated with more positive outcomes and areas of strength later in life (e.g., Conti-Ramsden et al., 2018; Winstanley et al., 2018). Unfortunately, identifying Arabic-speaking children with DLD poses a serious challenge to SLTs in Palestine and other Arabic-speaking countries. In general, there is a shortage of appropriate Arabic language assessments. Hence, Arabic-speaking children with DLD are often at risk of being unrecognized and unsupported. To address this issue, the main goal of this thesis was to investigate potential clinical markers of DLD in Arabic.

Although limited, the available literature has identified statistically significant group differences between Arabic-speaking children with DLD and same-age TD peers on verb morphology production, nonword repetition and sentence repetition. Notably, these findings are only preliminary and replication of these outcomes is necessary to ensure their validity and to better define the linguistic deficits associated with DLD in Arabic. The observed group differences between Arabic-speaking children with and without DLD suggest that these verb morphology production, nonword repetition and sentence repetition are problematic for children with DLD acquiring Arabic. However, group differences are not sufficient to determine the potential of these measures as clinical markers. This is because group comparisons do not consider the heterogeneity of profiles of children with DLD. Hence, to assess the potential of these measures as clinical markers of DLD in Arabic, it is necessary to assess their diagnostic power in differentiating between children with and without DLD.

The main goal of this thesis is to define the linguistic deficits in Arabic-speaking children with DLD. The thesis investigates the performance of Palestinian Arabic-speaking children with and without DLD on tasks that assess verb morphology production, nonword repetition and sentence repetition. The thesis further examines the diagnostic accuracy of nonword and sentence repetition (i.e., cross-linguistic markers) in identifying DLD in Arabic.

Identifying clinical markers of DLD in Arabic would inform SLTs of the linguistic difficulties that should be considered during the language assessment of Arabic-speaking children with DLD. In turn, this may contribute to enhancing the clinical practice of identifying DLD in Arabic-speaking children. Examining the DLD profile in Arabic will extend our knowledge of the cross-linguistic manifestation of DLD. It will also expand our understanding of the core features of the disorder. Arabic is a Semitic language with a complex and rich grammatical system, making it an interesting exemplar to investigate theoretical frameworks of DLD that were primarily developed based on findings from English and/or Indo-European languages. The clinical marker tasks used in this thesis are theoretically based. Verb morphology taps into knowledge of language rules, whereas nonword repetition and sentence repetition are language processing measures. Hence, findings from Arabic could contribute to the debate of whether the language difficulties of children with DLD are a consequence of deficits in linguistic knowledge, processing capacity/speed or both.

In the following chapters, three studies are presented. Using a verb elicitation task, study 1 characterizes the expressive verb morphology deficits in Arabic-speaking children with DLD relative to same-age TD peers. Study 2 and Study 3 compare the performance of 4 to 6-year-old Arabic-speaking children with DLD and age-matched TD children on nonword repetition and sentence repetition, respectively. Studies 2 and 3 further examine the diagnostic accuracy of nonword and sentence repetition in identifying Arabic-speaking children with DLD.

Chapter 2: Expressive verb morphology deficits in Arabic-speaking children with Developmental Language Disorder

Abstract

Purpose: This study investigated the production of tense and subject-verb agreement by Palestinian Arabic–speaking children with developmental language disorder (DLD) in comparison to their typically developing (TD) peers in terms of (a) performance accuracy and (b) error patterns.

Method: Participants were 14 children with DLD aged 4;0–7;10 years and 32 TD children aged 3;0–8;0 years matched on nonverbal abilities. Children were asked to complete a picture-based verb elicitation task. The task was designed to measure the production accuracy of tense (present and past tense verbs), subject-verb gender agreement (masculine and feminine verbs), subject-verb number agreement (singular and plural verbs) and subject-verb person agreement (third-person person verbs).

Results: The DLD group scored significantly lower than the TD group on the verb elicitation task. The DLD group was considerably less accurate than the TD group in marking tense, specifically present tense. In terms of subject-verb gender agreement, the DLD group were less accurate than the TD group in using feminine verbs but showed comparable accuracy in using masculine verbs. In terms of subject-verb number agreement, the DLD and TD groups did not differ significantly in producing singular and plural verbs. The DLD group marked third person agreement with a high level of accuracy. The DLD group made substantially more tense and agreement errors than the TD group but both groups showed similar error types.

Conclusions: Expressive verb morphology was impaired in Palestinian Arabic–speaking children with DLD. Not all verb forms were problematic for the DLD group. The production of present tense and feminine verbs was specifically challenging for the DLD group. The DLD group used past tense and third-person verbs with a high levels of accuracy and did not have difficulties with using singular and plural verbs. In both groups, tense and agreement errors resembled the use structurally simpler verb form in place of structurally more complex form. The results are discussed in relation to the structural characteristics of Arabic.

2.1 Introduction

Children with developmental language disorder (DLD) exhibit morphosyntactic deficits often related to the use of tense and subject–verb agreement inflections (for a review, see Leonard, 2014a). The production of verb inflections, such as past tense *–ed*, present third-person singular *–s*, auxiliary and copula *be*, and auxiliary *do* forms, has been reported as problematic for English-speaking children with DLD (Leonard & Kueser, 2019; Rice & Wexler, 1996), and verb morphology difficulties are considered to be a clinical marker of DLD in English (e.g., Bedore & Leonard, 1998; Conti-Ramsden et al., 2001).

Cross-linguistic research shows that verb morphology is differentially impaired across languages. For example, children with DLD acquiring Germanic languages(e.g., Dutch, Swedish, German and Norwegian) are reported to be less accurate than their typically developing (TD) peers in marking tense and agreement and especially past tense marking, yet their accuracy of using verb inflections is higher than that reported for English-speaking children with DLD (for a review, see Krok & Leonard, 2015). For children with DLD acquiring Romance languages, such as Spanish and Italian, verb morphology is not as problematic; the main difficulties seem to be using function words, such as articles, and unstressed direct object pronouns (e.g., Bedore & Leonard, 2001; Bortolini et al., 1997). Hebrew-speaking children with DLD have difficulties marking agreement in past tense, but not in marking present tense (e.g., Dromi et al., 1999; Leonard & Dromi, 1994).

In summary, verb morphology deficits vary between languages, especially when languages are typologically different. Therefore, studies of grammatical morphology should be language specific. This study aims to extend this line of research by characterizing verb morphology deficits in children with DLD acquiring Palestinian Arabic (PA).

2.1.1 The verb paradigm in Palestinian Arabic

In the Arab world, Modern Standard Arabic (MSA) is the language of literacy tasks and is used in formal contexts, such as news. A unique feature of the Arabic language is diglossia (Haeri, 2000). Each Arab country has a distinctive dialect of Arabic that is used for everyday social interactions. This article focuses on the colloquial dialect of Palestine: PA.

MSA and its dialectal varieties are characterized by their nonconcatenating templatic morphology that is based on a system of roots and patterns (McCarthy & Prince, 1988; Ryding, 2005).The root is an invariable sequence of three to five consonants, and it carries lexical meaning. The pattern consists of one or more vowels, and it carries grammatical meaning. Patterns (vocalic infixes) are discontinuously inserted within the consonantal root to form words and stems (Tucker, 2011).In PA, for example, the root *drs* denotes a meaning of "studying." By shifting different patterns and consonantal affixes around this root, we can derive different words such as *daras* "he studied," *madrasa* "school," or *dars* "lesson." MSA is null-subject language, and verbs are conjugated to represent different grammatical categories, including tense and aspect (past/present and perfective/imperfective), number (singular, dual, and plural), person (first, second, and third), gender (masculine and feminine), mood (indicative, subjunctive, jussive, energetic, and imperative), and voice (passive/active; Benmamoun, 2000).

Three verb forms are distinguished by traditional Arabic grammarians: perfective, imperfective, and imperative verbs. There is debate as to whether Arabic verbs are considered to be tense specific where perfective and imperfective verbs refer to past and non-past actions, respectively, or aspect specific where perfective and imperfective verbs refer to complete or noncomplete actions (for a review, see Ouali, 2018). According to Ouali (2018), there seems to be a consensus in recent literature that Arabic is tense specific language. Table 2.1 presents the verb paradigm in PA.

2.1.1.1 Past Tense

In PA, the perfective verb is used to refer to past and completed actions (Abu-Ghazaleh, 1983, p. 125) and will be referred to as past tense. Past tense is an abstract morpheme, that is, not realized by an overt affix (Benmamoun, 2000). The past tense verb consists of a stem (root + vocalic infixes) and takes only suffixes that denote subject–verb agreement (Benmamoun, 2000). The suffix is a discontinuous unit that simultaneously reflects agreement for person, gender, and number. For example, the suffix *–ti* in *daras**ti*** "you studied" denotes agreement for a second-person feminine singular subject (form 4; see Table 2.1). The third-person masculine singular *daras* "he studied" is unmarked, that is, it does not take any suffixes (form 6; see Table 2.1). It is homonymous with the past tense verb stem. It is important to note here that, unlike MSA, PA verb paradigm is smaller as the subject–verb number agreement has no dual category and the plural agreement suffix *–u* has no gender distinction (e.g., forms 8, 16, and 19 in Table 2.1; Jarrar et al., 2014).

Table 2.1. *Verb paradigm in Palestinian Arabic for the root d-r-s (studying)*

			Past Tense			Present Tense			Imperative		
Person	Number	Gender	Form	Affixes	Verb + Affixes	Form	Affixes	Verb + Affixes	Form	Affixes	Verb + Affixes
1st	Singular	Neutral*	1	*-it*	*daras**it***	9	*b-a-*	***ba**drus*			
1st	Plural	Neutral	2	*-na*	*daras**na***	10	*b-ni-*	***bni**drus*			
2nd	Singular	Masculine	3	*-it*	*daras**it***	11	*b-ti-*	***bti**drus*	17	*ʔi-*	***ʔi**drus*
2nd	Singular	Feminine	4	*-ti*	*daras**ti***	12	*b-ti--i*	***bti**drus**i***	18	*ʔi--i*	***ʔi**drus**i***
2nd	Plural	Neutral	5	*-tu*	*daras**tu***	13	*b-ti--u*	***bti**drus**u***	19	*ʔi--u*	***ʔi**drus**u***
3rd	Singular	Masculine	6	*Ø*	*daras*	14	*b-yi-*	***byi**drus*			
3rd	Singular	Feminine	7	*-at*	*dars**at***	15	*b-ti-*	***bti**drus*			
3rd	Plural	Neutral	8	*-u*	*daras**u***	16	*b-yi--u*	***byi**drus**u***			

Note. *The gender category "neutral" indicates that the affix attached to the verb has no gender distinction.

2.1.1.2 Present Tense

The imperfective verb is used to refer to an ongoing activity, which could be in the present, past, or the future time (Benmamoun, 2000). In PA, the imperfective verb has three moods: indicative, subjunctive, and imperative (Abu-Ghazaleh, 1983; Kimary Shahin, 2007).In this section, we focus on its indicative mood, which occurs in sentences with present tense interpretation (henceforth, present tense).The present tense is composed of a stem *drus* (root + vocalic affix) with its subject–verb agreement being realized by a prefix or a combination of a prefix and a suffix (circumfix morpheme).

In the PA present tense verb, the temporal information is carried by the present progressive clitic *b*–, which attaches to the prefix (Abu-Ghazaleh, 1983; Jarrar et al., 2014; Shahin, 2007). Person agreement is mainly realized by the prefix. Gender is also realized by the prefix, except for the second-person singular feminine where gender is expressed by the suffix –*i* (form 12; see Table 2.1). Plural number agreement is realized by the suffix –*u*, except for the first person where the number is realized by the prefix *bni*– (Benmamoun, 2000;form 10 in Table 2.1). More than one subject–verb agreement feature can be realized by one prefix. For instance, the prefix *byi*– in ***byi****drus* "he is studying" indicates a third-person masculine subject (person and gender agreement). In other instances, the subject–verb agreement features are realized by a circumfix affix, an un-analyzable unit of a prefix and a suffix. An example is the circumfix *byi*—*u* in ***byi****drus****u*** "they are studying," where it denotes third-person plural agreement (no gender distinction).

Finally, it is clear that the verb forms we described differ from each other in terms of markedness, that is, the morphological realization of grammatical categories (e.g., Corbett, 1991, 2000; Leech, 2006). In Arabic subject–verb agreement, contrasts in number agreement (singular vs. plural) and gender agreement (masculine vs. feminine) are asymmetrical in terms of their morphological realization. That is, one member of the contrast is overtly coded by an affix and therefore is "marked," whereas the other member has no overt coding (zero affixes) and is therefore considered as an unmarked form. For example, if we look at the opposition of singular–plural in number agreement, the singular verb is not overtly realized by any affixes (e.g., *daras* "he studied"), whereas the plural verb is realized by the affix –*u* (e.g., *daras****u*** "they studied"). The singular verb is therefore considered as the unmarked/default form, while the plural is the marked form. The same applies to gender agreement (only in past tense) where the feminine verb (e.g., *daras**at*** "she studied") is marked whereas the masculine form (e.g., *daras* "he studied") is unmarked.

2.1.1.3 The Imperative

The imperative verb and the present tense verb share the same stem (e.g., see forms 14 and 17 in Table 2.1). Unlike the present tense verb, the imperative lacks the present progressive clitic *b*– and the initial prefix, which indicates person and gender agreement. The imperative only occurs in the second person, yet the person agreement feature is unmarked (Al-Aqarbeh, 2011). Although PA has a prefix for second-person present tense verbs (e.g., *bti*– or *bit*–), this prefix is dropped in the imperative verb. Gender and number agreement of the imperative verb is denoted by the suffix (see forms 17–19 in Table 2.1).

There is little agreement on what is the default tense form in Arabic. While some researchers argue that the default form is the imperative (Abdallah & Crago, 2008; Morsi, 2009; Omar, 1973; Qasem & Sircar, 2017), others identify it as the imperfective verb stem (Aljenaie, 2010; Benmamoun, 1999).Fahim (2017) states that the default verb can take more than one form, including the imperative, subjunctive, or a variant of the imperfective verb stem. The imperative does not have a time reference, and it is considered nonfinite (Ryding, 2005). Similarly, Benmamoun (2000) states that the imperfective verb occurs in different contexts, such as sentences with past, present, or future interpretation, as well as in embedded nonfinite sentences. This evidence clearly shows that the imperfective does not morphologically carry any temporal or aspectual information (Benmamoun,1999, 2000). Although there are slight morphological differences between the two forms (primarily in their prefixes), they are very similar, which could be the cause of inconsistency among studies. By removing the affixes of the imperative (e.g., form 17; see Table 2.1) and imperfective indicative (present tense; e.g., form 11; see Table 2.1), it can be seen that both forms share the same stem, suggesting that the imperative is derived from the imperfective verb (Benmamoun, 1999; Shahin, 2010; Soltan, 2007).

2.1.2 Typical and Atypical Verb Morphology Acquisition in Arabic

Few studies have examined typical language acquisition in Arabic. In a longitudinal study, Omar (1973) described the acquisition of phonology, syntax, and morphology in 37 Egyptian Arabic–speaking children aged 6 months to 15 years. The study reported that children started using verbal agreement morphology around the age of 2;3 (years;months). Masculine and singular verbs emerged earlier than feminine and plural verbs, respectively. Omar further observed that, in the early stages of verb production, Egyptian Arabic–speaking children predominantly used the singular masculine verb as the default verb agreement category.

In a longitudinal study on PA, Abdu and Abdu (1986) documented the milestones of lexical development of their two children from around the age of 1 year up until 6 years. Their data on the acquisition of verbs indicated a certain order in which verb forms emerge in PA. In line with Omar's (1973) findings on verb agreement, masculine and singular verbs emerged earlier than feminine and plural verbs, respectively. Additionally, third-person verbs appeared before first-person verbs, with second-person verbs appearing last. This order was limited to past tense verbs, as no particular order was noted for present tense verbs.

Similar findings are reported by Aljenaie (2001), who followed the development of verb tense and agreement in four Kuwaiti Arabic–speaking children aged 1;17–2;6 for 6 months using spontaneous speech, elicited production, and imitation tasks. All four children began using present and past tense verbs at the age of 2;0 years. However, the order in which these forms emerged in the children's language could not be determined due to the variability in the data. Agreement marking emerged in the following pattern: masculine verbs appeared before feminine verbs, singular verbs appeared before plural verbs, while first-person verbs appeared first, followed by third-person and second-person verbs. Furthermore, Aljenaie (2001) noted that children showed a tendency to use unmarked forms in contexts where verb infections were required. In past tense contexts, the unmarked form was the third-person masculine singular, whereas in the present tense context, the unmarked form was described as being either the imperative masculine verb or a form that was homophonous to the stem of the target verb (Aljenaie, 2001).The use of the imperative was also noted in the language of Yemini- (Qasem & Sircar, 2017) and Egyptian Arabic–speaking TD toddlers (Omar, 1973).

In another longitudinal study, Aljenaie (2010) examined spontaneous language samples of three Kuwaiti Arabic–speaking children aged 1;8–3;1. An analysis of agreement errors revealed that masculine verbs were used to substitute feminine verbs. These findings suggest that children show a preference for the less marked, more neutral masculine form over the feminine counterpart, which is strongly and consistently marked by inflections for gender (Aljenaie, 2010, p. 852). Regarding tense errors, Kuwaiti Arabic–speaking children used the imperfective bare verb, a nonfinite form, in place of fully inflected verbs (Aljenaie, 2001, 2010).

This supports the view that the imperfective verb stem is most likely the default tense form in Arabic (Benmamoun, 1999, 2000).

Basaffar and Safi (2012) investigated the developmental patterns of tense and verb agreement in 2- to 4-year-old Hijazi Arabic–speaking children. Using experimental tasks alongside a spontaneous language sample analysis, they replicated the findings reported by Aljenaie (2001). Basaffar and Safi (2012) concluded that children produced present and imperative forms with higher accuracy than past and future forms. However, Basaffar and Safi (2012) did not report any descriptive statistics, statistical and error analyses, or any clear guidelines for the scoring system of the children/s responses. This limits the interpretation and generalizability of their findings.

Research into morphosyntactic difficulties in Arabic-speaking children with DLD has been scarce. Drawing on her dissertation data from 2002 (Abdallah, 2002), Abdallah and Crago (2008) analysed speech samples obtained from Hijazi Arabic–speaking children with DLD aged 4;0–5;3 years. Children with DLD were less accurate than their age- and language-matched peers in marking tense in general. The DLD group scored significantly higher for past tense than for present tense forms, which suggests that these children's difficulties with tense were more pronounced in present tense verbs. Not all subject-verb agreement categories were problematic for the DLD group. Present tense, feminine, and third-person verbs, which were structurally more complex, were more problematic than unmarked verb forms, such as past tense and masculine verb forms (Abdallah & Crago, 2008). Importantly, both TD children and children with DLD used the imperative in place of the target tense forms. In a few instances, children with DLD used an incorrect tense form (e.g., present tense for past tense). When agreement errors occurred, one agreement feature was affected (e.g., third-person masculine singular replaced third-person feminine singular). Abdallah and Crago (2008) characterized agreement errors as follows: singular verbs were used in place of plural verbs, masculine verbs were used in place of feminine verbs, and first-person verbs were used in place of third-person verbs.

Morsi (2009) found that Egyptian Arabic–speaking, 6-year-old children with DLD were less accurate than their age- and language-matched peers in the production of verbal tense and agreement, with tense being more challenging than agreement. Morsi stated that, for the DLD group, present tense production was more difficult than past tense production, and the imperative was used as the default form when tense errors occurred.

Drawing on her dissertation data from 2005 (Fahim, 2005), Fahim (2017) analyzed spontaneous speech samples of three Egyptian Arabic–speaking children with DLD aged 3;1–4;6 and six TD children aged 1;0–4;0.Fahim concluded that only subject-verb agreement marking was impaired in Egyptian-speaking children with DLD, while tense marking was less affected (based on their high accuracy in past tense marking). Furthermore, Fahim identified three tense error patterns in the speech of children with and without DLD. The first error pattern involved the use of a default verb form in place of the tense verb. The form was described to resemble the imperative or the subjunctive. The second error pattern involved a verb with the correct tense but incorrect agreement. The third error involved the production of nonadult target forms (pseudowords) in place of the target verbs.

A different pattern of results emerged in Shaalan's (2010) dissertation, which reported that Qatari Arabic–speaking children with DLD (aged 4;6–9;4) were less accurate in producing tense and agreement inflections than TD children. Specifically, past tense was more problematic than present tense for the DLD group. Shaalan (2010) stressed that these results were preliminary, as they were only based on a few items (N = 12) and noted that further research was required.

The results of the Arabic studies have generally determined tense and verb agreement aspects that are challenging for children with DLD. There is little agreement among the studies on which aspect of verb morphology is more problematic for children with DLD: tense or agreement. Also, it is inconclusive what the default form in Arabic is as both the imperative and the imperfective bare verb forms have been suggested. These questions require further investigation. The different findings may be attributed to several reasons. First, the low participant number (N = 3) in Fahim's (2017) and Morsi's (2009) studies does not allow for the generalization of their results. Second, there were methodological differences in terms of task used: Abdallah and Crago's (2008) and Fahim's (2017) studies analyzed speech samples, whereas Morsi (2009) and Shaalan (2010) used a structured elicitation task for the target verb inflections. This could have resulted in differences in the number and type of verb inflections included in the analyses.

2.1.3 The present study

This study aims to extend previous Arabic studies by conducting a systematic investigation of verb morphology use by children with and without DLD acquiring PA. Determining which verb forms are potential linguistic markers of DLD in PA would inform and enhance the current assessment practices of DLD in Palestine. Furthermore, data from Arabic children with DLD could be used to examine the assumptions of theoretical accounts of DLD and provide insights into possible underlying mechanisms of the disorder (see section 5.2.1) . This study examines the production of tense and subject–verb agreement in PA-speaking children with DLD, as compared with TD children, by investigating (a) the production accuracy and (b) error patterns of verb tense and agreement marking.

We predict that, compared to TD children, children with DLD will achieve lower overall accuracy on the verb elicitation task. Children with DLD will have more difficulties using marked verb forms compared to less marked ones. Specifically, the use of present tense verbs is expected to be more challenging than past tense verbs (Abdallah & Crago, 2008; Fahim, 2005; Morsi, 2009). Feminine and plural verbs are predicted to be more problematic than masculine and singular verb forms (Abdallah & Crago, 2008). Children with and without DLD are predicted to use the imperative verb (Abdallah & Crago, 2008; Fahim, 2017; Morsi, 2009) or the imperfective bare verb as tense default forms (Aljenaie, 2010; Benmamoun, 1999). Finally, children with and without DLD will use less marked verbs (masculine and singular verbs) as default agreement forms in place of more marked, feminine, and plural verbs (Abdallah & Crago, 2008; Aljenaie, 2010).

2.2 Methods

2.2.1 Participants

The study was approved by the Research Ethics Committee at University of Reading. Sixty-four PA-speaking children were recruited: 14 children with DLD (10 boys), aged between 48 and 94 months, with a mean age of 66 months (*SD* = 15.47), and 32 TD children (19 boys), aged between 36 and 96 months, with a mean age of 62 months (*SD* = 16.88). The groups did not differ significantly on chronological age, $t(44) = 0.83$, $p = .413$, $d = 0.27$. The TD and DLD groups were matched on nonverbal cognitive abilities as measured by raw score on the Colored Progressive Matrices (Raven, 2007), as this test is not standardized on PA-speaking children, $t(42) = -0.81$, $p = .423$, $d = 0.26$, variance ratio = 1.11. Table 2.2 summarizes the raw scores of the two groups on several background measures. See Appendix A for individual scores.

Table 2.2. *A summary of the demographic characteristics, developmental milestones and background measures for the TD and DLD groups*

	Group	
	TD N=32	DLD N=14
Demographic Characteristics	% (*N*)	
Mother's education		
Primary school	9.38 (3)	14.29 (2)
High school	31.25 (10)	28.57 (4)
University/college degree	46.87 (15)	35.74 (5)
Postgraduate degree	12.5 (4)	21.43 (3)
Working mother	39.47 (15)	50 (7)
Family history of communication disorders	6.25 (2)	42.56 (6)*
Developmental Milestones Age in months	***Mean (SD) Range***	
First word	12.69 (2.46) 9 - 18	24.64 (6.65)* 15 - 36
Follow simple commands	17.59(3.44) 12 - 24	20.14 (5.95) 12 - 36
walking	12.66(1.45) 10 - 15	12.5(1.7) 10 -16
Background measures raw scores	***Mean (SD) Range***	
MPU	4.91 (1.24) 2.41 - 7.61	3.58 (1.04)*** 2.19 - 6.27
A-QU-LITMUS-NWRT (Out of 30)	26.84 (4.34) 16 - 30	15.57 (4.13)** 9 - 23
CPM (Out of 36)	16.67 (4.39) 8 - 23	15.5 (4.62) 9 – 23

Note. TD = Typically Developing. DLD = Developmental Language Disorder. SD = standard deviation. MPU = Mean Morpheme per Utterance. A-QU-LITMUS-NWRT = Arabic version of a Quasi-Universal Nonword Repetition Test. CPM = Colored Progressive Matrices.

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

Children with DLD were recruited through four private speech therapy clinics located in Ramallah city. They were previously diagnosed with DLD by qualified speech and language therapists (SLTs) who used nonstandardized language assessment tasks. Based on a screening of clinical reports, all children in this group had primary language deficits, had no obvious nonverbal difficulties, used speech as their primary means of communication, and had no diagnoses of any speech disorder interfering with intelligibility. All children were receiving language intervention services at the time of the study. The TD control children were recruited through one daycare, two kindergartens, and one school in Ramallah city and had no reported history of language delay/impairment and demonstrated age-appropriate language skills as determined by parental/teachers' reports. Parents completed a questionnaire that included questions about demographics (e.g., maternal education), child's health and general development, language acquisition milestones, and family history of language difficulties. The questionnaire was used to ensure that all children were monolingual Arabic speakers and had no evidence or reported history of hearing loss, cognitive and/or neurological impairments, speech motor disorders, and diagnoses of other developmental disorders (e.g., autism). Based on questionnaire results, alongside teacher reports, four children did not meet the eligibility criteria for the TD group and were not tested for the study. Groups did not differ significantly in maternal education: $\chi^2(3, N = 46) = 1.03$, $p = .793$. Children with DLD had a significantly higher frequency of family history of communication disorders, $\chi^2(1, N = 46) = 6.72$, $p < .001$, and produced their first words significantly later, $t(14.57) = 6.53$, $p < .001$, $d = 2.39$. See Table 2.2 for details.

Because the diagnosis of DLD in Palestine is based on informal assessments, scores on standardized language assessments were not available. Two nonstandardized tasks were used to verify that children with DLD had language skills that were considerably below the level expected for their chronological age.

1. Spontaneous narratives of 100 utterances were elicited using a wordless picture book, Frog, Where Are You? (Mayer, 1969), to calculate the mean morpheme per utterance (MPU). MPU is equivalent to the mean length of utterance (Brown, 1973) in English. MPU is a measure of grammatical development and takes into account the highly synthetic nature and rich morphology of Semitic languages (Dromi & Berman, 1982). MPU is calculated by dividing the total number of morphemes by the total number of utterances produced in the narrative task. We followed the guidelines of counting

Arabic morphemes that were developed by Shaalan and Khater (2006). These guidelines were adapted from the MPU calculation rules in Hebrew (Dromi & Berman, 1982). Previous studies on Arabic (Abdallah & Crago, 2008; Shaalan, 2010) have also used this measure to confirm the presence of developmental language impairment.

2. The Arabic version of a Quasi-Universal Nonword Repetition Test (A-QU-LITMUS-NWRT; dos Santos et al., n.d; for a description of the test, see Abi-Aad & Atallah, 2020) was administered. The test included 30 nonwords and was scored using binary scoring approach (correct/incorrect), with the maximum score being 30. The task was found to be problematic for children with DLD whose first language (L1) is Lebanese Arabic (Abi-Aad & Atallah, 2020).The task was also documented to have good diagnostic accuracy in identifying Palestinian children at risk of DLD (Taha & Chondrogianni, 2017).

The mean MPU score for the DLD group was significantly lower than the TD group: $t(44) = -3.51$, $p < .001$, $d = 1.23$. The DLD scored substantially below the TD group on the A-QU-LITMUS-NWRT: $t(44) = -8.22$, $p < .001$, $d = 2.63$. Norms for these two measures are not established in Arabic. Therefore, mean raw scores are reported (see Table 2.2).

2.2.2 Verb Elicitation Task

An elicitation task was developed to test children's production of tense: (a) present tense and (b) past tense, subject-verb gender agreement: (c) feminine verbs and (d) masculine verbs, subject-verb number agreement: (e) singular verbs and (f)plural verbs as well as person agreement: (g) third-person verbs.

Seventy-two pictures were divided into 30 pairs of experimental items and 12 filler items (singular and plural noun pairs). Each pair of the experimental items consisted of the present and past tense forms of the target verb. The experimental items were categorized into eight paired items for masculine singular verb forms, seven paired items for feminine singular verb forms, and 15 paired items for plural verb forms. Because present tense inflections vary in stress assignment, 50% of the present tense verbs had a stressed tense prefix, and 50% had an unstressed tense prefix (see Appendix B for test items).

Each verb was represented by a pair of pictures showing a sequence of events that the child was asked to describe. The first photograph depicted a person or a group performing an activity, and the second photograph depicted the same person or group having finished the activity. The test items depicted actions from familiar daily routines. The task was piloted with

10 TD children aged between 40 and 67 months, with a mean age of 58 months (*SD* = 9.36), to ascertain that children of this age could easily identify the verbs in the photographs. Results showed that 96.38% (SD = 8.21) of the children were able to correctly name the target verbs.

2.2.3 Procedure

Children were assessed individually in a quiet room within their nursery, school, or speech and language therapy clinic. All assessments were conducted in one session by the first author (a qualified Arabic-speaking SLT). Each session lasted approximately 1 hr and was audio-recorded using a Sony ICD-PX370 digital voice recorder. The tasks were administered in the following order: Coloured Progressive Matrices, narrative task, A-QU-LITMUS-NWRT task, and verb elicitation task. Four practice items were given to familiarize the children with the verb elicitation task, and items were presented in the same order for all participants. Throughout the task, children received praise for their efforts but were not provided with any feedback about the accuracy of their productions. The examiner pointed at each item and presented the child with a question that created an obligatory context for the use of the target verb inflections in present tense and past tense, as seen in the examples below:

1. Present tense

a. Researcher: *ish byisawwi il-walad halla ?*
What do-PRES-3MS the-boy now?
"What is the boy doing now?"

b. Child: *il-walad byiyakul buza.*
The-boy eat-PRES-3MS ice-cream.
"The boy is eating ice cream."

2. Past tense

a. Researcher: *il-walad xallas, ish sawa il-walad?*
The-boy finish-PAST-3MS, what do-PAST-3MS the-boy?
"The boy finished, What did the boy do?"

b. Child: *il-walad akal buza.*
The-child eat-PAST-3MS ice-cream.
The child ate ice cream.

2.2.4 Scoring

The children's responses were transcribed orthographically online and were audio-recorded for further analysis. Children's productions were scored using three methods:

- Overall-item score: The child's response was scored as correct if it was in the correct tense and had the correct person, number, and gender agreement. That is, the child's response was identical to the target. If the response differed from the target verb in any of these elements (e.g., correct tense, person, and number agreement but incorrect gender agreement), it was scored as incorrect. Correct responses received a score of 1, while incorrect responses received a score of 0. The maximum overall score the child could achieve on the task was 60.
- Tense ac score : Tense accuracy was determined based on the context of the picture (present tense vs. past tense). The child's response was scored as correct and received a score of 1 if it matched the target tense, regardless of subject-verb agreement accuracy. In case of an incorrect response, the substitute verb form was recorded for further error analysis.
- Subject-verb agreement score: subject-verb agreement in Arabic is fusional. Therefore, determining the accuracy of subject-verb agreement is not transparent. Inspection of our data revealed the following: (a) children tended to omit different parts of the same prefix. For instance, the third-person masculine singular verb ***byi****drus* “he is studying” was produced as ***yi****drus*, which is a third-person masculine imperfective bare verb, or ***i****drus*, a second-person masculine imperative verb. (b) Children treated the discontinuous circumfix *byi—u* of the third-person plural present tense as separate affixes. Omitting part of the circumfix meant that some, but not all, of the agreement features of the verb were lost. For example, in the verb ***byi****drus****u*** “they are studying,” an omission of *–u* will only change number agreement from plural to singular. However, third-person agreement will not change since the prefix *byi–* is preserved. To account for this pattern, we followed Abdallah and Crago's (2008) scoring approach. Each of the agreement features of the child's response (person, number, and gender) was checked against the agreement features of the target verb (subject in the picture), irrespective of tense accuracy. Each agreement category was scored as correct or incorrect. Hence, we had three scores: person agreement accuracy, number agreement accuracy, and

gender agreement accuracy. Errors in each element were recorded for further error analysis. To illustrate the scoring system, we provide an example in Table 2.3. In the example, the child used the imperative in place of the target present tense verb. Therefore, tense score was 0. In terms of subject-verb agreement, the child produced a second-person, masculine singular verb in place of a third-person feminine singular verb. Therefore, subject-verb person and gender agreement scores received a score of 0 but number agreement received a score of 1. Given that the child's production differed from the target verb in at least one feature, the overall score of the production was 0.

Table 2.3 *An example of the scoring of the responses on the verb elicitation task*

	Verb + affixes	Affixes	Tense	Person Agreement	Number Agreement	Gender Agreement
Target	***bti**drus*	*bti–*	Present	3rd person	Singular	Feminine
Child's production	***i**drus*	*i–*	Imperative	2nd person	Singular	Masculine
Accuracy			incorrect	incorrect	correct	incorrect
Score			0	0	1	0

2.2.5 Reliability

The responses of ten randomly selected children (21% of the sample) on the verb elicitation tasks were scored by a second SLT. This was done to calculate interrater reliability. The agreement between the first author and the second rater was 100% for the overall scores, 98% for tense scores, 100% for gender agreement scores, 100% for number agreement scores, and 97% for the person agreement scores.

2.2.6 Analysis

Statistical analysis was carried out using RStudio software Version 3.6.0 (RStudio Team, 2019). Raw scores were converted to percentages. For each of the tense and agreement accuracy scores, mixed-design analyses of variance (ANOVAs) were conducted, with the target grammatical category as a within-subject variable and group as the between-subjects variable.

Significance levels were set at $p < .05$. Significant interactions were followed by simple effects analysis. Bonferroni corrections for multiple comparisons were applied (Field, 2009, p.373).Type 1 error was controlled for by dividing the significance value ($p < .05$) by the number of comparisons ($N = 4$). Hence, the significance level for all simple effects analysis was $p < .0125$.

2.3 Results

2.3.1 Analysis 1: The Production Accuracy of Verb Tense and Agreement Marking

Overall, the DLD group scored significantly lower than the TD group on the verb elicitation task, $t(16.91) = -3.89$, $p < .001$, $d = 1.36$. Table 2.4 summarizes the accuracy of the verb forms examined in the task.

2.3.1.1 Tense Accuracy

Tense accuracy scores were analysed using a 2 × 2 mixed-design ANOVA, with group as a between-subjects factor (two levels: DLD and TD) and verb tense as a within-subject factor (two levels: past tense and present tense). Analysis revealed a significant main effect of group, $F(1, 44) = 22.36$, $p < .001$, $\eta^2 = .34$ and verb tense, $F(1, 44) = 23.85$, $p < .001$, $\eta^2 = .35$. Also, the group by verb tense interaction was significant, $F(1, 44) = 18.04$, $p < .001$, $\eta^2 = .29$.

The TD group was significantly more accurate in marking past tense than present tense: $t(31) = 2.79$, $p < .0125$, $d = 0.49$. Similarly, the DLD group produced past tense more accurately than present tense : $t(13) = 3.97$, $p < .0125$, $d = 1.06$. The TD group was more accurate than the DLD group in using present tense, $t(14.87) = -3.49$, $p < .0125$, $d = 1.27$, and past tense, $t(44) = -3.36$, $p < .0125$, $d = 1.07$.

Furthermore, we examined whether the production accuracy of present tense verbs varied based on whether the prefix was stressed or not. Children with DLD used present tense verbs with a stressed prefix with 73.33% accuracy ($SD = 29.12$). This was slightly higher than their accuracy of producing verbs with unstressed prefixes, which was 67.13% ($SD = 22.57$). However, this difference was not statistically significant, $t(13) = -1.41$, $p = .18$, $d = 0.38$.

Table 2.4. *Mean Percentages correct (with standard deviations) of the TD and DLD groups for the target morphemes*

	Group	
	TD N=32	**DLD** N=14
Overall accuracy	94.64 (9.06)	77.14 (15.71)***
Tense accuracy	96.09 (6.51)	81.42 (14.93)**
Present tense	94.06 (9.94)	70.24 (24.72)**
Past tense	98.13 (4.47)	92.38 (6.97) **
Agreement accuracy	97.34 (4.86)	85.12 (12.75)*
Gender agreement	98.96 (2.15)	93.10 (7.33)*
Masculine agreement	100.00 (0)	97.32 (4.72)
Feminine agreement	97.77 (4.6)	88.27 (12.09)**
Number agreement	98.7 (4.55)	95.36 (7.11)**
Singular agreement	100.00 (0)	98.81 (2.48)
Plural agreement	97.40 (6.21)	91.91 (8.54)
Person agreement *3rd person*	99.06 (2.71)	92.14 (10.55)**

Note. **TD** = Typically Developing. **DLD** = Developmental Language Disorder.
* = $p < .05$, **= $p < .01$, ***= $p < .001$

2.3.1.2 Subject–Verb Agreement Accuracy

A composite percentage score of subject–verb agreement was calculated for number, gender, and person agreement. Subject–verb agreement accuracy scores were analysed using a 2 × 2 mixed-design ANOVA, with group as a between-subjects factor (two levels: DLD and TD) and verb tense as a within-subject factor (two levels: past and present). There was a main effect of group, $F(1, 44) = 22.5$, $p < .05$, $\eta^2 = .33$. The main effect of tense was nonsignificant, but the interaction between tense and group was significant, $F(1, 44) = 8.39$, $p < .001$, $\eta^2 = .16$. Based on simple effects analysis, the TD group marked subject–verb agreement at a similar level of accuracy for past tense ($M = 97.4\%$, $SD = 5.53$) and present tense ($M = 97.29\%$, $SD = 6.07$), $t(31) = 0.09$, $p = .923$, $d = 0.01$. The DLD group presented a different pattern, showing higher accuracy in marking subject–verb agreement in past tense verbs ($M = 97.92\%$, $SD = 6.07$) compared to present tense verbs ($M = 89.52\%$, $SD = 9.41$), $t(13) = 2.36$, $p < .05$, $d = 0.62$. Furthermore, the TD group was significantly more accurate than the DLD group in marking subject–verb agreement in present tense verbs, $t(14.87) = -3.49$, $p < .0125$, $d = 1.27$, but not in past tense verbs, $t(17.07) = -2.92$, $p = .02$, $d = 1.0$.

2.3.1.3 Subject–Verb Agreement: Gender Agreement Accuracy

This analysis was only conducted for singular verbs as there is no gender distinction in plural verbs in PA. Gender agreement accuracy scores were analysed using a 2 × 2 × 2 mixed-design ANOVA, with group as a between-subjects factor (two levels: DLD and TD) and verb tense (two levels: past and present) and gender category (two levels: masculine and feminine) as within-subject factors. There were significant main effects of group, $F(1, 44) = 17.36$, $p < .001$, $\eta^2 = .28$, and gender, $F(1, 44) = 18.52$, $p < .001$, $\eta^2 = .3$. The group × gender interaction was significant, $F(1, 44) = 9.83$, $p < .01$, $\eta^2 = .18$.

The TD group showed higher accuracy in marking masculine verbs relative to feminine verbs, $t(31) = -2.74$, $p < .01$, $d = 0.49$. The same was observed in the DLD group, $t(13) = -3.31$, $p < .0125$, $d = 0.88$. The TD and DLD groups did not differ significantly in their production accuracy of masculine verbs, $t(13) = -2.12$, $p = .06$, $d = 0.84$., but the DLD group was significantly less accurate than the TD group in using feminine verbs, $t(14.68) = -2.85$, $p < .0125$, $d = 1.04$. There were no significant interactions between group and tense; gender and tense; and group, gender, and tense. Further analysis was conducted for the DLD group to examine whether the production accuracy of the present tense feminine prefix *bti-* and its allomorph *bit-* was affected by stress assignment. The DLD group produced present tense verbs with a stressed prefix (M = 78.57%, SD = 32.31), with significantly higher accuracy than present tense verbs with unstressed prefix (M = 61.43%, SD = 29.83), $t(13) = -2.28$, $p < .05$, d = 0.61.

2.3.1.4 Subject–Verb Agreement: Number Agreement Accuracy

The number agreement accuracy scores were analyzed with a 2 × 2 × 2 mixed-design ANOVA, with group as a between-subjects factor (two levels: DLD and TD) and verb tense (two levels: past and present) and number category (two levels: singular and plural) as within-subject factors. There were significant main effects of group, $F(1, 44) = 7.36$, $p < .01$, $\eta^2 = .14$, and number, $F(1, 44) = 16.76$, $p < .001$, $\eta^2 = .28$. The Group × Number interaction was significant, $F(1, 44) = 4.29$, $p < .05$, $\eta^2 = .11$. Simple effects analysis revealed that the TD group did not differ in the accuracy of marking singular and plural verbs, $t(31) = -2.37$, $p < .0125$, d = .42. In contrast, the DLD group was significantly less accurate in marking plural verbs compared to singular verbs, $t(13) = -3.64$, $p < .0125$, d = .97. The TD and DLD groups were not significantly different in their accuracy of marking singular verbs, $t(13) = -1.79$, $p = .094$, d = .6, or plural verb forms, $t(19.26) = -2.44$, $p = .044$, d = .74. There were no significant interactions between group and tense; number and tense; and group, number, and tense.

2.3.1.5 Subject–Verb Agreement: Person Agreement Accuracy

Person agreement score was based on the accuracy of marking verbs in third person and were analysed with a 2 × 2 mixed-design ANOVA, with group as a between-subjects factor (two levels: DLD and TD) and verb tense (two levels: past and present) as a within-subject factor. There was a main significant effect of group, $F(1, 44) = 12.26$, $p < .001$, $\eta^2 = .22$, with the TD group outperforming the DLD group in person agreement accuracy. There was a main effect of tense, $F(1, 44) = 7.53$, $p < .05$, $\eta^2 = .15$. In general, marking third person in past tense verbs (M = 98.62%, SD = 3.34%) was easier than marking present tense verbs (M = 95.29, SD = 11.06). The Group × Tense interaction was not significant, $F(1, 44) = 2.72$, $p = .08$, $\eta^2 = .02$.

2.3.2 Analysis 2: Error Patterns in Verb Tense and Agreement Marking

2.3.2.1 Tense errors

We compared children with DLD and TD children on the type and frequency of the forms they used in place of the target tense. The frequency of tense substitutes in the DLD group was almost as twice as that of the TD group (see Table 2.5). The tense substitutes were either finite forms or nonfinite/tenseless forms. Finite substitutes involved the use of the incorrect tense (e.g., past tense for present tense). The nonfinite errors involved the use of the imperfective bare verb and the imperative in place of the target tense.

Table 2.5. *Frequency of tense substitutes*

Target	Substitute type		Group	
			TD	DLD
			N	*N*
Present tense	Non-finite	*Imperative*	15	51
		Imperfective	42	59
	Finite	*Past tense*	15	19
		Total	**72**	**129**
Past tense	Non-finite	*Imperative*	3	15
		Imperfective	10	4
	Finite	*Present tense*	7	15
		Total	**20**	**34**

Note. **TD** = Typically Developing. **DLD** = Developmental Language Disorder.

The imperfective bare verb was most commonly used as a substitute for present tense by the DLD group, followed by the imperative and incorrect tense (e.g., past for present). Similarly, the most common present tense substitute in the TD group was the use of imperfective, followed by the imperative and incorrect tense. The frequency of present tense errors was significantly higher in the DLD group compared to the TD group , $\chi^2(2, N = 201) = 7.05$, $p < .05$.

The DLD group used the imperative and the present tense as substitutes for past tense verbs. In rare occasions, they used the imperfective bare verb. On the other hand, the TD group predominantly used the imperfective verb as a default form for past tense, followed by the use of present tense. The TD group rarely used the imperative as a default form in place of past tense. The frequency of past tense substitutes significantly differed between the TD and DLD groups, $\chi^2(2, N = 54) = 10.56$, $p < .001$.

2.3.2.2 Subject–Verb Agreement

For present tense verbs, the frequency of agreement errors in the DLD group was 4 times that of the TD group (see Table 2.6). Inspection of the data in Table 2.6 revealed that some of the agreement errors were associated with tense errors. The majority of the agreement errors were related to the use of the imperative verb and affected person agreement only. The omission of the prefix *byi–* often resulted in the third-person present tense verb being substituted by the second-person imperative verb (tense and person errors). This type of error barely occurred in the TD group. There were few instances where gender and/or number was also affected. An example of this was the use of the second-person masculine imperative instead of the third-person feminine present tense (tense, person, and gender errors).

There were also agreement errors that occurred despite using the correct tense. The majority of errors in the TD and DLD groups affected the third-person plural present tense. Correct agreement for this form requires the use of the circumfix (e.g.*, byi––u* in *byidrusu* "they are studying"). In both groups, it was noted that the plural morpheme *–u* was omitted, which resulted in the third-person singular verb (number agreement error). The third-person feminine singular present tense form had the second highest rate of errors in both groups. In both groups, this form was substituted by its masculine counterpart (gender agreement error).

In general, the frequency of errors that affected past tense production was lower than present tense production. As seen in Table 2.7, some of the agreement errors in past tense were associated with tense errors. The majority of these errors were associated with the imperative and only affected person agreement, for instance, when the third-person plural past tense was replaced with the second-person plural imperative (person and tense error). In a few occurrences, gender agreement was also affected. An example of this was the use of the second-person masculine imperative in place of the third-person feminine past tense (tense, person, and gender errors).

When past tense was used correctly, the majority of agreement errors affected third-person plural past tense. Both the TD and DLD groups showed omissions of the plural suffix –*u*, which resulted in the third-person singular past tense as a substitute (number error). The third-person feminine past tense had the second highest number of errors in both groups. The omission of the feminine suffix –*at* resulted in the third-person masculine as a substitute (gender error).

Table 2.6. *Frequency of subject-verb agreement errors in present tense verbs*

		TD			DLD		
		Target forms			Target forms		
Actual productions		PRES-3MS ***b-yi-**drus*	PRES-3FS ***b-ti-**drus*	PRES-3P ***b-yi-**drus**u***	PRES-3MS ***b-yi-**drus*	PRES-3FS ***b-ti-**drus*	PRES-3P ***b-yi-**drus-**u***
Nonfinite forms	IMPR-2FS *ʔidrus**i***		3			8	2
	IMPR-2MS ***ʔi**drus*	4	2		15	2	
	IMPR-2P ***ʔi-**drus-**u***						23
	IMPF-3MS ***yi-**drus*					1	
Wrong tense	IMPF-3FS ***ti-**drus*				3		
	IMPF-3P ***ti-**drus**u***					1	
	PAST-3MS *daras*			1		3	1
	PAST-3FS *dars-**at***						1
Correct tense	PRES-3MS ***b-yi-**drus*		1	5		5	13
	PRES-3FS ***b-ti-**drus*			3			0
Total		**4**	**6**	**9**	**18**	**20**	**40**

Note. TD = Typically Developing. **DLD** = Developmental Language Disorder. **PRES-3MS** = Present 3rd person masculine singular. **PRES-3FS** = Present 3rd person feminine singular. **PRES-3P** = Present 3rd person plural. **IMPR-2FS** = Imperative 2nd person feminine singular. **IMPR-2MS** = Imperative 2nd person masculine singular. **IMPR-2P** = Imperative 2nd person plural. **IMPF-3MS** = Imperfective 3rd person masculine singular. **IMPF-3FS** = Imperfective 3rd person feminine singular. **IMPF-3P** = Imperfective 3rd person plural. **PAST-3MS** = Past 3rd person masculine singular. **PAST-3FS** = Past 3rd person feminine singular.

Table 2.7. *Frequency of Subject-verb agreement errors in past tense verbs*

		TD			DLD		
		Target forms			Target forms		
		PAST-3MS *daras*	PAST-3FS *dars-**at***	PAST-3P *daras-**u***	PAST-3MS *daras*	PAST-3FS *dars-**at***	PAST-3P *daras-**u***
Actual productions							
Nonfinite forms	IMPR-2MS ***ʔi**-drus*	2	3		3	2	
	IMPR-2FS ***ʔi**-drus-**i***		1			4	
	IMPR-2P ***ʔ-i**drus-**u***			1			5
	IMPF-3MS ***yi**-drus*					1	
Wrong tense	PRES-3MS ***b-yi**-drus*		1				2
Correct tense	PAST-3MS *daras*		2	12		5	8
	PAST-3FS *dars-**at***			3	1		5
	PAST-3P daras-**u**				1		
Total		**2**	**7**	**16**	**5**	**12**	**20**

Note. **TD** = Typically Developing. **DLD** = Developmental Language Disorder. **PAST-3MS** = Past 3rd person masculine singular. **PAST-3FS** = Past 3rd person feminine singular. **PAST-3P** = Past 3rd person plural. **IMPR-2MS** = Imperative 2nd person masculine singular. **IMPR-2FS** = Imperative 2nd person feminine singular. **IMPR-2P** = Imperative 2nd person plural. **IMPF-3MS** = Imperfective 3rd person masculine singular. **IMPF-3FS** = Imperfective 3rd person feminine singular.

2.4 Discussion

This study examined verb morphology production in PA-speaking children with DLD and their TD peers. Using a novel verb production task, we aimed to compare children with and without DLD in terms of their (a) accuracy rates and (b) error patterns of marking tense and subject–verb agreement.

2.4.1 The Production Accuracy of Verb Tense and Agreement Marking

As predicted, there was a significant difference between children with and without DLD in the percentage of correct use of tense and subject–verb agreement verb inflections, with the DLD group scoring significantly lower than the TD group on the verb elicitation task. This suggests that PA-speaking children with DLD have difficulties in using verbal tense and agreement forms. These findings corroborate the well-documented evidence that verb morphology production is an area of vulnerability for children with DLD acquiring Arabic (Abdallah & Crago, 2008; Morsi, 2009; Fahim, 2017), just as it is for other languages, such as English (e.g., Rice & Wexler, 1996), German (e.g., Rothweiler et al., 2012), Swedish (e.g., Hansson & Leonard, 2003), Hebrew (e.g., Leonard& Dromi, 1994), and Italian (e.g., Bortolini et al., 1997).

Overall, the percentage of correct tense marking in the DLD group (82%) was significantly lower than that in the TD group. When the accuracy scores of the groups for both tense forms were contrasted, a remarkable pattern emerged. Despite significant group differences, TD children and children with DLD produced past tense verbs with a high level of accuracy, scoring 98% and 92%, respectively. Conversely, the DLD group had significant difficulties with their use of present tense, with a mean accuracy of 70%. The specific difficulty with present tense production was reported previously for Arabic-speaking children with DLD (e.g., Abdallah & Crago, 2008; Morsi, 2009). This is contrast to findings of Germanic languages (e.g., English, German, Swedish, Danish among others) where children with DLD are reported to have greater difficulties with the past tense (Krok & Leonard, 2015).

The results from Arabic also differ from studies which showed that Hebrew-speaking children with DLD had greater difficulties with using past tense than present tense verbs. These cross-linguistic differences could be attributed to differences in the structural complexity of present and past tense forms across different languages. To illustrate, the higher number of errors exhibited by Hebrew-speaking children with DLD in using past tense relative to present tense has been attributed to the higher number of agreement features required for the past inflection (Dromi et al., 1999). Following this view, in PA, the past tense form is less marked and structurally simpler than the present tense (as discussed in the introduction). For example, the verb *daras* "he studied" is formed by combining the vocalic pattern *a–a* with the root *d–r–s* (there is no overt marking of tense), whereas the present form ***byi**drus* "he is studying" entails the insertion of a vocalic pattern *–u–* plus the addition of a prefix *byi–,* where the politic *b–* indicates present tense. It is also important to note that the prefix in present tense feminine verbs has two variations, *bit*- and *bti*-, the latter contains a consonant cluster. It has been well-documented that children with DLD have difficulties with the production of syllables containing consonant clusters (e.g., Marshall & van der Lely, 2007). Hence, the articulatory complexity of producing present tense prefixes may attribute to the difficulties with this form. No measures of articulatory production were taken in this study and this notion requires further investigation.

In terms of subject–verb agreement, children with DLD produced 85% of the verbs with the correct marking all target subject-agreement categories. This was significantly lower than the TD group who showed an almost ceiling effect, with their subject-verb agreement accuracy being 97%. Interestingly, the overall accuracy for subject-verb agreement marking in the DLD group was higher than for marking tense. This suggests that marking of tense was more problematic than marking subject–verb agreement for our sample. Abdallah and Crago (2008) also reported that preschool-age, Hijazi Arabic–speaking children had higher accuracy scores in marking subject–verb agreement (77%) compared to tense (68%). Difficulty with subject–verb agreement is not surprising, as the subject and verb must agree on several grammatical categories, including person, number, and gender. Furthermore, agreement in PA is fusional, where more than one agreement category is denoted by a single inflection. For example, the suffix *–at* in *dars**at*** "she studied" denotes third person, feminine gender, and singular number simultaneously. In other instance, agreement categories are denoted by a circumfix affix, where a prefix and a suffix are required. An example of this is the circumfix *byi—u* in ***byi**drus**u***, where it indicates third-person plural agreement (no gender distinction). Having to express more than one agreement category simultaneously using less transparent morphemes could be contributing factors in making these forms more challenging (Dromi et al., 1999).

Examination of subject-verb gender agreement marking revealed that the DLD group was similar to the TD group in producing masculine verbs but were less accurate in producing feminine verbs. This pattern was also found in Hijazi Arabic–speaking children with DLD (Abdallah & Crago, 2008). Indeed, in the typical acquisition of Arabic, masculine verb forms are acquired earlier than feminine verb forms, both in production (Aljenaie, 2000, 2001) and comprehension (Al-Akeel, 1998). Furthermore, masculine verb forms are less marked compared to feminine forms (e.g., *daras* "he studied" vs. *dars**at*** "she studied").

Looking at subject-verb number agreement marking, the DLD group was similar to the TD group in producing singular and plural verbs. However, the DLD group was less accurate in their use of plural verbs compared to singular verbs. This can be attributed to the order in which these forms appear in typical development. Singular verb forms are acquired earlier than plural verb forms, both in production (Abdu & Abdu, 1986; Aljenaie, 2001; Basaffar & Safi, 2012; Omar, 1973) and comprehension (Al-Akeel, 1998; Moawad, 2006). Moreover, singular number agreement is unmarked by any overt inflections in present and past tense verbs, whereas plural number agreement is marked by the suffix *–u* (e.g., *daras* "he studied" vs. *daras**u*** "they studied").

In regard to subject-verb person agreement, though there were significant differences between the TD and DLD groups, both groups marked third-person agreement with more than 90% of accuracy. This high level of accuracy can be attributed to the fact that third-person verbs are the first to emerge in the language of TD children acquiring Arabic (Abdu & Abdu, 1986). Our findings are in contrast to the findings of Abdallah and Crago (2008), who reported that Hijazi Arabic–speaking children with DLD had a difficulty with person agreement as they produced third-person verbs with 66% of accuracy (compared to 92% in our study). This difference can be attributed to age differences: In our study, the mean age of the DLD group was 66 months, with the oldest child being 94 months, whereas in Abdallah and Crago's study, the mean age of the DLD group was 57 months, with the oldest child being 63 months.

An interesting observation emerged regarding stressed and unstressed affixes (for a description of stress patterns in PA, see Watson, 2011). Despite the lack of significant statistical differences, the DLD group produced present tense verbs with the stressed prefix more accurately than verbs with the unstressed prefix. For instance, consider the present tense feminine inflection *bti–* and its allomorph *bit–*. The DLD group used present tense feminine verbs with a stressed prefix (e.g., *'**bti**k.tub* "she is writing") with 79% of accuracy compared to 61% of accuracy for verbs with an unstressed prefix (e.g., ***bit**.'naʃ.ʃif* "she is drying").This discrepancy could possibly be attributed to the lower acoustic salience unstressed prefixes (see section 5.2.1).

Furthermore, the past tense feminine agreement morpheme –at as in *ˈdar.s**at*** "she studied" was challenging for the DLD group in our study. This inflection occurs at the end of the word as part of an unstressed syllable, making the suffix *–at* more likely to be missed by children with DLD possibly due to its lower acoustic salience. This suffix was often omitted from the past feminine verb forms, resulting in a masculine verb *da.ras* "he studied." The plural inflection *–u* as in *ˈda.ra.s**u*** "they studied" was not problematic for the DLD group. The plural inflection always occurs in a final unstressed syllable (Watson, 2011), which would have lower acoustic salience relative to the other syllables in the verb. Stressed syllables are typically louder and longer, making them have a high perceptual salience. Although the accuracy of using inflections was higher when they were stressed compared to being unstressed, the scores of the DLD group on the stressed inflections were relatively low. This suggests that, even though children with DLD may have difficulties in perceiving morphemes of low acoustic saliency, this is unlikely to be the only factor that underpins their difficulties with verb morphology production, and further research is needed to address this issue.

2.4.2 Error Patterns in Verb Morphology Production

Qualitative analysis revealed that the target tense forms were substituted by either finite forms (incorrect tense) or nonfinite/tenseless forms (imperative and the imperfective bare verb). Interestingly, the TD and DLD groups appear to display the same tense substitution patterns, but they differ in the frequency of their use. As predicted, the most frequent tense substitution patterns in the DLD group were the use of the imperative as well the imperfective bare verb. These two nonfinite forms occurred with equal frequency. On the other hand, the use of the imperfective bare stem was the most common substitute noted in the TD group, whereas the imperative was used less frequently in this group. The use of incorrect tense (e.g., past tense for present tense) was the least occurring tense error in both groups.

A considerable body of research has shown that the verb morphology error patterns displayed by children with DLD are similar to those observed in younger TD children acquiring the same language (Leonard, 2014). In fact, according to the Extended Optional Infinitive (e.g., Rice & Wexler, 1996; Rice et al., 1995), children with and without DLD go through an Optional Infinitive stage in which they treat marking of tense and agreement as being optional in obligatory contexts (e.g., Rice & Wexler, 1996). For example, English- and German-speaking children with DLD tend to use infinitives or bare stem forms instead of the target tense (Rice & Wexler, 1996). Arabic has no infinitive forms. Yet, a stage similar to Optional Infinitive seems to exist in this language. Children with and without DLD in our study used the imperative and imperfective bare verb forms instead of target tense. The use of the imperative has been observed in the language of TD toddlers acquiring Yemini Arabic (Qasem & Sircar, 2017), Egyptian Arabic (Fahim, 2017; Omar, 1973), and Kuwaiti Arabic (Aljenaie, 2001) as well as children with DLD acquiring Hijazi Arabic (Abdallah & Crago, 2008) and Egyptian Arabic (Fahim, 2017; Morsi, 2009). The imperfective bare stem has been observed in the language of TD children acquiring Kuwaiti Arabic (Aljenaie, 2010) and children with and without DLD acquiring Egyptian Arabic (Fahim, 2017). In accordance with Extended Optional Infinitive, the use of the imperative and imperfective bare verb forms as default forms is extended for a longer period in Arabic-speaking children with DLD. Both of these forms are described as being nonfinite (Ajlenaie, 2010) or tenseless (Benmamoun 1999, 2000). Children with and without DLD in our study also used finite forms instead of the target. Our findings thus emphasize that the typology of a language impacts both on the type of structures affected by DLD and on the type of errors that characterize the disorder. Our findings also expand on Paradis and Crago's (2001) proposal that the term default form refers to the optional use of either nonfinite or finite forms instead of target tense, which is observed in children with and without DLD.

A closer look at the types of errors in subject–verb agreement reveals an interesting pattern. The use of the masculine verb instead of the feminine verb was the most dominant gender agreement error in the DLD and TD groups. The error involved the omission of the suffix –at of past tense feminine verbs or the prefix *bti–*/*bit–* of present tense feminine verbs. This type of error has been reported to Arabic-speaking children with typical language development (Aljenaie, 2001, 2010; Omar, 1973) and with DLD (Abdallah, 2002; Abdallah & Crago, 2008; Fahim, 2005).

For the TD and DLD groups, the most dominant number agreement error was the omission of the plural suffix *–u* of the past tense or the suffix *–u* of the circumfix *byi—u* in the present tense verb. This pattern was observed in the TD and DLD groups. This omission error resulted in the unmarked singular verb being a substitute of the marked plural verb. The use of singular verbs in place of plural verbs has also been documented in Arabic-speaking children with and without DLD (Abdallah, 2002; Abdallah & Crago, 2008; Aljenaie, 2001, 2010; Omar, 1973). It can be seen that, in line with our prediction, gender and number agreement errors involved the use of the unmarked form instead of the marked form. In this case, the unmarked masculine and singular verbs were used instead of the marked feminine and plural verbs, respectively. This pattern has been also reported for Kuwaiti Arabic–speaking TD children (Aljenaie, 2001, 2010) and Hijazi Arabic–speaking children with DLD (Abdallah & Crago, 2008). These findings are in support of Omar's (1973) suggestion that the third-person masculine singular may be the default verbal agreement form in Arabic.

We only examined the subject–verb agreement for third-person verbs. Person agreement errors were primarily associated with tense errors. This occurred in cases where the imperative was used instead of the target tense. This pattern differs from the findings of Abdallah and Crago (2008), who documented that Hijazi Arabic–speaking children with DLD used first-person verbs in place of third-person verbs. The pattern also differs from studies reporting that the third-person verbs emerged earlier than second-person verbs (Abdu & Abdu, 1986; Aljenaie, 2001, 2010; Basaffar & Safi, 2012). In the DLD group, the imperative was mostly used instead of present tense verbs (N = 51) and much less frequently in place of past tense verbs (N = 15). Third-person agreement is realized by the prefix of the present tense verb or the suffix of the past tense verb, whereas the imperative second-person agreement is unmarked by any affixes. Therefore, it appears person agreement errors represent the use of the unmarked second-person imperative instead of the marked third-person present/past tense verb. Based on the current data and the test items, it is difficult to determine whether the difficulty is in marking tense or person agreement. To determine this, an additional examination of first- and second-person verb production is needed.

It is important to note that Abdallah and Crago (2008) reported that when Hijazi Arabic–speaking children with and without DLD made tense or agreement errors, the inaccurate production differed from the target verb by one feature only. Inspection of our data reveals a similar pattern. Apart from the use of the imperative (tense and person error), the majority of errant productions of the TD and DLD groups differed from the target by one feature. These errors are referred to as being “near misses” and have been documented in richly inflected languages such as Hebrew and Spanish (for a review, see Leonard, 2014). Another important observation is that most errors in the TD and DLD groups were made in forms in which agreement is realized by a circumfix morpheme. In our study, this form was the third-person plural present tense verb in which tense, person, and number agreement are expressed by the circumfix *byi—u*. The children in our study treated the circumfix affixes as separate units. The most common error was the omission of the prefix *byi–* while retaining the suffix *–u*. A similar pattern was noted in Kuwaiti Arabic in which the third-person plural present tense verb is expressed with the circumfix *yi—oon*. Aljenaie (2001) found that the TD Kuwaiti Arabic–speaking children tended to omit the prefix *yi–* and maintain the suffix *–oon*. The second error pattern in our study involved omission of the plural suffix *–u* while retaining the prefix, and this pattern was documented in Hijazi Arabic–speaking children with DLD (Abdallah & Crago, 2008) and was also observed in TD Kuwaiti Arabic–speaking children (Aljenaie, 2010). For a discussion of these findings in relation to prominent DLD theoretical accounts, see section 5.2.1.

2.4.3 Clinical Implications

Given the lack of standardized Arabic assessments for PA, the diagnosis of DLD is based on informal evaluation procedures that are combined with subjective clinical judgments, which may lead to variations and inconsistencies across SLTs as to which structures are targeted in the assessment of DLD. The results of our study provide SLTs with a description of specific verb morphology difficulties in Arabic-speaking children with DLD. Significant differences between children with DLD and TD controls were found in using present tense and verbs with feminine inflections. The findings indicate that SLTs should consider targeting these structures in the assessment and intervention of PA children with DLD.

2.4.4 Limitations

One of the limitations was the small sample size of the DLD group. This is due to the limited number of clinics in Ramallah city from which this group was recruited. Future studies are recommended to include larger sample sizes. The study provides results about the deficits of verb morphology production only and no data on children's comprehension of verb morphology. To achieve a full understanding of the underlying mechanisms of DLD, other aspects of verb morphology should be examined. These should include comprehension and grammaticality judgment tasks, tasks investigating first- and second-person morphemes, and tasks that target derivational and inflectional morphology. It will important for future studies to examine the diagnostic accuracy of the verb elicitation task in differentiating between children with and without DLD. This will determine the clinical usefulness of the task in identifying DLD in Arabic.

2.5 Conclusion

The findings show that PA-speaking children with DLD present with deficits in the production of verb morphology relative to TD children. Inflected verbs with increased markedness, including present tense and feminine verb forms, were more challenging for the DLD group than past tense and masculine verb forms, respectively. The latter forms were produced with a high accuracy by the DLD group. The production of singular, plural and third-person verbs was not impaired in Arabic-speaking children with DLD. For the TD and DLD groups, the most frequent tense and agreement error patterns included omissions of the target morphemes. The omission of target morphemes often resulted in the children producing structurally simpler (less marked) verb forms instead of marked verb forms. Also, although it seemed that the DLD group was more accurate with some stressed than unstressed forms, the scores of the DLD group were still lower than the TD group. Future studies would need to include larger sample sizes to increase statistical power and generalizability of the findings; investigate other aspects of verb morphology, including both production and comprehension; consider other language domains, such as syntax, phonology, and semantics; and employ longitudinal designs to provide more in-depth knowledge of Arabic language acquisition.

Acknowledgments

This study was funded by the University of Reading International Studentship (GS18-082) awarded to the first author. We would like to thank the children and their parents who participated in the study. We also thank the teachers and the speech and language therapists who facilitated recruitment of the children and, in particular, Hana' Al-amleh and Alaa Warasneh for their help in recruiting children with DLD.

Chapter 3: Nonword repetition performance of Arabic-speaking children with and without Developmental Language Disorder: A study on Diagnostic Accuracy

Abstract

Purpose: This study evaluates the effectiveness of a nonword repetition (NWR) task in discriminating between Palestinian Arabic–speaking children with developmental language disorder (DLD) and age-matched typically developing (TD) children.

Method: Participants were 30 children with DLD aged between 4;0 and 6;10 (years;months) and 60 TD children aged between 4;0 and 6;8 matched on chronological age. The Arabic version of a Quasi-Universal NWR task was administered. The task comprises 30 nonwords that vary in length, presence of consonant clusters (CCs) and wordlikeness ratings. Responses were scored using an item-level scoring method. Receiver operating characteristic curve analysis was conducted to determine the best cutoff point with the highest sensitivity and specificity values, and likelihood ratios were calculated.

Results: Children with DLD scored significantly below age-matched TD peers on the NWR task. The DLD group found the repetition nonwords with two more challenging than nonwords with no or only one CC. For the TD and DLD groups, three-syllable nonwords were repeated less accurately than two- and one-syllable nonwords. Also, high word-like nonwords were repeated more accurately than nonwords with low wordlikeness ratings. The best cutoff score on the NWR correctly classified 93% of the TD children and 93% of children with DLD. The likelihood ratios indicated that the NWR task is informative of the presentence/absence of DLD.

Conclusions: NWR was an area of difficulty for Palestinian Arabic–speaking children with DLD. Nonwords with one and two CCs were significantly more challenging for the DLD group relative to the TD group, suggesting that phonological complexity was sensitive to the language abilities of both groups. The effects of nonword length and wordlikeness ratings on performance suggest that the task taps into verbal short-term memory and information stored in long-term memory. The NWR appears to hold promise for clinical use as it is a useful indicator of DLD in Arabic. These results need to be further validated using population-based studies.

3.1 Introduction

Developmental language disorder (DLD) affects approximately 7% of children at school entry (Norbury et al., 2016), and it refers to difficulties in understanding and/or using language without a known biomedical etiology. These difficulties interfere with everyday life and educational achievement and are likely to persist into school age and beyond (Bishop et al., 2016, 2017). Given the negative impact of DLD on the quality of life of affected children, early identification of the disorder is imperative.

Clinical markers are tasks that can reliably capture the difficulties experienced by children with DLD and exclude those with typical language development. Therefore, these tasks play an important role in accurate identification and appropriate treatment of DLD. Cross-linguistic evidence shows that nonword repetition (NWR) may be a reliable clinical marker of DLD in monolingual and bilingual children speaking a variety of languages (for a review, see Chiat, 2015; Schwob et al., 2021). Our study aims to investigate NWR abilities of Palestinian Arabic–speaking children with DLD aged 4–6 years relative to chronological age–matched typically developing (TD) peers. Importantly, the study will evaluate the diagnostic accuracy of NWR as a potential clinical marker of DLD in Arabic. Exploring the diagnostic accuracy will inform clinicians to what extent NWR can accurately distinguish Palestinian Arabic–speaking children with and without DLD. We begin with an overview of the cross-linguistic evidence for NWR deficits in children with DLD, followed by a review of the usefulness of NWR tasks as possible diagnostic markers of DLD, and factors that may influence performance on NWR tasks.

3.1.1 NWR deficits in children with DLD: Cross-linguistic evidence

NWR tasks assess the ability to encode, temporarily store, retrieve, and imitate an unfamiliar string of phonemes that conform to the phonotactics of the child's native language, yet lack any meaning. NWR resembles a crucial skill that underlies early word learning: children's ability to spontaneously repeat the new, unfamiliar words they hear. NWR has been reported to correlate with TD children's concurrent vocabulary size (e.g., Gathercole, 2006; Melby-Lervåg et al., 2012) and to predict vocabulary acquisition (e.g., Gathercole et al., 1997).

Studies have consistently reported that English-speaking children with DLD are significantly less accurate in repeating nonwords compared to their TD peers and that these group differences persist across development (for a review, see Graf Estes et al., 2007). The finding that NWR is impaired in children with DLD has been replicated in many languages, including

Italian (Bortolini et al., 2006), Spanish (Girbau, 2016; Girbau & Schwartz, 2007), French (Thordardottir et al., 2011), Dutch (Rispens & Parigger, 2010), Swedish (Kalnak et al., 2014), Slovak (Kapalková et al., 2013), and Turkish (Topbaş et al., 2014) among others.

In contrast, Cantonese-speaking children with DLD (age range: 4;2–5;7 years; months) have been reported to perform as well as age-matched TD children on an NWR task, suggesting that NWR is not a clinical marker of DLD in this language (Stokes et al., 2006). As the NWR task in Stokes et al.'s (2006) study was based on the phonotactic rules of Cantonese, these findings were attributed to the phonologically less complex nature of Cantonese compared with other languages. According to Stokes et al., Cantonese is a tonal language with a small phonemic inventory, basic syllabic structure (consonant–vowel [CV] only), and only a limited set of syllabic combinations are allowed. Additionally, syllables in multisyllabic words are equally stressed (i.e., quite salient). Therefore, it could be that the nonwords used in Stokes et al. were not as complex as the nonwords used in other languages with more complex syllabic structures and stress variations (e.g., English). Notably, a subsequent study found that 5-year-old Cantonese-speaking children with DLD scored below their age-matched TD controls on NWR (Wong et al., 2010). Although the between-group difference was only marginally significant ($p = .06$), Wong et al. (2010) argued that Cantonese-speaking children's weak performance on the nonword (and word) repetition tasks relative to age norms suggests that these children have an impairment in this domain. The contradictory results of the two Cantonese studies were attributed to differences in the NWR tasks and scoring methods (for a discussion, see Wong et al., 2010). Recently, Pham and Ebert (2020) found that Vietnamese-speaking children with DLD performed poorly on NWR relative to same-age TD peers. In line with the results of Wong et al.'s study and contrary to those of Stokes et al., Pham and Ebert (2020) found that NWR could discriminate between Vietnamese-speaking children with and without DLD, which suggests NWR tasks may have potential in detecting DLD in Asian tonal languages.

3.1.2 Factors Influencing NWR Performance

It is well documented that nonword length, that is, the number of syllables, affects how accurately children repeat nonwords (e.g.,Coady & Evans, 2008). TD children, as well as children with DLD, typically show accurate repetition of short nonwords (i.e., one and two syllables). As the nonwords increase in length (three or more syllables), the repetition accuracy decreases for both groups, particularly for children in the DLD group (Archibald & Gathercole, 2006; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Jones et al., 2010; Weismer et al., 2000). According to Chiat (2015), this length effect has been replicated in all languages studied to date.Phonological complexity is another factor that influences NWR accuracy. Phonologically complex nonwords with consonant clusters (CCs) are repeated less accurately than phonologically simple nonwords that only contain singleton consonants. Although articulatory complexity affects children with and without DLD (Coady & Evans, 2008; Edwards & Lahey, 1998; Gathercole & Baddeley, 1990), children with DLD are more adversely affected by the presence of CCs relative to TD peers (Briscoe et al., 2001; Leclercq et al., 2013; Munson et al., 2005).

Long-term language knowledge also plays a role in NWR. NWR accuracy appears to be influenced by two closely related factors: Wordlikeness (the extent to which a nonword resembles a real word based on native speakers' judgment) and phonotactic probability (an objective measure of the frequency of the occurrence of a specific sound or sound combination in a given language). Nonwords that sound like real words in a given language receive high ratings from adults as being word-like. Nonwords with high word-like ratings are repeated by children more accurately than nonwords that are rated as less word-like (Archibald & Gathercole, 2006; Briscoe et al., 2001; Coady et al., 2010; Gathercole, 2006; Munson et al., 2005). High word-like nonwords overlap with real lexical items in long-term memory thus will be more easily repeated than nonwords with low wordlikeness ratings (Bowey, 2001; Metsala, 1999; Snowling et al., 1991; Szewczyk et al., 2018). Furthermore, nonwords containing high phonotactic probability sequences are repeated more accurately than nonwords containing low phonotactic probability sequences (Munson et al., 2005). Some studies have found that wordlikeness and phonotactic probability have a larger effect on NWR accuracy of children with DLD relative to TD peers (Jones et al., 2010; Leclercq et al., 2013; Munson et al., 2005). For instance, Munson et al. (2005) reported that the difference in NWR accuracy between children with DLD and TD children was larger on items with low phonotactic ability than on those with high phonotactic probability. However, others have found no differences between children with and without DLD (Coady et al., 2010).

3.1.3 NWR as a Clinical Marker of DLD

The statistically reliable difference between children with and without DLD on NWR tasks is important. However, it does not inform us about its clinical usefulness for the identification of DLD. This requires determining its diagnostic accuracy. Diagnostic accuracy is indexed by measures of sensitivity, that is, the proportion of children with a DLD diagnosis correctly identified by the task (true positive rate), and specificity, that is, the proportion of children without a disorder correctly identified by the task (true negative rate). A threshold score should be set as a cutoff point for the analysis of sensitivity and specificity. The classification accuracy of a cutoff point with specificity and sensitivity values above 80% is considered acceptable, with values above 90% being excellent (Plante & Vance, 1994). Dollaghan and Campbell (1998) recommend also calculating positive likelihood ratio (LR+), that is, the probability to be identified as impaired if impaired, and negative likelihood ratio (LR−), that is, the probability to be identified as unimpaired if unimpaired. Following the guidelines of Sackett et al. (1991), Dollaghan (2007) indicated that values of LR+ ≥ 10.0 and LR− ≤ 0.1 can be interpreted with high confidence to rule in or rule out the disorder, respectively, whereas values of LR+ ≥ 3.0 and LR- ≤ 0.3 are suggestive but insufficient to rule in or rule out the disorder, respectively.

The findings of studies that have examined the use of NWR in distinguishing English-speaking children with DLD from TD children are inconclusive (for a review, see Pawłowska, 2014). Dollaghan and Campbell (1998) found that an overall score of 70% of less on the Nonword Repetition Test (NRT) was 25 times more likely to come from a child with language impairment than from a TD child, suggesting that the NRT had a high degree of accuracy in differentiating children with and without language impairment. However, using the same cutoff point with 7- to 8-year-old children, Weismer et al. (2000) found the LR+ to be 2.78, indicating that the diagnostic accuracy of the NRT was “intermediate” and not sufficient to identify language impairment in this age group. Subsequent studies have reported high levels of sensitivity and specificity of NWR in identifying DLD in preschool-age children (Deevy et al., 2010) and in 7-year-old children (Redmond et al., 2011). Some studies have found lower levels of sensitivity and acceptable levels of specificity of NWR in identifying DLD (Conti-Ramsden, 2003; Conti-Ramsden et al., 2001), while other studies have documented low values for sensitivity and specificity (Archibald & Joanisse, 2009). The discrepancy of results across studies from English-speaking populations may be due to the variability in reference standards used to identify children with DLD, the structure of NWR tasks used, and their scoring methods (for a review, see Graf Estes et al., 2007; Pawłowska, 2014).Some studies have followed a one-gate design by recruiting unselected population samples (Weismer et al., 2000), others have followed a two-gate design by recruiting preselected TD and DLD groups (e.g.,Conti-Ramsden, 2003; Conti-Ramsden et al., 2001; Deevy et al., 2010; Gray, 2003; Redmond et al., 2011). Pawłowska (2014) argued that one-gate studies include children with DLD across the ability spectrum, some of which could have borderline scores, whereas two-gate studies include children with a prior diagnosis of DLD who are likely to have severe language difficulties as they were enrolled for intervention. Hence, TD and DLD group differences in two-gate studies are likely to be larger than in one-gate studies leading to variations in diagnostic accuracy levels. The diagnostic accuracy of NWR has also been examined in languages other than English (see Table 3.1 for a summary). Most studies have documented good sensitivity and specificity values of above 80%, showing the clinical value of NWR in distinguishing children with and without DLD across languages.

Table 3.1. *Summary of cross-linguistic findings on the diagnostic accuracy of nonword repetition in identifying DLD in monolingual children*

Reference	Language	*TD* N	*TD* Age in years	*DLD* N	*DLD* Age in years	Sensitivity %	Specificity %	LR+	LR-
Ahufinger et al. (2021)	Portuguese	75	7;0 – 11;11	75	7;0 – 11;11	47	99	35.92	.54
Armon-Lotem & Meir (2016)	Hebrew	38	6 (.17)	14	6;1 (.33)	93[a]	66	2.71	.11
Armon-Lotem & Meir (2016)	Russian	20	6;1 (.17)	14	5;10 (.25)	86	90	8.57	.16
Bortolini et al. (2006)	Italian	11	3;7 - 5;5	11	3;7 - 5;6	82	82	4.56	.22
Dispaldro et al. (2013)	Italian	17	3;11-5;8	17	4;1 - 5;7	100	1	ND[b]	ND
Girbau (2016)	Spanish	20	8;1- 10;3	20	8;0 - 9;11	100	85	6.67	0
Guiberson & Rodríguez (2013)	Spanish	23	4;1 (.82)	21	3;11(.81)	71	74	2.74	.39
Kalnak et al. (2014)	Swedish	86	9;4 (1.3)	61	9;3 (1.2)	90	98	38.8	.10
Kapalková et al. (2013)	Slovak	16	4;3 - 5;6	16	4;2 - 5;6	94	100	ND	.06
Kazemi & Saeednia (2017)	Farsi	31	4;8 (.7)	20	4;5 (.74)	90	96	27.9	.10
Pham & Ebert (2020)	Vietnamese	194	5;8 (.4)	10	5;5 (.3)	90	79	4.53	.13
Thordardottir et al. (2011)	French	78	4;1 - 5;11	14	4;6 - 5;11	85	88	6.77	.18
Topbaş et al. (2014)	Turkish	120	4;4 - 8;0	20	4;2 - 8;3	89	87	6.85	-.02

Note. **TD**: Typically Developing. **DLD:** Developmental Language Disorder. **LR+**: Positive Likelihood Ratio. **LR-** : Negative Likelihood Ratio. **ND**: not defined.

[a] Sensitivity and Specificity and LR values are reported for the best cutoff points.

[b] When the specificity is 100, the LRs are undefined.

In Arabic, Wallan (2018) examined the clinical utility of the adapted Verbal Short Term Memory test (Pickering & Gathercole, 2001), which included digit recall, word list recall, and nonword list recall tasks. The nonword list recall was administered to a "language concern" (LC) group, which included children whose parents/teachers had concerns about their language development (*N* = 14, age range: 2;10–5;11) and a group of TD children matched on age and nonverbal IQ. The "LC" group scored slightly lower than the TD group on the nonword list recall task. Wallan found that this task failed to distinguish between the two groups and attributed the poor diagnostic accuracy of the task to the limited range of scores in the TD children. The poor diagnostic accuracy of nonword list recall in Wallan's study can also be explained in relation to the reference standard according to which children were placed in an "LC group." The sole reliance on parental/teachers' reports as an indicator of language status could mean that some of the children in the "LC" group did not have language impairment of clinical significance.

Previous studies revealed that, on average, Arabic-speaking children with DLD scored below their age-matched TD peers on NWR tasks (Abi-Aad & Atallah, 2020; Balilah, 2017; Khater, 2016; Saiegh-Haddad & Ghawi-Dakwar, 2017; Shaalan, 2010, 2020). However, group differences are not sufficient to conclude that poor NWR is a clinical marker of DLD in Arabic-speaking children, due to the high degree of variability in individual DLD profiles. Therefore, the extent that NWR can be an accurate indicator of the presence or absence of DLD in Arabic remains unclear. Exploring the diagnostic accuracy is thus necessary as it considers the individual differences among children with DLD. Examination of diagnostic accuracy can also determine the accuracy of NWR in differentiating between Arabic-speaking children with DLD from TD peers.

In Palestine, the identification of DLD is an ongoing challenge, as no standardized language assessments are available. As a result, Palestinian children with DLD are particularly vulnerable to being misdiagnosed or just missed altogether. Diagnostic tools are needed to facilitate effective and efficient identification of DLD in Arabic. In response to this issue, this study attempts to provide speech and language therapists (SLTs) with evidence of the potential of NWR as a screening measure. This, in turn, can help enhance the accuracy of assessment procedures when diagnosing DLD in Palestinian children.

3.1.4 Aims

Existing studies have provided important insights about the potential of NWR as a clinical marker of DLD in Arabic. However, information about the clinical usefulness of this measure is yet to be determined. In this study, the Arabic version of a Quasi-Universal Nonword Repetition task (dos Santos et al., n.d.) was employed to address the following questions:

1. How do children with DLD compare to age-matched TD children in terms of their NWR performance accuracy?
2. How accurate is NWR performance in distinguishing Palestinian Arabic–speaking children with DLD from their age-matched TD peers?

3.2 Method

3.2.1 Participants

This study received ethical approval by the University of Reading Research Ethics Committee. There were 90 participants in two groups: a group of 60 TD children and 30 children with DLD. All participants were monolingual native speakers of Palestinian Arabic. According to parents and teachers' reports, all participants had normal hearing, and no symptoms or history of neurological deficits, oral–motor impairments, or social–emotional/behavioral difficulties. See Table 3.2 for demographic information.

The TD children (27 females and 33 males) aged between 4;0 and 6;8 (M = 63.85 months, SD = 10.16 months) were recruited from three kindergartens in the same geographical area as the DLD group. Additional inclusionary criteria for this group were (a) age-appropriate language skills as reported by their caregivers and (b) no history of speech-language therapy. The children with DLD (eight girls and 22 boys) aged between 4;0 and 6;10 (M = 61.50 months, SD = 11.27 months) were recruited from five private speech therapy clinics in Ramallah City, Palestine. Each child in the TD group was within 2 months of age of a child in the DLD group. The two groups were matched on chronological age, $t(53.04) = -0.96$, $p = .34$, $d = .22$.

Table 3.2. *Participants' characteristics*

	Group	
	TD	*DLD*
Family characteristics	%(*N*)	
Mother's education		
High school	20(12)	33.33(10)
University degree/college diploma	75(45)	53.34(16)
Postgraduate degree	5(3)	13.33(4)
Family history of communication disorders	6.67(4)	30(9)**
Language milestones	Age in months *Mean*(*SD*)	
Babbling	6.22(1.69)	6.33(1.71)
First word	11.72(2.06)	20.43(6.94)***
Word combinations	19.44(3.53)	35.60(9.37)***
Follow simple commands	18.89(5.07)	26.13(7.33)***

Note. **TD**: Typically Developing. **DLD:** Developmental Language Disorder.
* $p < .05$, ** $p < .01$, *** < .001

All 30 children in the DLD group had been diagnosed with DLD by qualified SLTs independent of this study and were receiving language intervention at the time of testing. The diagnosis of DLD in Palestine is made based on qualitative assessment supported by the clinical judgment of the SLTs. Therefore, it was crucial to ensure that the children with DLD met the criteria for DLD as set out by Bishop et al. (2016, 2017). A brief interview with each of the children's SLT was done to confirm that (a) their language disorder was not limited to expressive phonology, but also affected other language components such as semantics morphosyntax and pragmatics among others; (b) their hearing was normal according to audiology reports; (c) and their language disorder was not associated with any biomedical conditions (e.g., neurological and genetic syndromes).

A weakness in expressive morphosyntax is a hallmark of children with DLD (Leonard, 2014). Particularly, Arabic-speaking children with, or at risk of, DLD are known to have difficulties with sentence repetition (Shaalan, 2010; Wallan, 2018), the production of verb inflections (Abdallah & Crago, 2008; Fahim, 2017; Shaalan, 2010), and noun plurals (Abdalla et al., 2013; Shaalan, 2010).Accordingly, three non-standardized language tasks were administered to verify the language status of the TD children and to ascertain that the children with DLD had language skills that were considerably below those expected for their chronological age. These included the following (a) Palestinian Arabic LITMUS Sentence Repetition Test (LITMUS-SR-PA-72; Taha et al., 2021a): the task assesses the production of language-specific structures that are impaired in Arabic-speaking children with DLD and language-independent structures that are documented to be impaired in children with DLD across languages; (b) Arabic Verb Elicitation Test (AVET): a picture-naming task that examines the production of verb tense and agreement inflections (Taha et al., 2021b) ; (c) Arabic Noun Pluralization Test (ANPT; Taha, 2019): a picture-naming task that examines the production of noun plural types. Additionally, we calculated (d) mean morpheme per utterance (MPU). MPU is an index of grammatical development that accounts for the highly synthetic nature and rich morphology of Semitic languages (Dromi & Berman, 1982). MPU is equivalent to the mean length of utterance (Brown, 1973) in English. A language sample of 100 utterances was obtained using the wordless storybook Frog, Where Are You (Mayer, 1969). Using this sample, we followed the guidelines of Shaalan and Khater (2006) for MPU calculations in Arabic. The MPU score reflects the total number of morphemes divided by the total number of utterances produced in the narrative task. Clinically, low mean length of utterance scores are viewed as supporting evidence for the diagnosis of language impairment in children (Rice et al., 2010).In addition to the language tasks, the Coloured Progressive Matrices (CPM; Raven, 2007) was administered to assess the children's nonverbal abilities.

Given that all the measures are not standardized, the results of the TD group (mean and standard deviation) were used to calculate the *z* scores for all participants (see Table 3.3). Each child in the DLD group scored at or below −1.5 SD of the TD mean on at least two of the linguistic measures (LITMUS-SR-PA-72, ANPT, AVET & MPU).See Appendix C for the individual raw scores of all participants on the background measures. Groups were compared using raw scores. Children with DLD scored significantly below the TD children on the LITMUS-SR-PA-72, $t(47.46) = -15.64$, $p < .001$, $d = 3.63$; AVET, $t(31.67) = -9.98$, $p < .001$, $d = 2.52$; the ANPT, $t(84.58) = -12.56$, $p < .001$, $d = 2.58$; and MPU, $t(72.49) = -11.28$, $p < .001$, $d = 2.42$. The raw scores on the CPM did not differ significantly between the groups, $t(51.59) = -1.26$, $p = .214$, $d = .29$

Table 3.3. *A summary of the raw and z scores of the TD and DLD groups on the background measures*

Measures	Group							
	TD				*DLD*			
	Raw scores		*Z scores*		*Raw scores*		*Z scores*	
	M (SD)	Range	M (SD)	Range	M (SD)	Range	M(SD)	Range
LITMUS-SR-PA-72 (out of 100)	82.78(13.95)	30.56 – 100	0(1)	-3.11 – 1.32	24.78(17.76)	1.39 – 56.94	-4.16(1.27)	-5.83 – -1.85
AVET (out of 100)	96.63(5.81)	73.96 –100	0(1)	-3.90 – . 58	60.83(19.21)	14.58 – 89.58	-6.16(3.31)	-14.12 – -1.21
ANPT (out of 100)	74.67(24.68)	20 – 100	0(1)	-2.22 – 1.03	21.99(14.97)	0 – 73.33	-2.14(.61)	-3.03 – -.05
MPU	5.35(.97)	3.15 – 7.48	0(0)	-2.27 – 2.20	3.25(.75)	1.89 – 4.61	-2.17(.78)	-3.57 – -.76
CPM (out of 36)	15.89(3.68)	9 – 23	0(1)	-1.87 – 1.94	14.76(3.99)	9 – 23	-.30(1.09)	-1.87 – 1.94

Note. **TD** = typically developing; **DLD** = developmental language disorder; **LITMUS-SR-PA-72** = Palestinian Arabic LITMUS Sentence Repetition Test; **AVE**T = Arabic Verb Elicitation Test; **ANPT** = Arabic Noun Plurals Test; **MPU** = mean morpheme per utterance; **CPM** = Coloured Progressive Matrices (Raven, 2007).

3.2.2 The nonword repetition task

The design of the NWR task used in this study was motivated by the Crosslinguistic Nonword Repetition (CL-NWR) Framework (Chiat, 2015), which was established within the COST Action IS0804 “Language Impairment Testing in Multilingual Settings” (LITMUS; Armon-Lotem et al., 2015). The goal of the CL-NWR Framework was to design NWR tasks containing nonwords of minimal language-specific features such that these tasks can discriminate between children with and without DLD regardless of their language background (Chiat, 2015). The framework is composed of three types of tests that vary in the phonological characteristics of nonwords, one of which is the Cross-linguistic (Quasi-Universal) NWR test (CL-NWRT; Chiat, 2015). The test examines phonological short-term memory and was constructed to be maximally compatible with languages with diverse phonological systems. Specifically, the test contains 16 nonwords varying in length from two to five syllables. The syllables are of CV structure, a simple syllable type that is relatively universal. The syllables of nonwords were composed using a set of consonants /p, b, t, d, k, g, s, z, l, m, n/ and vowels /a, u, i/ that are the most common sounds across languages (Chiat, 2015).

Within the CL-NWR, dos Santos and Ferré (2018) developed the French LITMUS Nonword Repetition Test (LITMUS-NWRT). The test aimed to assess phonology with a particular focus on the effects of phonological complexity. Three phonological aspects (based on French phonology but also applicable to a large number of different languages; dos Santos & Ferré, 2018) were systematically manipulated including syllable structure, segmental complexity, and sequential complexity. In line with the CL-NWR Framework (Chiat, 2015), the LITMUS-NWR task contained a set of language-specific nonwords and a set of language-independent (quasi-universal) nonwords. The latter set was created using phonemes and phonotactic rules compatible with a large number of languages (Maddieson et al., 2011). Furthermore, this set was adapted into Lebanese Arabic by dos Santos et al. (n.d.) resulting in the Arabic version of the Quasi-Universal LITMUS-NWRT (A-QU-LITMUS-NWRT). The set was adapted to identify Lebanese bilingual children whose first language was Arabic and second language was French/English.

With regard to syllabic structure complexity, the items of the A-QU-LITMUS-NWRT had 13 syllabic structures made of three-syllable types. The first type was CV syllable structure, which was the same structure used in the CL-NWRT (Chiat, 2015). The A-QU-LITMUS-NWRT also included CCV and CVC syllables, which were not present in the CL-NWRT (Chiat, 2015). While syllables with CV structure are common across all languages, syllables with CCs or codas are not. The inclusion of these structures was justified by their known effects on NWR performance in languages that permit them, in this case: French, Arabic, and English (e.g., Coady & Evans, 2008; dos Santos & Ferré, 2018; Shaalan, 2010).

Segmental complexity of the nonwords was varied for the consonants. This resulted in a smaller set of consonants compared to the CL-NWRT (Chiat, 2015). The nonwords were created using only four consonants /k, f, b, l/ and three vowels /a, u, i/. The stops /p/ (in the Arabic version /b/) and /k/ were contrasted for their place of articulation with /k/, a dorsal stop, being more complex than /b/, which is a labial stop (dos Santos & Ferré, 2018). These two stops were contrasted with the fricative /f/ of which the manner of articulation is considered to be more complex. Moreover, the liquid /l/ was chosen to enable the formation of nonwords with branching onsets that are permitted across many of the world's languages (dos Santos & Ferré, 2018). Importantly, these consonants are acquired early in the phonological systems of most languages (Abi-Aad & Atallah, 2020; dos Santos & Ferré, 2018). In Arabic, /k/ and /f/ are acquired by ages 2;10, /b/ is acquired by ages 3;4, and /l/ by ages 3;10 (Amayreh & Dyson, 1998). Additionally, sequential complexity (sequentiality) was taken into account. According to dos Santos and Ferré (2018), sequentiality could increase item complexity at two levels: consonant sequences and syllable sequences (for further details, see dos Santos & Ferré, 2018).

The A-QU-LITMUS-NWRT contained 30 nonwords varying in length from one to three syllables. Given that the main purpose of the QU-LITMUS-NWRT was to assess effects of phonological complexity, the influence of working memory was restricted by limiting the length of nonwords to three syllables (Abi-Aad & Atallah, 2020; dos Santos & Ferré, 2018). Hence, the nonwords in the current task are shorter (up to three syllables) compared to those in the CL-LITMUS-NWR test (Chiat, 2015), which increased the nonwords' syllable number (up to five syllables) rather than syllable complexity to be compatible with languages that lack complex syllables.

According to Abi-Aad and Atallah (2020), the A-QU-LITMUS-NWRT has quasi-universal prosody to control for familiarity with lexical phonology of the target. That is, the syllables of the nonwords receive equal stress and they are produced with even length and pitch, with the exception of the final syllable lengthening, which typically marks the end of an utterance (Chiat, 2015). In this way, language-specific prosodic patterns were avoided.

Lastly, given that wordlikeness affects NWR performance (Archibald & Gathercole, 2006), a familiarity questionnaire (Abi-Aad & Atallah, 2020) was used to obtain familiarity ratings for the nonwords from 30 Palestinian Arabic–speaking adults (10 males, M_{age} = 25.32 years, *SD* = 5.79). After hearing the auditorily presented nonwords, participants were asked to rate each nonword on a 5-point scale, where 1 = this word is very unlike an Arabic word and 5 = this is a very Arabic-like word. Nonwords with an average score above 2.5 were considered to be of high wordlikeness, and those equal or below 2.5 were considered to be of low wordlikeness. There were seven nonwords in the high wordlikeness category (*M* = 3.43, *SD* = .74) and 23 nonwords in the low wordlikeness category (*M* = 1.65, *SD* = .33). The items on the A-QU-LITMUS-NWRT (dos Santos et al., n.d.) are presented in Appendix D.

3.2.3 Procedure

Written informed consent was obtained by the parents of all participating children before testing. Children were participating in a larger study and completed a battery of tests in two separate sessions each lasting approximately 1 hr. In the first session, CPM, a narrative task, ANPT and LITMUS-SR-PA-72 were administered; in the second session, A-QU-LITMUS-NWRT and AVET were administered. All tests were conducted by the first author who is a qualified SLT and a native speaker of Palestinian Arabic. Each child was tested individually, in a quiet room, in their kindergarten or the speech and language therapy clinic they were attending.

The NWR task was administered in the form of a stringing beads game. Children were given wooden animal beads and were given the following instruction in Arabic: "Now, you will put the wooden animal block next to your ear and listen to the funny word it will say. Listen carefully and repeat the funny word immediately and exactly as you heard it. After you repeat the funny word, you will insert the bead in the thread. Then, you will pick up another animal bead and listen to another funny word" and so on. The nonwords were produced live by the researcher. Live presentation is less consistent compared to the use of audio-recorded nonwords. However, it is a more natural approach, and it is more relevant to clinical practice in that it is similar to tasks employed in speech and language therapy sessions (Chiat & Roy, 2007). The use of an interactive game alongside the live presentation of nonwords has been used in previous studies and shown to be effective in motivating children and maintaining their attention (Chiat & Roy, 2007; Kapalková et al., 2013).To ensure consistency of the delivery of the stimuli across children, the first author practiced the production of the items and conducted the test with all children.

Two practice items were provided before the test was administrated. The practice nonwords were repeated until the children understood what they had to do. The experimental nonwords were presented in a fixed randomized order to all children. Each experimental nonword was only presented once unless there was an interruption to the first presentation (e.g., loud noise, the child being distracted). If the child self-corrected himself/herself, the final response was scored regardless of its accuracy. To keep the children motivated, they were praised with "well done" or "bravo" for their responses irrespective of their accuracy. The children's responses were audio-recorded and were transcribed phonetically off-line by the first author for analysis.

3.2.4 Coding and Scoring

Following the CL-NWR Framework (Chiat, 2015), children's responses were scored using item-level scoring. Each repeated nonword was scored as correct if it contained all the consonants and vowels of the target in the correct order. This scoring method did not allow for typical developmental phonological errors. Repetitions that included any additions, omissions, or substitutions were scored as incorrect. Correct repetitions received a score of 1 while incorrect repetitions received a score of 0. The maximum raw score was 30. Item-level (binary) scoring is a straightforward scoring method for SLTs to use in clinical settings. Item-level scoring is commonly used for NWR tests such as the Children's Test of Nonword Repetition (Gathercole et al., 1994) and the Early Repetition Battery (Seeff-Gabriel et al., 2008).

Calculating the percentage phonemes correct (PPC) is also a common scoring method for NWR tests. Roy and Chiat (2004) compared the item-level scores and PPC scores in a sample of English-speaking children. They concluded that the two scoring methods were equally able to differentiate between TD and clinical samples, but item-level scoring was less time consuming. Kapalková et al. (2013) explored several NWR scoring methods in a sample of Slovak-speaking children. She found that item-level scores did not discriminate between 3-, 4-, and 5-year-old TD children, allowing for the use of one cutoff point for all age groups Item-level scoring was more accurate than a vowel scoring method in differentiating children with and without DLD (Kapalková et al., 2013). Furthermore, in Spanish-speaking children, item-level scores have yielded better levels of diagnostic accuracy compared to the PPC scores (e.g., (Guiberson & Rodríguez, 2013; Gutiérrez-Clellen & Simon-Cereijido, 2010; Windsor et al., 2010). Across languages, item-level scores on NWR tasks have sufficiently discriminated children with language impairments from TD peers (Dispaldro et al., 2013; Kalnak et al., 2014; Kapalková et al., 2013; Kazemi & Saeednia, 2017; Roy & Chiat, 2004; Topbaş et al., 2014).To calculate interrater reliability, a second native Palestinian Arabic–speaking SLT independently scored the audio-recorded responses of 25 children (27% of the sample). The intraclass correlation coefficient (absolute) was found to be excellent (intra-class correlation coefficient = .93).

3.3 Results

3.3.1 Analysis 1: Group Differences

All statistical analyses were performed using R Studio software, Version 3.6.3 (R Core Team, 2020). All raw scores were converted to percentages. To address the first research question, we examined the differences in accuracy scores of the TD and DLD groups. Table 3.4 summarizes the overall performance of the two groups on the A-QU-LITMUS-NWRT as well as their scores across nonwords that vary in terms of length, phonological complexity, and wordlikeness.

The dependent variable was NWR accuracy (where “correct” response = 1 and “incorrect” = 0). Given that this is a binary outcome with assumed binominal distribution, data were analysed using mixed-effects logistic regression models (Baayen et al., 2008) with lme4 package (Bates et al., 2015). The independent variables were nonword length (three levels: one, two, and three syllables), the presence of CC (three levels: none, one, and two CCs), wordlikeness (two levels: high word-like and low word-like), and group (two levels: TD and DLD). Age was entered as a covariate. All independent variables were contrast-coded and entered as fixed effects. The model included by-participant and by-item random intercepts (random effects). This was done to account for the non-independence of the data (repeated measures; Baayen et al., 2008). Fitted models were compared in terms of Akaike information criterion (AIC) and Bayes information criterion, with reduced AIC and Bayes information criterion values indicating a better model fit (Tabachnick & Fidell, 2007). This was supplemented by likelihood ratio tests to determine if the inclusion of a predictor significantly improved the model fit (Baayen et al., 2008; Tabachnick & Fidell, 2007).

Table 3.4. *Mean percentages of correct nonwords (with standard deviations) of the TD and DLD groups on the A-QU-LITMUS-NWRT.*

	Group	
	TD	*DLD*
Overall performance	93.61(10.61)	52.22(19.89)***
Nonword length		
One syllable	98.89(4.19)	79.44(23.44)***
Two syllables	95.24(10.49)	53.57(22.02)***
Three syllables	88.17(19)	34(23.43)***
Presence of consonant clusters		
none	94.68(10.29)	66.15(19.75)***
One CC	93.33(11.05)	44(24.36)***
Two CC	88.75(24.54)	23.33(36.51)***
Wordlikenss		
High wordlikeness	98.96(4.77)	80(24.91)***
Low wordlikeness	92.79(11.73)	47.95(20.19)***

Note. TD = typically developing; DLD = developmental language disorder; A-QU-LITMUS-NWRT = Arabic version of the Quasi-Universal LITMUS Nonword Repetition Test; CC = consonant cluster.****p* < .001.

First, we examined whether the inclusion of the random effects structure was permitted. This was done by comparing a baseline generalized linear model without the random intercepts (null model) with a baseline mixed-effects model that only included the random intercepts. Relative to the null model (*AIC* = 2731), the baseline mixed-effects model provided a substantially better fit for the data (*AIC* = 1708, $\chi^2(2) = 1027$, $p < .001$). Therefore, the inclusion of the random intercepts was justified.

Next, we implemented a step-wise step-up procedure for building the mixed-effects model. Age was entered first as a covariate. Next, the predictors: group, nonword length, CCs, and wordlikeness variables were entered into the model, respectively, followed by their interactions. A summary of the model fitting procedure is provided in Appendix E. The fit of the final model (M8) was significantly better than the intercept-only baseline model (*AIC* =

1596, $\chi^2(12) = 1157$, $p < .001$). The output of the final model is presented in Table 3.5. The significance level of the main effects of the fixed factors was obtained using the Anova() function. The estimated marginal means (EMMs) were obtained using the emmeans package (Lenth, 2020), with all pairwise comparisons corrected using Tukey's honesty significant difference adjustment.

There was a main effect of age, $\chi2(1) = 7.24$, $p < .01$. There was a main effect of group, $\chi^2(1) = 114.53$, $p < .001$, with the TD group (*EMM* = 4.42, *SE* = .40) scoring higher than the DLD group on the task (*EMM* = .36, SE = .42, $p < .001$). The Group × Age interaction was not significant, $\chi^2(1) = 1.60$, $p = .207$.There was a main effect of nonword length, $\chi2(2) = 32.72$, $p < .001$, such that three-syllable nonwords (*EMM* = 1. 06, *SE* = .44) were repeated less accurately compared to one-syllable (*EMM* = 3.54, *SE* = .48, $p < .001$) and two-syllable nonwords (*EMM* = 2.58, *SE* = .39, $p < .001$). The difference in the repetition accuracy of one- and two-syllable nonwords was not significant ($p = .106$). The Group × Nonword Length interaction was not significant, $\chi^2(2) = 0.79$, $p = .673$.

There was a significant effect of the number of CCs, $\chi^2(2) = 11.41$, $p < .01$, such that nonwords with two CCs (*EMM* = 2.26, *SE* = .68) were repeated less accurately compared to nonwords with no CCs (*EMM* = 3.22, *SE* = .34, $p < .01$) but were comparable to nonwords with one CC (*EMM* = 2.70, *SE* = .36, $p = .084$). The repetition accuracy of nonwords with no or one CC did not differ significantly ($p = .376$).

The Group × Number of CCs interaction was significant, $\chi^2(2) = 9.98$, $p < .01$. The interaction is illustrated in Figure 3.1, which plots the proportion of correctly repeated nonwords as a function of number of CCs for the TD and DLD groups. It can be observed that, for the DLD group, the repetition accuracy decreased more significantly with an increased number of CCs. This reduction in accuracy appears to be much less pronounced for the TD group. Post hoc comparisons showed that, within the DLD group, nonwords with two CCs (*EMM* = −0.86, *SE* =. 75) were repeated less accurately than nonwords without CCs (*EMM* = 1.48, *SE* = .40, $p < .05$) or with one CC (*EMM* = .46, *SE* = .42, $p < .05$). There was no difference in repetition accuracy of nonwords with one or two CCs ($p = .879$).

Table 3.5. *Parameter estimates of the final logistic mixed-effects model (M8)*

Parameters	β	SE (β)	Z statistic
Fixed Effects			
Intercept	.25	1.24	.20
Age	.05**	.02	2.69
Group: TD (compare with DLD)	3.48***	.42	8.21
Nonword length:2 Syllables (compared with 1 syllable)	-.96*	.46	-2.10
Nonword length: 3 syllables (compared with 1 syllable)	-2.48***	.47	-5.29
CC: 1 CC (compared with no CC)	-1.02**	.35	-2.93
CC:2 CC (compared with no CC)	-2.34***	.71	-3.32
Wordlikeness: low wordlikeness (compared with high wordlikeness)	-1.25*	.52	-2.39
Group X CC interaction			
Group: TD x CC number: 1 CC	1.01**	.32	3.14
Group: TD x CC number: 2 CC	.76	.53	1.44
Random Effects	**Variance**	**SD**	
Participant (Intercept)	2.18	1.48	
Item (Intercept)	.57	.76	
Observations 2730, participants: 90, items: 30			

Note. TD = typically developing; DLD = developmental language disorder; CC = consonant cluster.

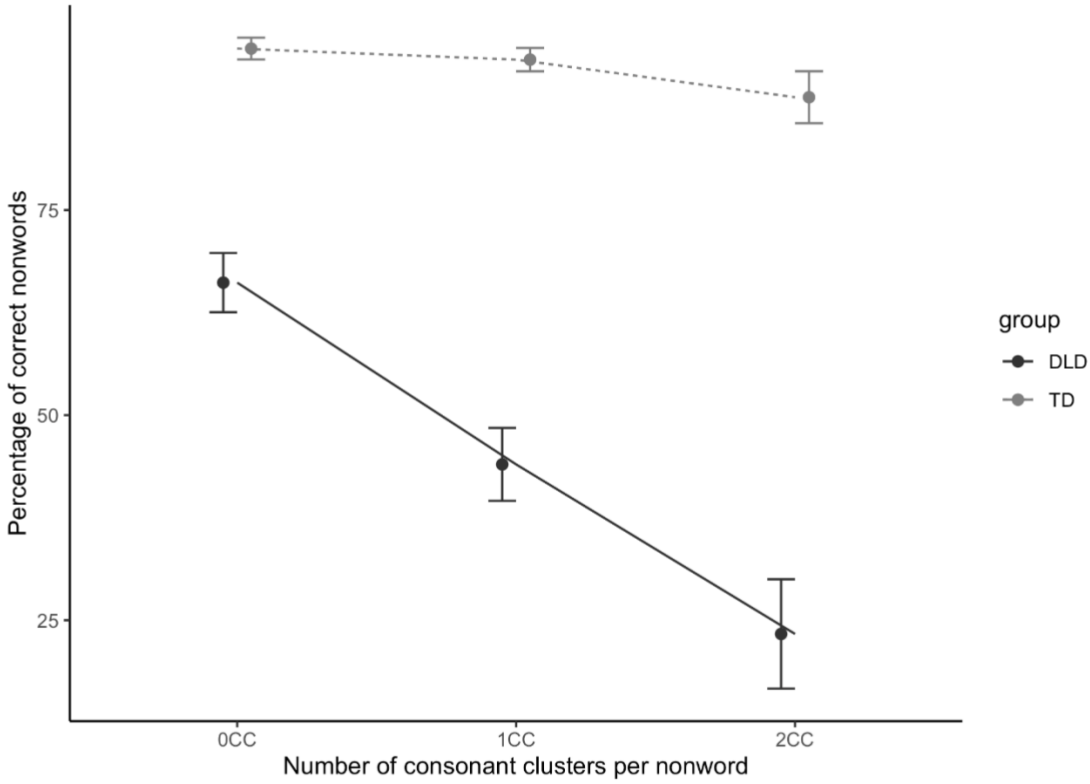

Figure 3.1. *Nonword repetition accuracy across nonwords with different numbers of consonant clusters for the typically developing (TD) children and children with developmental language disorder (DLD).*

Within the TD group, the repetition accuracy of nonwords with two CCs (*EMM* = 3.38, *SE* = .74) was not significantly different to nonwords without CCs (*EMM* = 4.96, *SE* = .43, $p < .433$) or with one CC (*EMM* = 4.94, *SE* = .41, $p = .422$). There was no difference in repetition accuracy of nonwords without CCs and nonwords with one CC ($p = 1$). The TD group outperformed the DLD group in repeating nonwords with one, two, or no CCs (for all comparisons, p < .001). The effect of wordlikeness was significant, $\chi^2(1) = 5.72$, $p < .05$. Highly word-like nonwords (*EMM* = 3.01, *SE* = .55) were repeated more accurately than nonwords that were less word-like (*EMM* = 1.77, *SE* = .32, $p < .05$). Group × Wordlikeness interaction was not significant, $\chi^2(1) = 0.37$, $p = .542$.

3.3.2 Analysis 2: Diagnostic Accuracy of the NWR Task

To address the second research question, we assessed the diagnostic accuracy of the QU-LITMUS-NWRT. Receiver operating characteristic (ROC) curve was generated using the *pROC* package (Robin et al., 2011). ROC curves plot the true positive rate (sensitivity) as a function of false-positive rate (1 – specificity) for every possible cutoff score (Gonçalves et al., 2014). Consequently, the optimal cutoff score with the highest sensitivity and specificity values is determined. Also, the area under the curve (AUC) was computed. AUC is an index of the test classification accuracy, and it reflects the probability that a randomly selected child with DLD will have a lower score than a randomly selected TD child. According to Carter et al. (2016), AUC values range from 0.5 to 1.0. An AUC of 1.0 indicates a perfect test, .90–.99 is an excellent test, .8–.89 a good test, .7–.79 a fair test, and lower than .7 is a non-useful test. Sensitivity, specificity, and LRs were calculated for the final cut-off score.

Figure 3.2 presents the ROC curve for the QU-LITMUS-NWRT using item-level scoring. Based the ROC analysis, the optimum cut-off score was 81.67% (equivalent to a raw score 24 out of 30). The diagnostic accuracy of the cut-off score was excellent: AUC = .99 (95% confidence interval (CI) [0.94, 1]); sensitivity = .93 (95% CI [0.83, 0.10]); specificity = .93 (95% CI [0.87, 0.98]); LR+ = 13.93 (95% CI [5.41, 36.26]); LR− = .07 (95% CI [0.02, 0.27]).

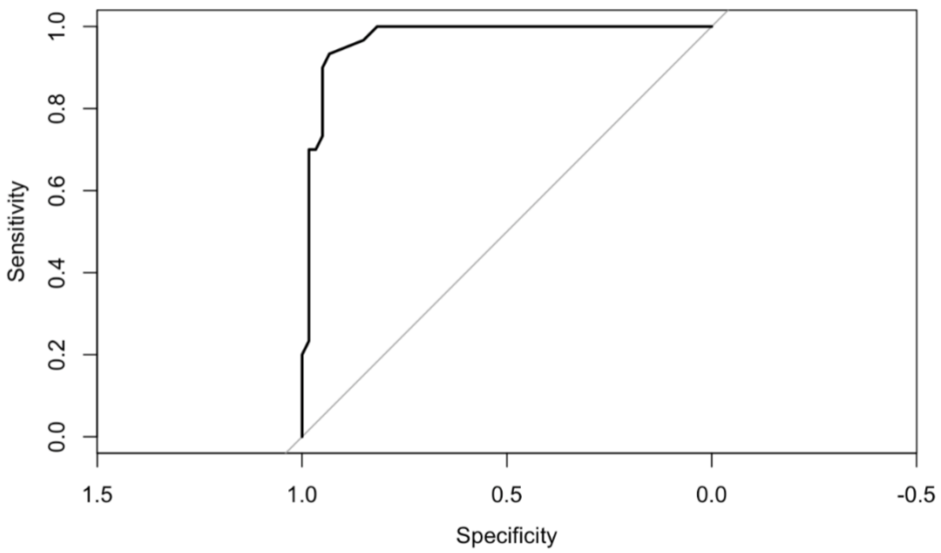

Figure 3.2. *Receiver operating characteristics curve for the item-level scoring method.*

3.4 Discussion

This is the first study to examine the diagnostic accuracy of NWR for the identification of DLD in Arabic. This study found that 4- to 6-year-old Palestinian Arabic–speaking children with DLD performed below the level of age-matched TD controls on the QU-LITMUS-NWRT. Nonword length and wordlikeness ratings appeared to influence NWR accuracy of TD and DLD groups, whereas the presence of CC influenced the NWR accuracy of the DLD group only. The A-QU-LITMUS-NWRT was found to have excellent diagnostic accuracy in distinguishing children with DLD from TD peers, indicating that it is a promising measure that clinicians could include within their assessment battery to establish DLD diagnosis in Arabic-speaking children.

3.4.1 Poor nonword repetition in Arabic-speaking children with DLD

The accuracy scores of the DLD group were substantially lower than those of the TD group on the A-QU-LITMUS-NWRT (52% vs. 93%). This result aligns with existing literature documenting that children with DLD have considerable difficulty in repeating nonwords compared to age-matched TD peers across languages (Ahufinger et al., 2021; Armon-Lotem & Meir, 2016; de Bree et al., 2007; Girbau, 2016; Graf Estes et al., 2007; Kapalková et al., 2013; Pham & Ebert, 2020; Topbaş et al., 2014). Our findings are also consistent with previous studies that showed poor performance of Arabic-speaking children with or at risk of DLD on language-specific NWR tasks (Balilah, 2017; Khater, 2016; Saiegh-Haddad & Ghawi-Dakwar, 2017; Shaalan, 2010). It should be noted that these studies used NWR tests that were language-specific, that is, followed Arabic phonotactics, while, in this study, we used a quasi–language-independent NWR test. The fact that there were significant group differences on the QU-LITMUS-NWRT suggests that the test is as sensitive as language-specific Arabic NWR tests to the language difficulties of Arabic-speaking children with DLD. There was a main effect of age on NWR accuracy in the TD and DLD groups suggesting that scores on the A-QU-LITMUS-NWRT improved with age. The effect of age replicates studies that have reported that older children outperformed younger children on NWR tasks (e.g., Chiat & Roy, 2007; Guiberson & Rodríguez, 2013; Kapalková et al., 2013; Roy & Chiat, 2004; Weismer et al., 2000).

Several item characteristics appeared to influence task performance. For both groups, repetition accuracy decreased as the nonwords increased in length. Accuracy fell significantly for three-syllable nonwords compared to one- and two-syllable nonwords. The nonsignificant Group × Nonword Length interaction suggests that the effect of length on NWR was equivalent across for both groups. This result contradicts studies showing that, as nonwords increase in length, repetition accuracy decreases for TD and, to a greater degree, DLD groups (Archibald & Gathercole, 2006; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Jones et al., 2010; Weismer et al., 2000). Particularly, research shows differences between TD and DLD groups are larger when repeating nonwords of three or more syllables (Archibald & Gathercole, 2006; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Jones et al., 2010; Weismer et al., 2000). The additional disadvantage noted in DLD groups in repeating long nonwords has been explained in the light of a limitation in their phonological short-term memory (e.g., Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990). However, as mentioned above, the developers of the QU-LITMUS-NWRT aimed to limit the effect of length on NWR as their focus was to evaluate the effects of phonological complexity (e.g., presence of CCs) on NWR. Hence, the fact that the test had relatively short nonwords of one, two, and three syllables could have contributed to the lack of interaction between the two variables. Previous research with Gulf Arabic–speaking children has documented similar findings when using an NWR task containing two- and three-syllable nonwords (Shaalan, 2010).

The number of CCs in nonwords seemed to affect the repetition accuracy of the DLD group only. The DLD group repeated nonwords with two CCs less accurately than nonwords with one or no CCs. This is in line with earlier studies showing that nonwords with CCs are more difficult to repeat than nonwords with singleton consonants in children with DLD (Briscoe et al., 2001; Coady & Evans, 2008; Graf Estes et al., 2007; Leclercq et al., 2013; Munson et al., 2005). It is suggested that the increased articulatory complexity of nonwords with CC places higher demands on speech motor output processes since their production involves the coordination of many articulatory movements within syllables. This, in turn, increases the likelihood of articulation errors occurring (Archibald et al., 2013). Given that articulatory control skills were not measured in this study, such a conclusion needs further examination.The TD and DLD groups in our study showed a higher repetition accuracy of high word-like nonwords than low word-like nonwords. This result extends previous research indicating that knowledge stored in long-term memory supports NWR (Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990; Jones et al., 2010; Munson et al., 2005). A nonsignificant interaction between group and wordlikeness ratings revealed that wordlikeness affected both groups similarly, although the scores of the DLD group were lower than those of the TD group on high and low word-like nonwords.

3.4.2 Poor NWR as a Possible Clinical Marker of Arabic DLD

The A-QU-LITMUS-NWRT (dos Santos et al., n.d.) showed an overall excellent diagnostic accuracy in differentiating 4- to 6-year-old, Palestinian Arabic–speaking children with DLD from their age-matched TD peers. ROC analyses using item-level scores revealed that a cutoff score of 81.67% on the task had the best overall classification accuracy (93%). The sensitivity and specificity of the cutoff score were equal to 93%, showing a good value in terms of diagnostic accuracy (Plante & Vance, 1994). These results mean that the A-QU-LITMUS-NWRT correctly identified 28 out of 30 children with DLD as having DLD (sensitivity) and 56 out of 60 TD children as being TD (specificity).

Our findings are in contrast to those of Wallan (2018) who found that a nonword list recall task had inadequate diagnostic accuracy in distinguishing Arabic-speaking children with LCs from their TD peers. The nonword list recall task in Wallan's study correctly identified 89% of TD children but only 56% of the children with LC. The difference in results can be attributed to several reasons. Firstly, in the task used by Wallan, children were asked to repeat a list of up to four nonwords, whereas the A-QU-LITMUS-NWRT used in our study was less demanding as children repeated one nonword at a time.Secondly, the performance of the TD and LC groups on the nonword recall list was approximately similar with both groups showing floors effects in Wallan's study (2018). Out of a maximum score of 4 points, the mean score for the TD group was 1.63 (SD = .47) and 1.16 (SD = .35) for the DLD group. This suggests that the nonword recall task used by Wallan was difficult even for the TD children. In our current study, performance of the TD group was close to the ceiling and significantly higher than the DLD group, showing a large effect size (d = 2.62).

Importantly, none of the children in the LC group (N = 16) in Wallan's study had a confirmed diagnosis of DLD. Although children in the LC may have weaker language skills compared to their TD peers, the level of their language ability might have not been low enough for a DLD diagnosis. On the other hand, the children in our study had a DLD diagnosis and were receiving language intervention at the time of the study. This means that the DLD group in our study may have had more severe language difficulties compared to the LC group in Wallan's (2018) study. The less demanding nature of the A-QU-LITMUS-NWRT compared to the nonword list recall used in Wallan's study and the more stringent criteria for the DLD children recruited for our study may have enlarged the differences between the TD and DLD groups in our study, positively influencing the diagnostic accuracy of the task. We further calculated the LRs for the A-QU-LITMUS-NWRT. The LR+ was 13.93, and the LR− was equal to .07. Based on Dollaghan (2007), values of LR+ ≥ 10.0 and LR− ≤ 0.1 can be interpreted with confidence. Thus, based on the A-QU-LITMUS-NWRT alone, one can conclude that a child who scores below the cutoff (81.67%) may have DLD and a child who scores above it may not. Although the 95% CIs for the LRs include values that fall beyond the threshold mentioned above, they remain within the informative range. This points to the diagnostic value of the A-QU-LITMUS-NWRT for the identification of DLD in Arabic. The finding that NWR has a good level of accuracy in identifying children with DLD and excluding TD children is not trivial. It replicates the existing literature that reported good diagnostic accuracy for NWRTs in identifying children with DLD acquiring typologically different languages (Armon-Lotem & Meir, 2016; Dispaldro et al., 2013; Kalnak et al., 2014; Kapalková et al., 2013; Kazemi & Saeednia, 2017; Thordardottir et al., 2011; Topbaş et al., 2014). The excellent identification accuracy of the A- QU-LITMUS-NWRT and its consistency with the DLD literature provides strong evidence that NWR should be considered as a potential clinical marker of DLD in Arabic-speaking children.

3.4.3 Clinical Implications

Our findings form a stepping-stone into advancing the diagnostic procedures for identifying Arabic-speaking children with DLD in the Palestinian context and other Arab countries where speech and language therapy remains a relatively underdeveloped field. SLTs face difficulty in diagnosing DLD in Arabic due to the poor availability of appropriate language assessments. When examining the language abilities of Arabic-speaking children, the sole reliance on qualitative assessments and/or subjective clinical judgment might not provide sufficient or reliable evidence regarding the presence or absence of DLD. As a result, Palestinian Arabic–speaking children with DLD encounter an increased risk of under-identification and misdiagnosis.This study offers information that can contribute to a more accurate evaluation of Arabic-speaking children with DLD. Our findings show that poor NWR has good discriminatory power in distinguishing between Arabic-speaking children with and without DLD. Consequently, our results highlight the importance of considering NWR abilities besides the informal language measures when diagnosing DLD in Arabic. Particularly, the study highlights the potential of the Arabic version of the A-QU-LITMUS-NWRT as a useful indicator/index of DLD that is quick to administer. Previous Arabic studies showed that children with DLD perform poorly on NWR tasks. An important contribution of our study is that we can specify what the threshold performance should be for a child to be considered for further assessment. For the A-QU-LITMUS-NWRT, a cutoff point of 81.67%, equivalent to a score of 24, could be used to determine whether a child's language abilities need further assessment.The A- QU-LITMUS-NWRT was constructed using early acquired sounds and syllabic structures that are common across all Arabic dialects (Watson, 2002) as well as across many languages (Maddieson, 2006). This means that the use of the test can be extended beyond identifying DLD in monolingual children acquiring Palestinian Arabic to other Arabic dialects. The design of A-QU-LITMUS-NWRT makes it suitable to be used with bilingual children whose first language or second language is Arabic once its diagnostic accuracy in identifying DLD in this population is explored.

3.4.4 Limitations and Future Directions

Although promising, our findings are preliminary and should be interpreted with caution. Our study followed a two-gate design in which preselected TD and DLD groups were recruited. Two-gate designs are very common in diagnostic studies; however, they could lead to a spectrum bias (Pawłowska, 2014; Redmond et al., 2019).Children with DLD in this study were receiving language intervention and may not be representative of Palestinian Arabic–speaking children with DLD in terms of severity. Population-based one-gate designs are needed to validate our results.The diagnostic accuracy of the NWR task should be considered with relevance to the reference standards of DLD employed in this study. The first reference standard was the receipt of speech and language therapy intervention Children with DLD were diagnosed prior to the current study. To verify the DLD status of the children, our second reference standard was poor performance (below 1.5 SDs) on at least two morphosyntactic measures. These tasks only assess expressive morphology, and their use as a reference standard might be limited with children with DLD whose language difficulties do not involve grammar (e.g., semantics). Notably, reference standards that are used to estimate diagnostic accuracy are not interchangeable (Redmond et al., 2019). Hence, if different reference standards are used, the diagnostic accuracy of the current task may vary.

Live administration of the QU-LITMUS-NWRT was engaging for the children. However, live administration could be associated with inevitable variations in rate, pitch, and loudness when the examiner delivered the test to different children. This could have influenced the children's performance in the test. Therefore, future studies should consider the use of audio-recorded stimuli to ensure consistency of delivery of the test.Future studies should examine the nonword repetition errors in Arabic-speaking children with DLD. This will provide more insights into the nature of phonological production difficulties in this group. Although it has been reported that oral motor planning influences NWR performance (e.g., (Archibald et al., 2013), no measures of this ability were taken as part of this study. Future studies on NWR in Arabic should take this measure into account as it could provide us with insights about the underlying cause of NWR difficulties in Arabic-speaking children with DLD. It also needs to be pointed that there was an imbalance between the number of nonwords in the categories of wordlikeness and CCs. Although we reported the significant and insignificant interactions (group and wordlikeness, and group and number of CCs), they are likely to have been conflated with nonword length, which limits the interpretation of the analysis of these interactions.

3.4.5 Conclusion

This study offers valuable implications for the assessment of DLD in Palestinian Arabic–speaking children. Children with DLD were found to perform poorly on the A-QU-LITMUS-NWRT (dos Santos et al., n.d). In the current study, the A- QU-LITMUS-NWRT was found to have high diagnostic accuracy, suggesting that it should be considered as a clinical marker of DLD in Arabic-speaking children aged 4–6 years. The test could be used by SLTs—alongside other language measures—to improve the accuracy of identifying DLD in Arabic. However, the adaptation of the task for clinical use requires further validation of its diagnostic accuracy. The use of one-gate designs incorporating reference standards that cover different language domains will be needed to include a more representative, heterogeneous group of children with DLD.

Acknowledgments

This work was funded by University of Reading International Studentship (GS18-082). We would like to thank the children and their parents who participated in the study. We also thank the teachers and speech and language therapists who facilitated the recruitment of the children.

Chapter 4: Sentence Repetition as a Clinical Marker of Developmental Language Disorder: Evidence from Arabic

Abstract

Purpose: Research on the typical and impaired grammatical acquisition of Arabic is limited. This study systematically examined the morpho-syntactic abilities of Arabic-speaking children with and without Developmental Language Disorder (DLD) using a novel sentence repetition task. The usefulness of the task as an indicator of DLD in Arabic was determined.

Methods: A sentence repetition task was developed in Palestinian Arabic (LITMUS-SR-PA-72) and administered to 30 children with DLD (*M* = 61.50 months, *SD* = 11.27) and 60 age-matched TD children (*M* = 63.85 months, *SD* = 10.16). The task targeted grammatical structures known to be problematic for Arabic-speaking children with DLD (language-specific) and children with DLD across languages (language-independent). Responses were scored using binary, error and structural scoring methods.

Results: Children with DLD scored below TD children on the LITMUS-SR-PA-72 in general, as well as in the repetition of language-specific and language-independent structures. The frequency of morpho-syntactic errors was higher in the DLD group relative to the TD group. Despite the large similarity of the type of morpho-syntactic errors between the two groups, there were some atypical errors exclusively produced by the DLD group. The three scoring methods showed good diagnostic power in the discrimination between children with and without DLD.

Conclusion: Sentence repetition was an area of difficulty for Palestinian Arabic-speaking children with DLD. The DLD group demonstrated difficulties with language-specific and language-independent structures, particularly complex sentences with non-canonical word order. Most grammatical errors made by the DLD group resembled those of the TD group and were mostly omissions or substitutions of grammatical affixes or omissions of function words. SR appears to hold promise as a good indicator for the presence or absence of DLD in Arabic. Further validation of these findings using population-based studies is warranted.

4.1 Introduction

Developmental language disorder (DLD) is a condition where the child has significant impairment in understanding and/or using spoken language, such that it impairs everyday social functioning and educational progress; this difficulty is not associated with an obvious cause and is likely to persist beyond childhood (Bishop et al., 2017). Research has focused on identifying the psycholinguistic phenotypic markers that are characteristically associated with DLD and can be used as indicators of the disorder (e.g., Rice & Wexler, 1996). These can either be a) distinct grammatical behaviors that are observed in spontaneous and elicited language, for example deficits in marking verb tense and agreement in English (e.g., Ash & Redmond, 2014) and omission of articles and object clitics in Spanish and Italian (e.g., Guasti et al., 2016; Jackson-Maldonado & Maldonado, 2017), or; b) poor performance on language-based processing tasks such as non-word (see Chiat, 2015) and sentence repetition (see Marinis & Armon-Lotem, 2015).

Sentence repetition (SR) tasks have gained traction as reliable screening measures for identifying DLD in monolingual and bilingual children in different languages (e.g., Armon-Lotem & Meir, 2016; Conti-Ramsden et al., 2001; Fleckstein et al., 2018). To date, little is known about the usefulness of SR in identifying DLD in Arabic. This study investigates the morpho-syntactic abilities of Palestinian Arabic-speaking children with DLD and their typically developing (TD) peers using a novel SR task. First, we compare the two groups on accuracy and error patterns in the repetition of grammatical structures known to be problematic for children with DLD acquiring Arabic and other languages. Then, we assess the accuracy of SR for discriminating Palestinian Arabic-speaking with DLD from TD peers.

4.1.1 Sentence Repetition as a measure of morpho-syntactic abilities

The exact mechanisms underlying SR have been debated. The central question has been whether performance on SR tasks reflects linguistic knowledge (Klem et al., 2015; Polišenská et al., 2015), or memory capacity (e.g., Alloway & Gathercole, 2005). Early accounts proposed that, if sentence length exceeds the individual's immediate memory, repetition of the stimulus will involve linguistic representations in long-term memory in addition to short-term memory. Such repetitions are suggested to be filtered through the individual's productive linguistic system (Slobin & Welsh, 1973). This view was supported by later studies suggesting that, when a sentence is long enough to tap into the individual's grammatical system, grammatical reconstruction takes place. Thus, after hearing a sentence, individuals use recently activated lexical items to create a conceptual message of the sentence in short-term memory to regenerate the sentence using morpho-syntactic representations they are holding in long-term memory (Potter & Lombardi, 1990, 1998). Short sentences, however, are imitated in a parrot-like fashion, exclusively relying on short-memory rather than linguistic competence (Vinther, 2002).

Conversely, Riches (2012) proposed that the roles of short- and long-term memory in SR are not length-dependent, but they work effectively together at all sentence lengths. This is supported by evidence showing that, when the sentence length is constant, increasing syntactic complexity of sentences results in a greater number of errors in SR (Frizelle & Fletcher, 2014; Kidd et al., 2007; Riches et al., 2010). Also, Riches (2012) found that the best predictor of SR was syntactic knowledge as indexed by a priming task. Similarly, Polišenská et al., (2015) have suggested that SR is more dependent on morpho-syntax and lexical phonology, and less so on semantics or prosody. Together, these findings support the view that SR is a measure of underlying syntactic competence. It is generally agreed that children find it difficult to imitate structures that they do not know (Devescovi & Caselli, 2007) and that there is an overlap between SR errors and errors made in spontaneous contexts (Riches, 2012). This makes SR a valuable tool for evaluating grammatical structures that might not otherwise be present in spontaneous speech (Seeff-Gabriel et al., 2010) and in characterizing the typical and impaired acquisition of linguistic structures in a given language.

4.1.2 Diagnostic accuracy of sentence repetition tasks

The quality of a clinical marker as an indicator of the presence or absence of DLD can be determined based on diagnostic accuracy metrics. Sensitivity refers to the proportion of children with the disorder (i.e., with DLD) correctly identified by the task, and specificity refers to the proportion of children without a disorder (i.e., TD) correctly identified by the task. Plante & Vance (1994) recommend that Sensitivity and Specificity values of 90% and above indicate good classification accuracy of the test, values of 80% to 89% indicate fair diagnostic accuracy and values below 80% indicate unacceptably high rates of misidentification. Alternative measures of diagnostic accuracy include positive likelihood ratio (LR+) i.e., the probability of being correctly identified as having DLD if the child has DLD, and Negative Likelihood Ratio (LR-), i.e., the probability of being correctly identified as unimpaired if the child has typical language

(Sackett et al., 1991). Likelihood ratios have an advantage over sensitivity and specificity because they are less likely to change due to variations in the prevalence of the disorder (Dollaghan & Campbell, 1998). Dollaghan (2007) suggested that values of LR+ ≥ 10.0 and LR- ≤ 0.1 indicate that the test can indicate, with confidence, the presence or absence of the disorder, while values of LR+ ≥ 3.0 and LR+ ≤ 0.3 indicate that the test is suggestive but insufficient to rule in or rule out the disorder, and values of LR+ <3.0 and LR- > 3.0 indicate the test does not discriminate between presence or absence of the disorder.

SR has been shown to be a reliable clinical marker of DLD in English-speaking children (for a review see Pawłowska, 2014). Conti-Ramsden et al., (2001) found that SR, compared to a third-person singular task, past tense marking and nonword repetition, was the most accurate in identifying English-speaking children with DLD aged 10;5 to 11;1 years old, with sensitivity and specificity values of 90% and 85%, respectively. More recently, Redmond et al. (2019) revealed that SR discriminated 7-year-old English-speaking children with and without DLD with sensitivity and specificity values greater than 80%, indicating the potential of the task as a diagnostic tool for DLD (Redmond et al., 2019). Several studies have examined the diagnostic accuracy of SR in identifying children with DLD who speak languages other than English (for a summary, see Table 4.1). The sensitivity and specificity of SR tasks in most cross-linguistic studies varied between 80% and 90% indicating fair to good levels of accuracy in discriminating between children with and without DLD (for a review, see Rujas et al., 2021).

Table 4.1. *Summary of diagnostic accuracy of sentence repetition tasks in identifying DLD across languages*

Reference	Language	*TD*		*DLD*		Sensitivity %	Specificity %	LR+	LR-
		N	Age in years	N	Age in years				
Armon-Lotem & Meir (2016)	**Hebrew**	38	6 (.17)	14	6;1 (.33)	100[a]	87	7.60[b]	0
Armon-Lotem & Meir (2016)	**Russian**	20	6;1 (.17)	14	5;10 (.25)	86	90	8.57	.16
Christensen (2019)	**Danish**	37	7;9 (1.5)	16	7;9 (1.1)	94	97	34.7	.06
Christensen (2019)	**Danish**	50	12;5 (.8)	11	12;3 (1.1)	91	98	45.5	0.09
Stokes et al (2006)	**Cantonese**	15	4;1 – 6;9	14	4;2 – 5;7	77	97	25.66	.24
Pham & Ebert (2020)	**Vietnamese**	194	5;8 (.4)	10	5;5 (.3)	90	71	3.13	.14
Thordardottir et al (2011)	**French**	78	4;1 - 5;11	14	4;6 - 5;11	92	86	6.46	.09
Leclercq et al (2014)	**French**	34	10.2 (1.4)	34	9.11 (1.2)	97	88	8.08	.03
Theodorou et al (2017)	**Cypriot Greek**	22	4;5-8;7	16	4;11-8;1	75	82	4.11	.3

Note. **TD** = Typically Developing. **DLD =** Developmental Language Disorder. **LR+** = Positive Likelihood Ratio. **LR-** = Negative Likelihood Ratio.

4.1.3 DLD in Arabic: characteristics of morpho-syntactic deficits

Arabic-speaking children with DLD have difficulties with verb morphology production (e.g., Abdallah & Crago, 2008; Fahim, 2017; Taha et al., 2021b). In a recent study, Taha et al (2021b) reported that 4 to 7-year-old Palestinian Arabic-speaking children with DLD were significantly less accurate than age-matched TD peers in producing the following forms: past tense masculine singular verbs (e.g., *daras*, study-PAST-3MS, "he studied"), past tense feminine singular morpheme *–at* (e.g., *daras**at***, study-PAST-3FS, "she studied"), past tense plural morpheme *–u* (e.g., *daras**u***, study-PAST-3PL, "they studied"), present tense masculine singular morpheme *byi–* (e.g., ***byi**drus*, study-PRES-3MS, "he is studying"), present tense feminine singular morpheme *bti–* (e.g., ***bti**drus*, study-PRES-3FS, "she is studying") and the present tense plural circumfix morpheme *byi—u* (e.g., ***byi**drus**u***, study-PRES-3PL, "they are studying"). The tense errors of the DLD group resembled the use of finite (i.e., wrong tense) or non-finite/tenseless forms (i.e., imperative and imperfective verbs) in place of the correct tense. The pattern of subject-verb agreement errors comprised of the use of the singular verbs in place of the plural verbs and the use of the masculine verbs in place of the feminine verbs.

Compared to age-matched TD children, Arabic-speaking children with DLD exhibit difficulties with inflecting Arabic noun plurals (Abdallah et al., 2013; Fahim, 2005; Shaalan, 2010). This includes the use of suffixes for the regular masculine sound plural (MSP; e.g., the suffix ***-:in*** as in *najjari**:n*** "carpenters") and feminine sound plurals (FSP; e.g., the suffix ***-a:**t* as in *ward**a:t*** "flowers"), and the use of irregular broken plural forms (BP; e.g., *dafadiʕ "frogs").* Analysis of error patterns revealed that children with DLD tended to either use a singular noun instead of the plural form (e.g., *tawala* "table" for *tawl**a:t** "tables"*), or a non-morphological form such as a periphrastic expression of number (e.g. *tamanja kalb* "eight dog" for *klab "*dogs") or quantifiers (e.g. *kteer arnab* "many rabbit" for *aranib* "rabbits"; Abdallah et al., 2013).

Another characteristic of DLD in Arabic is the omission of bound pronouns (Abdallah, 2002; Faquih, 2014; Shaalan, 2010). Using an elicitation task, Faquih (2014) found that the production of bound pronouns is impaired in Hijazi Arabic-speaking children with DLD aged 3;2 to 6;9 years compared to TD children. Specifically, Faquih (2014) reported that only a few children in the DLD group produced third person masculine possessive pronouns (e.g., *ktab**o***, book-POSS-3MS, "his book") and feminine singular possessive pronouns (e.g, *ktab**ha***, book-POSS-3FS, "her book") and failed to produce any third-person plural possessive pronouns (e.g, *kutub**hum***, book—BP-POSS-3PL, “their book”). Errors made by the DLD group were characterized by pronoun omission, or substitution of a bound pronoun with the free possessive pronoun /ħag/ “mine” (Faquih, 2014).

As examples 1a and 1b illustrate below, Arabic has a flexible word order where VSO and SVO structures are commonly used (Mohammad, 2000). Through syntactic movement, the object could be moved to a pre-verbal position resulting in an OVS or OVS structure. One can add an object clitic to the verb to refer to the fronted object (see 1c). This process is called clitic left dislocation (CLD; Lalami, 1996). The production and comprehension of sentences with CLD are reported to be challenging for Qatari Arabic-speaking children with DLD. Shaalan (2010) found that children with DLD scored significantly lower on sentences with CLD than age-matched TD peers.

1. (a) *biyakul il-walad buza* [VSO]
   Eat-PRES-3MS the-boy ice-cream
   “The boy is eating ice-cream”
   (b) *il-walad biyakul buza* [SVO]
   The-boy eat-PRES-3MS ice-cream
   “The boy is eating ice-cream”
   (c) *buza akalha il-walad* [$OV_{cl}S$]
   Ice-cream eat-PAST-3MS-CL-3FS the-boy
   “it is the Ice-cream, the boy ate ”

Shaalan (2010) also showed that Qatari Arabic-speaking children with DLD aged 4;10-8;11 years scored significantly lower than age- and language-matched TD peers when repeating subject relative clauses, suggesting that subject relatives may pose a difficulty for Arabic-speaking children with DLD. The task included only one object relative clause, and although the DLD group repeated this item less accurately (35%) than the TD group (77%), more evidence is needed to determine whether this form is problematic for Arabic-speaking children with DLD. Examples of subject and object relatives in PA are provided in examples 2a and 2b, respectively.

2. (a) *hay il-binit illi ʃafat il-arnab* [subject relative]
   This the-girl that see-PAST-3FS the-rabbit
   “This is the girl that saw the rabbit”
   (b) *hay il-bisse illi il-sulħafa 3adatha* [object relative]

This the-cat that the-turtle bite-PAST-3FS-RES-3FS

"This is the cat that the turtle bit"

Recently, Wallan (2018) developed two SR tasks: A novel SR targeting grammatical structures in Arabic, and an Anomalous Sentence Repetition (ASR) test including sets of semantically anomalous and syntactically anomalous sentences. The tasks were administered to a group of Najdi Arabic-speaking TD children between 2;6 and 5;11 years of age and a group of children with reported language concerns (LC). The LC group performed poorly on the SR and the ASR tasks relative to age and non-verbal IQ matched TD children. Wallan (2018) also found that the SR task correctly identified 81% of children with LC and 93% of TD children. Although the SR had a good level of accuracy in discriminating children with and without LC, the results should be considered in light of the study caveats. None of the children in the LC group was clinically assessed, or had a confirmed diagnosis of DLD. It is unclear whether the language difficulties of the LC group were associated with other co-morbidities or differentiating conditions (e.g., hearing loss) which could have contributed to the poor performance on the SR tasks. Thus, the diagnostic accuracy of SR in identifying DLD in Arabic remains unknown.

4.1.4 The present study

Although existing findings on the morpho-syntactic difficulties in Arabic DLD are informative, they remain preliminary. Most of the studies included small numbers of children with DLD (e.g., *N* = 14 in Faquih, 2014 and Taha et al., 2021b; *N* = 12 in Abdallah et al., 2013; *N* = 10 in Abdallah & Crago, 2008). In some studies, the number of items used to examine the target grammatical structures was very limited (e.g., object relatives and passives in Shaalan (2010)'s study were only assessed using one item each). While two studies have shown that Arabic-speaking children with language impairment (as a group) perform poorly on SR tasks (Shaalan, 2010; Wallan, 2018), the diagnostic accuracy of the task in discriminating between children with and without DLD at the individual level is yet to be established.

There is a scarcity of norm-referenced tests that are available in Arabic (see ELO-L for Lebanese Arabic, Zebib et al., 2019 and ALEF for Gulf-Arabic, Rakhlin et al., 2021). In Arabic-speaking contexts, speech and language therapists (SLTs) rely on informal assessment tasks (i.e., parental interview, language sample analysis) to establish DLD diagnosis. Thus, diagnostic decisions are not always consistent and vary according to the subjective judgment and clinical experience of the SLTs. Consequently, Arabic-speaking children with DLD continue to be at risk of being under/misdiagnosed. Tasks with good discriminatory power are needed to help facilitate the effective and efficient identification of DLD in Arabic. Accordingly, this study aims to examine the potential of SR as a clinical marker of DLD in Arabic-speaking children. We specifically address the following questions:

1. How do Arabic-speaking children with DLD compare to TD children in terms of their performance accuracy on SR?
2. How do Arabic-speaking children DLD compare to TD children in terms of the quantity and quality of their grammatical errors in SR?
3. What is the diagnostic accuracy of the SR for the identification of DLD in Arabic?

We predict that the scores of the DLD group on the SR task will be significantly lower than those of TD children. Based on Riches (2012)'s findings that errors in SR corresponded to errors made in other production tasks, we expect the morpho-syntactic errors made by the DLD group to mirror those reported in the Arabic literature for children with DLD in elicited or spontaneous language samples. We also predict that the SR task will show good accuracy in differentiating between children with and without DLD.

4.2 Methods

4.2.1 Participants

This study received approval from Reading University Ethics Committee. A total of 90 monolingual Palestinian Arabic-speaking children aged 4;0 to 6;10 years were recruited from Ramallah city, Palestine. According to a parental report, none of the children had a history of hearing loss, or cognitive, motor, behavioral or neurological impairments. See Table 3.2 for demographic information.

There were 30 children (22 boys, 8 girls) with DLD aged between 4;0 and 6;10 years (*M* = 61.50 months, *SD* = 11.27) recruited through five private speech and language therapy clinics. These children received a diagnosis of DLD by qualified SLTs and were enrolled in language intervention sessions at the time of the study. Given that the DLD diagnosis was based on informal assessments, it was imperative to confirm that these children met the criteria for DLD (Bishop et al., 2016, 2017). Screening of each child's clinical reports was done to confirm that they had: (1) Language difficulties affecting one or more language aspects (children with expressive phonological difficulties were included only if they also had difficulties in other language domains e.g., morpho-syntax, semantics); (2) Passed hearing tests, and; (3) Had language disorder that was not associated with any differentiating conditions (e.g., neurological or genetic disorders). There were 60 TD children (33 boys, 27 girls) aged 4;0 to 6;8 years (*M* = 63.85 months, *SD* = 10.16). They were recruited through three kindergartens and two schools. The additional inclusion criteria for this group were: 1) No parental concerns about the child's current language skill, and; 2) No history of language delay or intervention. Each TD child was within two months of age of a child with DLD. The two groups were matched on chronological age ($t(53.04) = -.96$, $p = .34$, $d = .22$) and did not differ in their non-verbal abilities as measured by the Colored Progressive Matrices (CPM, Raven, 2007; ($t(51.59) = -1.26$, $p = .214$, $d = 0.29$).

A battery of standardized language tasks was administered to all children to confirm their diagnostic status. The tasks examined language areas known to be problematic for Arabic-speaking children with or at risk of DLD. The tasks included: *a) Arabic Verb Elicitation Test (AVET)*: a picture-naming task which examines the production of verb tense and agreement inflections, *b) Arabic Noun Pluralization Test (ANPT)*: an elicitation task that examines the production noun plural types, *c)* Arabic version of the Quasi-Universal LITMUS *Nonword Repetition Test* (*A-QU-LITMUS-NWRT*; dos Santos et al., n.d): the task examines the repetition of nonwords with minimal language-specific features. Additionally, we calculated the d) *Mean Morpheme per Utterance (MPU*). MPU is a measure of a child's grammatical ability level in Semitic languages (Dromi & Berman, 1982). A narrative sample was obtained for each child using *Frog, where are you* story (Mayer, 1969) and the first 100 utterances were transcribed. MPU scores were calculated according to guidelines adopted by Shaalan and Khater (2006) for Arabic. The *MPU* is derived by diving the total number of morphemes by 100 i.e., the number of utterances produced in the narrative task. The results of the TD group (mean and standard deviation) were used to obtain z scores for all participants. All children with DLD scored at or below -1.5 SD below the mean on at least three of the language measures. All TD children scored above the −1.5 SD cutoff point on at least three language measures. The raw and standardized scores of the TD and DLD groups on the language measures are presented in Table 4.2. The average raw scores of the DLD group were significantly below those of the TD group on the AVET ($t(31.67) = -9.98$, $p < .001$, $d = 2.52$), the ANPT ($t(84.58) = -12.56$, $p <.001$, $d = 2.58$), QU-LITMUS-NWRT ($t(37.23) = -10.73$, $p < .001$, $d = 2.62$) and MPU ($t(72.49) =-11.28$, $p < .001$, $d = 2.42$). See Appendix C for Individual raw scores on these measures.

Table 4.2. *A summary of the raw and z scores of the TD and DLD groups on the background measures*

Measures	Group							
	TD				*DLD*			
	Raw scores		*Z scores*		*Raw scores*		*Z scores*	
	M (SD)	Range	M (SD)	Range	M (SD)	Range	M(SD)	Range
A-LITMUS-NWR (Out of 100)	93.79 (10.47)	40 – 100	0.02 (.99)	-5.05 – -60	52.16 (19.91)	3.33 – 86.67	-3.90 (1.87)	-8.50 – -.65
AVET (Out of 100)	96.63 (5.81)	73.96 –100	0(1)	-3.90 – . 58	60.83 (19.21)	14.58 – 89.58	-6.16 (3.31)	-14.12 – -1.21
ANPT (Out of 100)	74.67 (24.68)	20 – 100	0(1)	-2.22 – 1.03	21.99 (14.97)	0 – 73.33	-2.14 (.61)	-3.03 – -.05
MPU	5.35 (.97)	3.15 – 7.48	0(0)	-2.27 – 2.20	3.25 (.75)	1.89 – 4.61	-2.17 (.78)	-3.57 – -.76
CPM (Out of 36)	15.89 (3.68)	9 – 23	0(1)	-1.87 – 1.94	14.76 (3.99)	9 – 23	-.30 (1.09)	-1.87 – 1.94

Note. TD = Typically Developing. DLD = Developmental Language Disorder. ***A-QU-LITMUS-NWR =*** Arabic version of the Quasi-Universal LITMUS Nonword Repetition Test (dos Santos et al., n.d). ***AVET*** = Arabic Verb Elicitation Test. ***ANPT*** = Arabic Noun Plurals Test. MPU = Mean Morpheme per Utterance***.*** ***CPM*** =Colored Progressive Matrices (Raven, 2007).

4.2.2 Sentence Repetition task

The SR task was designed following the principles of the COST Action IS0804 "Language Impairment Testing in Multilingual Settings" (LITMUS; Armon-Lotem et al., 2015). According to Marinis and Armon-Lotem (2015), SR tasks should include grammatical constructions that are vulnerable for children with DLD in the target language (i.e., language-specific), as well as syntactically complex structures which are problematic for children with DLD across languages (i.e., language-independent structures). Based on the available research on DLD in Arabic, the language specific-structures were: tense and verb agreement morphology (Abdallah & Crago, 2008; Fahim, 2017; Taha et al., 2021b), noun plural morphology (Abdalla et al., 2013) and bound possessive pronouns (Faquih, 2014). The language-independent structures were syntactically complex sentences and included: passives, sentences with clitic left dislocation, object Wh-questions, subject and object relative clause, sentences with subordination and conditionals. Additionally, the task included bi-clausal sentences with coordination and complementizes which were syntactically simple control structures matching the syntactically complex sentences (i.e., language-independent) in length (Marinis & Armon-Lotem, 2015).

According to LITMUS-SR guidelines, sentences should be grouped into levels according to their length and syntactic complexity. Essentially, language-specific structures were assessed using syntactically simple (e.g., SVO structure) and short sentences (average of 8 syllables). The language-specific targets emerge early in development and are evident in the language of 4-year-old Arabic-speaking TD children (e.g., Abdallah et al., 2013; Abdu & Abdu, 1986; Al-Akeel, 1998; Aljenaie, 2000; Omar, 1973; Ravid & Farah, 1999). Hence, all language-specific structures were included in level 1. No data was available on the acquisition of the language-independent structures in Arabic. Therefore, the assignment of these structures to levels of difficulty followed the design of other LITMUS-SR tasks. This was done to ensure that our task was comparable to other SR tasks in other languages.

The initial version of the task was piloted with an additional group of 13 monolingual Palestinian Arabic-speaking TD children aged 4;1 to 6;5 years (*M* = 62.4 months, *SD* = 7.44). These children were not included in the main TD group of this study. Pilot findings revealed that the repetition accuracy of the target structures ranged from 54% to 100%. The average repetition accuracy differed significantly across levels ($F(1, 43) = 41.38$, $p < .001$), whereby the average accuracy of repeating level 1 structures (*M* = 94.16, *SD* = 5.18) was significantly lower than that of level 2 (*M* = 76.67, *SD* = 11.22) and Level 3 (*M* = 68.83, *SD* = 18.43); for all comparisons, $p < .05$). These results confirmed that the levels of the task were increasing in difficulty. Although conditional sentences were difficult for the TD children, we decided to retain these items as their repetition accuracy (*M* = 54.81, *SD* = 13.96) was above chance level, suggesting that these structures are not yet acquired but are emerging.

The vocabulary (verbs, nouns and adjectives) used in the task was limited to early acquired words and selected from children's story books. Age of acquisition data in Lebanese Arabic was available for only 52 of the words used in the task (Łuniewska et al., 2019; see Appendix G). As an additional measure, the list of words was judged as being appropriate for pre-school age children by five kindergarten teachers: all words included in the test received an overall agreement score of 80% or above as being familiar to pre-school age children.

The final version of the Palestinian Arabic LITMUS sentence repetition task (LITMUS-SR-PA-72) consisted of 72 sentences. The task examined a total of 13 structures (20 sub-structures). The structures were classified into three levels of increasing difficulty, while each level contained 24 sentences. All language-specific structures were included in level 1: past tense, present tense, noun plurals and bound possessive pronouns. Levels 2 and 3 included language-independent structures. Level 2 contained movement-derived structures such as passives, object Wh-questions and sentences with clitic left dislocation and control structures (bi-clausal sentences with coordination and complementizes). Level 3 included structures with embedding and these were conditionals, subordinate sentences and subject relatives. We also included object relatives which involve both movement and embedding. The order of sentences within each level was pseudo-randomized so that there were no two consecutive sentences of the same structure (for the full list of items, see Appendix F). The sentences varied in length from 3 to 7 words and 7 to 15 syllables. Moreover, there was a significant difference in length across the levels ($F(1,70) = 60.06$, $p < .001$); sentences in Level 2 (*M* = 10.83 syllables, *SD* = 2.32) and level 3 (*M* = 11.75 syllables, *SD* = 1.45) did not differ significantly in length ($p = .18$), but were significantly longer than sentences in level 1 (*M* = 7.92 syllables, *SD* = .88, $p < .001$).

4.2.3 Procedures

Each child was tested individually in a quiet room in the kindergarten, school or speech and language therapy clinic they attended. The children were participating in a larger research project and were assessed using a battery of tests across two one-hour sessions. In the first session, CPM, a narrative task, ANPT and LITMUS-SR-PA-72 were administered; in the second session, QU-LITMUS-NWRT and AVET were administered. Testing was performed by the first author who is a qualified SLT and native speaker of Palestinian Arabic. The administration of the LITMUS-SR-PA-72 followed the procedures suggested by Marinis and Armon-Lotem (2015). Live voice was used given the young age of the participants, as presenting sentences with live voice makes the task more engaging for the children (e.g. Devescovi & Caselli, 2007; Frizelle et al., 2017; Gavarró, 2017) and allows the examiner to build a better rapport with them. The live presentation of the task is more clinically relevant: sentence repetition tasks within standardized language tests are presented live by clinicians (e.g., Newcomer & Hammill, 2008; Seeff-Gabriel et al., 2008; Wiig et al., 2013). To achieve a consistent presentation of the task for all participants, the examiner practised reading the sentences at an average speed. Sentences were presented according to their level of difficulty, with sentences in level 1 being presented first, then sentences in levels 2 and 3, respectively. The LITMUS-SR-PA-72 task was introduced using a tower building game. Children were given a bucket full of coloured blocks. They were instructed to listen carefully to each sentence and to repeat it verbatim. Two practice sentences preceded the task and the child was given feedback on their repetitions to ensure their understanding of the task. The examiner read each sentence individually and only once. The sentence was read again if the child did not hear it due to ambient noise, or if being distracted. After each repetition of the experimental sentences, the child was verbally praised (e.g., good job) and was allowed to add a block to the tower. The task took approximately 20 minutes. Responses were audio-recorded for later transcription and scoring.

4.2.4 Scoring

The responses of the children on the *LITMUS-SR-PA-72* task were transcribed orthographically, coded and scored offline using different scoring systems as follows:

- Binary scoring: The child received a score of 1 if their repetition is identical to the target sentence and a score of 0 if their repetition contained any omission, substitution or addition of words and/or affixes of the target sentence. The maximum total binary score was 72.

- **Error scoring:** A 0 to 3 scoring scheme was employed based on the number of errors observed in the child's repetition. Identical repetitions of the target sentence were assigned a score of 3, repetitions containing one error were assigned a score of 2, repetitions containing two to three errors were assigned a score of 1 and repetitions containing 4 or more errors were assigned a score of 0. This yielded a maximum score of 216.
- **Structure scoring:** This method was based on whether or not the child maintained the grammatical structure targeted by the sentence. Repetitions containing the target grammatical affix or morpho-syntactic structure received a score of 1 whereas repetitions in which the target grammatical structure was omitted, substituted or changed were considered incorrect and received a score of 0. Compared to the binary and error scoring methods, structural scoring was more lenient as the child's repetition was not penalized for errors that did not affect the structure targeted by the sentence (i.e., lexical substitutions). In all scoring methods, phonological errors that were consistent with the child's speech were not considered errors. Dialectal variations in the repetition of words were also disregarded (e.g., *ke:ka* for *kaʕke).* If the child self-corrected and provided more than one response, their final response was scored irrespective of its accuracy. Errors were disregarded if they did not affect the sentence grammatical structure and included: the use of a shortened form of the word (e.g., *ʕa* for *ʕala* "on"), and the omission of the relative *illi* from relative clauses. Both errors did not affect the grammatical structure or meaning of the sentences.

4.2.5 Error analysis

The error analysis was applied to sentences that were ungrammatical; i.e., received a structural score of 0. While we did not have pre-defined error categories, error description was provided and the resulting structure was determined for each ungrammatical sentence. Example 3 below illustrates the scoring and error coding methods. When repeating item 46, the child omitted the Wh-word *mi:n* which is essential for the formulation of the target object-Wh question and also omitted the relative pronoun *illi*. Given that the repetition deviated from the target sentence, it received a binary score of 0. There were two omission errors hence the error score was 1. Finally, the child failed to repeat the object-Wh structure correctly, and so, the structural score was 0. In this case, the morpho-syntactic error category would be a combined omission of Wh-word and relative *illi* leading to a change of structure (sentence with clitic left dislocation).

Item 46	*mi:n il-be:bi illi taʕmato mama?* [object-Wh question] Who the-baby that feed-PAST-3FS-RES-3MS mom? "Who is the baby that mommy fed?"
DLD–3	*il-be:bi taʕmato mama* The-baby feed-PAST-3FS-CL-3MS mom? "..the baby mom fed (it)?"
Binary score	0
Error score	1
Structural score	0
Error type	Omission of Wh word *mi:n* and relative pronoun *illi* (1)
Actual production	Sentence with left clitic dislocation

4.2.6 Reliability

A second Arabic-speaking SLT independently scored 22% of the data (7 DLD and 11 TD). The intra-class correlation coefficient (ICC; absolute) indicated a high inter-rater reliability for the binary (ICC = .98), error (ICC = .91) and structural scoring methods (ICC = .89). Within each level, items of each grammatical structure were equally divided across odd and even items. The odd-even split-half reliability was determined, and the resulting Spearman-Brown coefficient was .96. Furthermore, the Cronbach's alpha for all test items was valued at .985. Both values indicate that the LITMUS-SR-PA-72 had a satisfactory-level of internal consistency reliability.

4.2.7 Statistical analyses

All statistical analyses were run using R software (version 4.0.3; R Core Team, 2020). Raw scores were used for first and second analyses (to address the first and second research questions respectively). Percentage scores were used for the third analysis to address the third research question.

To address the first research question, accuracy scores of the TD and DLD groups on the task were compared. A series of Generalized Linear Mixed Models (Baayen et al., 2008) were fitted to the data using the *lme4* package (Bates et al., 2015). The dependent measure was the accuracy of the grammatical structure of each of the child's repeated sentences. This was a binomial categorical variable (Two levels: 1 = correct; 0 = incorrect). We entered age and sentence length as covariates. The predictors were group, level, target structure, and their interactions. A step-wise-step up procedure was followed for building the mixed-effects models. The random effects were determined initially. First, we included by-participant and by-item random intercepts. This was done to account for the non-independence of the data (repeated measures; Baayen et al., 2008). The addition of random slopes of the within-subject variables was considered as recommended by Barr et al. (2013). However, their inclusion led to model non-convergence. Hence, the models did not include any random slopes. We compared a baseline generalized linear model without random effects (null model) with a baseline mixed-effects model that only included crossed random effects for items and participants. The latter model had a significantly better fit to the data (AIC = 2991; $\chi^2(2) = 2750$, $p < .001$), which warranted the inclusion of the random effects structures. Next, the covariates and the fixed effects and their interactions were entered incrementally to the baseline mixed-effects model. Likelihood Ratio Tests (using a chi-square statistics) were conducted to evaluate whether the inclusion of a fixed effect significantly improved the model's fit statistics (Meteyard & Davies, 2020). Only the fixed effects that significantly improved the model fit were retained in the model. Significant interactions were followed with pairwise comparisons using Bonferroni correction. These were obtained by the *emmeans* package (Lenth, 2020).

To address the second research question, TD and DLD groups were compared with regard to the types and frequency of errors they made when they did not succeed in producing the target grammatical structure. For each error type, the differences in error rates between TD and DLD groups were examined using Mann-Whitney test.

To address the third research question, we assessed the diagnostic accuracy of the LITMUS-SR-PA-72 task. Receiver Operating Characteristic (*ROC*) curve was generated using the *pROC* package (Robin et al., 2011). *ROC* curves plot true positive rate (sensitivity) as a function of false-positive rate (1 – specificity) for all possible cutoff points (Gonçalves et al., 2014) and the optimal cutoff score with the best sensitivity and specificity tradeoff is determined. The area under the *ROC* curve (*AUC*) was computed and it is a measure of test classification accuracy

(Xu, 2012). Carter et al. (2016) indicate that AUC values could range from .5 to 1.0. An *AUC* of 1.0 reflects a perfect test, .90– .99 refers to an excellent test, .8–.89 indicates a good test, .7–.79 is a fair test and any values lower than this indicate that the test is uninformative.

4.3 Results

4.3.1 Analysis 1: Performance accuracy

Figure 4.1 illustrates the average percentage scores of children with and without DLD on LITMUS-SR-PA-72. The DLD scored significantly lower than the TD group using binary (t(34.51) = -12.17, p < .001, d = 3.02), error (t(31.71) = -11.03, p < .001, d = 2.79), and structural scoring methods (t(31.03) = - 10.08 , p < .001, d = 2.56).

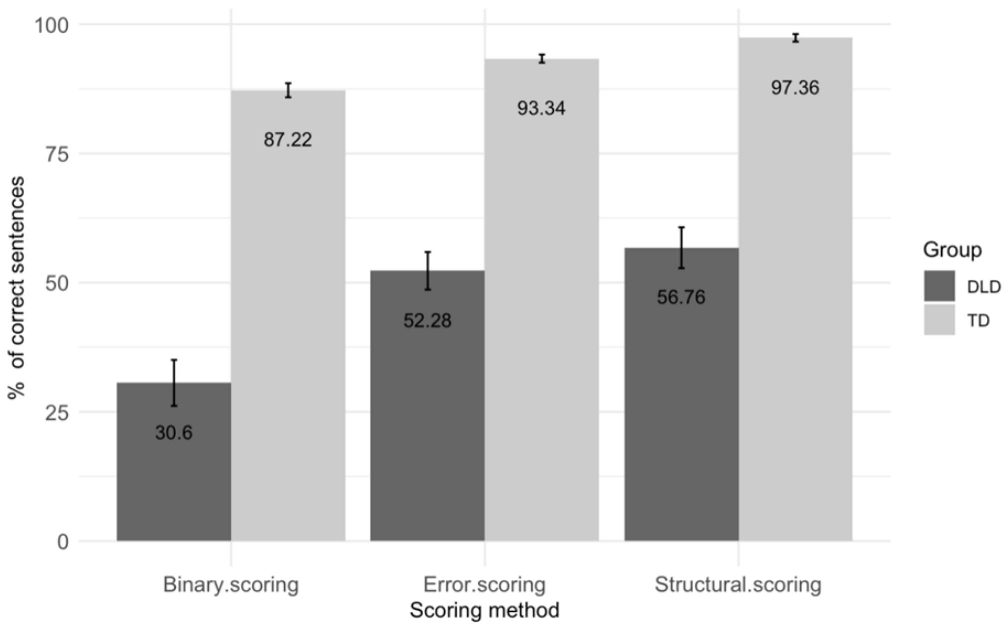

Figure 4.1. *Percentage scores of TD and DLD groups across binary, error and structural scoring methods on the LITMUS-SR-PA-72 task.*

Unlike binary and error scoring methods, structural scoring did not penalize the child for repetition errors that did not alter the grammatical structure assessed by the sentence. The structural scores index the child's ability to repeat the target grammatical structures, irrespective of their ability to exactly imitate all the words in the sentence. Given our focus is the children's grammatical ability, structural scores were used in the first and second analyses to investigate differences between the TD and DLD groups in repeating sentences of increasing grammatical complexity. A summary of structural scores for TD and DLD on the LITMUS-SR-PA-72 is displayed in Table 4.3.

Table 4.3. Percentage *structural scores of the TD and DLD groups across the grammatical targets of the LITMUS-SR-PA-72*

	TD	*DLD*	
	M(SD)	***M(SD)***	
Overall performance	97.36(5.71)	56.76(21.69)	***
Level 1	99.44(1.62)	83.33(13.08)	***
Past tense	99.17(3.66)	86.67(17.18)	***
Present tense	98.61(5.57)	57.78(33.82)	***
Noun plural	100(0)	92.22(12.17)	***
Possessive pronoun	100(0)	96.67(8.07)	***
Level 2	96.88(8.84)	47.22(25.13)	***
CLD	97.5(10)	67.5(25.55)	***
Sentences with complements	98.75(5.49)	56.67 (34.7)	***
Sentences with coordination	97.08(8.09)	41.67(36.16)	***
Wh- Object question	95.83(14.68)	42.92(36.95)	***
passive	96.25(12.02)	31.67(31.44)	***
Level 3	95.76(8.10)	39.72(24.65)	***
Conditional sentences	90.42(18.46)	23.33(32.78)	***
Object relatives	95(11.32)	33.75(39.41)	***
Subject relatives	98.33(6.29)	50.83(44.28)	***
Sentences with subordination	97.5(9.98)	45.83(37.76)	***

Note. TD = Typically Developing. DLD = Developmental Language Disorder.

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

The fit of the final model (*M*11) was significantly better than the intercept-only baseline model (*AIC* =2757, $\chi^2(17) = 268$, $p < .001$). The results of the final model are presented in Table 4.4. The inclusion of age ($X^2(1) = 6.89$, $p < .01$) and sentence length ($X^2(1) = 36.7$, $p < .001$) significantly improved the model fit. As shown in Table 4.4, age did not significantly predict SR performance (β = .03, *p* = .171) but sentence length did (β = - .20, *p* < .01). As the length of the target sentence increased (number of syllables), children were less likely to repeat it correctly. There was a main effect of group ($X^2(1) = 104$, $p < .001$) such that the TD group (*EMM* = 5.50, *SE* = .25) repeated sentences more accurately than the DLD group (*EMM* = .70, *SE* = .27, *p* < .001).

Table 4.4. *Parameter estimates of the final logistic mixed-effects model (M11)*

Parameters	β	SE (β)	Z statistic
Fixed Effects			
Intercept	.35	1.39	.25
Age	.03	.02	1.37
Sentence length	-.20	.07	-2.67**
Group: TD (compare with DLD)	4.84	.36	13.60**
Level 2 (compared with Level 1)	-.91	.90	-1.5*
Sentences with complements (compared with CLD)	.22	.51	.42
Conditionals (compared with CLD)	-1.24	.45	-2.73**
Sentences with coordinates (compared with CLD)	-.16	.63	-.25
Noun plurals (compared with CLD)	2.83	.44	6.38*
Object relatives (compared with CLD)	-.95	.38	-2.49**
Object Wh questions (compared with CLD)	-1.07	.40	-2.65**
Passives (compared with CLD)	-1.70	.45	-3.82***
Past tense (compared with CLD)	1.95	.40	4.92
Possessive pronouns (compared with CLD)	3.80	.55	6.87***
subject relatives (compared with CLD)	- .001	.39	-.89
Age X level	.02	.01	1.81***
Random Effects	**Variance**	**SD**	
Participant (Intercept)	1.92	1.39	
Item (Intercept)	.22	.47	
Observations 6480, participants: 90, items: 72			

Note. **TD =** Typically Developing. **DLD** = Developmental Language Disorder. **CLD =** sentences with clitic left dislocation
* = *p* < .05, ** = *p* < .01, *** = *p* <. 001

The group by age interaction ($X^2(1) = 2.15$, $p = .14$) and the group by sentence length interaction ($X^2(2) = 2.86$, $p = .24$) were non-significant. There was a main effect of level ($X^2(1) = 21.6$, $p < .001$). The level by group interaction was not significant ($X^2(1) = .58$, $p = .44$). The interaction between age and level was significant ($X^2(1) = 10.5$, $p < .01$). As Figure 4.2 shows, repetition accuracy of structures in all levels increased with age, but this effect was more prominent in levels 2 and 3 compared to level 1. When age was controlled, the proportion of correctly repeated structures in level 1 (*EMM* = 4.62, *SE* = .3) was higher than that of structures in level 2 (*EMM* =2.56, *SE* =.23, $p < .001$) and level 3 (*EMM* = 2.23, *SE* = .26, $p < .001$). There was no significant difference in the proportion of correctly repeated sentences between levels 2 and 3 ($p = .325$).

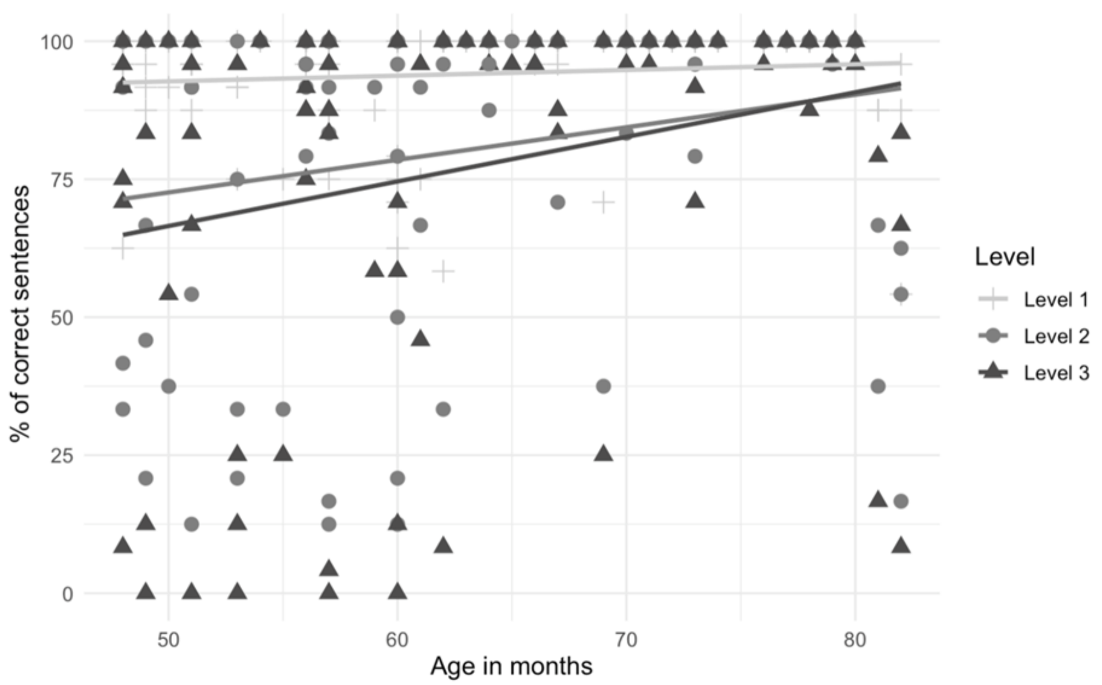

Figure 4.2. *A scatterplot showing the change in average structural scores of all children with age across levels 1, 2 and 3 of the LITMUS-SR-PA-72 task.*

Grammatical target had a significant effect on repetition accuracy ($X^2(11) = 88.3$, $p < .001$). Within level 1, the repetition accuracy of present tense verbs (*EMM* = 2.98, *SE* =.38) was significantly lower than that of possessive pronouns (*EMM* = 6.28, *SE* =.54), noun plurals (*EMM* =5.32, *SE* = .43) and past tense (*EMM* = 4.44, *SE*= .39; all comparisons; p <.001).

Within level 2, the probability of correct repetition did not differ significantly across sentences with Clitic Left Dislocation (*EMM* =3.10, *SE* = 2.35), complements (*EMM* =3.32 *SE* = .38), coordination (*EMM* = 2.94 =, SE= .48) and object-wh questions (*EMM* = 2.03, *SE* =.28; for all comparisons, $p > .05$). The repetition accuracy of passive sentences (*EMM* = 1.40, *SE* = .35) did not differ from that of object-wh questions (p =. 907) or sentences with coordination (p =.269), but was significantly lower than for sentences with complements or coordination (both comparisons, $p < .05$).

Within level 3, there were no significant differences between the repetition accuracy of sentences across sentences with conditionals (*EMM* = 1.45, *SE* = .45), subject relatives (*EMM* =2.78, *SE* = .28), object relatives (*EMM*= 1.83, *SE* = .29) and sentences with subordination (*EMM* = 2.78, *SE* = .38; for all comparisons $p > .05$). The group by grammatical target interaction was not significant ($X^2(12) = 12.1$, $p = .523$). The proportion of correct sentences repeated by the TD group was significantly higher than that of the DLD group across all of the target structures (for all comparisons, $p < .001$).We conducted an additional analysis to tease apart the effects of length (indexed by the number of syllables in the sentence) and grammatical complexity (indexed by level of complexity) on repetition accuracy. Levels 2 and 3 only differed in grammatical complexity but did not differ significantly in length. Therefore, we conducted a repeated measures ANOVA with repetition accuracy as the dependent variable, group as a between-subject variable and level (i.e., levels 2 and 3) as a within-subject variable. The main effects of group ($F(1,88) = 179$, $p < .001$), level ($F(1,88) = 6. 45$, $p < .05$) and their interaction ($F(1,88) = 5.67$, $p < .05$) were all significant. To unpack the interaction, post-hoc tests were conducted using Bonferroni corrected p-values to account for multiple comparisons (Field, 2009, p. 373). The TD group achieved significantly higher scores than the DLD group in repeating sentences within Level 2 ($p < .001$) and Level 3 ($p < .001$). Within the TD group, there was no significant difference in the average repetition scores of sentences in Level 2 and Level 3 (p = 1). In contrast, the DLD group scored significantly lower on sentences in Level 2 compared to sentences in level 1 ($p < .001$).

4.3.2 Analysis 2: Error patterns

Children with DLD were significantly more likely to produce ungrammatical structures relative to the TD group ($X^2(1, N = 90) = 1748, p < .001$). As illustrated in Table 4.5, the most common error in the repetition of past tense verbs included the omission of the entire verb from the sentence or substitution of the plural verb with a singular verb (e.g., *ʃirbib* drink-PAST-3MS "he drank" for *ʃirb**u*** drink-PAST-3PL "they drank"). These errors also affected present tense verbs. Additionally, when repeating present tense verbs, the DLD group showed an omission of the present progressive *b-* and/or gender/person agreement prefix *yi/ ti* of the present tense verb resulting in an imperative or imperfective verb (e.g., ***ti**qra*, read-IMPER-3FS "she study" or ***i**qra*, read-IMP-3MS "you study" for ***bti**qra*, read-PRES-3FS "she is studying"). In the DLD group, the imperative was used more frequently than the imperfective as a substitute of a present tense verb.

When repeating sentences with noun plurals, the DLD group substituted plural nouns with singular or dual nouns (e.g., *ta:be* "ball" or ta*:bt**e:n*** "two balls" for *tab**a:t*** "balls"). As for possessive pronouns, the DLD group showed omissions of the bound pronouns (e.g., *ʃaʕrat "hair" for ʃaʕrat**hum*** hair-CL-3FS "their hair"). Overall, the TD group made very few errors in level 1 structures (all comparisons, $p < .05$).

The DLD group was much more likely than the TD group ($p < .001$) to omit the passive prefix *in-* which resulted in changing the passive sentence to an active one (see Example 3b).

3. (a) **Item 34**: *il- ʃubak infataħ min il-hawa* [passive]

   The-window open-PASSIVE-3MS by the-wind

   "The window got opened by the wind"

   (b) **DLD-3**: *ʃubak fataħ hawa* [SVO]

   Window open-PAST-3MS wind

   "The window opened wind"

Table 4.5. Error types and Frequency of TD and DLD groups on LITMUS-SR-PA-72

		Group			
		TD	DLD		
Target structure	Error pattern (*actual production*)	*N*	*N*	*U*	z
Level 1					
Past tense	Omission of verb	0	9	1080	-3.37***
	Omission of plural suffix -u (*Singular for plural verb*)	0	6	1080	-3.37***
Present tense	Omission of prefix bti-/byi- (*imperative*)	1	36	1370	-5.77***
	Omission of b- clitic (*imperfective*)	2	8	1081	-2.75**
	Omission of verb	0	6	1020	-2.64***
	Omission of plural -u (*Singular for plural verb*)	0	5	1110	-3.69***
Noun plural	Omission of plural suffix (*Singular for plural noun*)	0	4	990	-2.20*
	Substitution of plural suffix (*Use of dual for plural noun*)	0	3	990	-2.20*
Possessive pronouns	Omission of possessive pronoun	0	3	1020	-2.64**
Level 2					
Passive	Omission of passive verb prefix in- (*Past tense verb*)	9	73	1587	-6.92***
Sentences with CLD	Omission of clitic pronoun and change to word order (*main clause*)	3	18	1240	-4.62***
	Omission of clitic pronoun and change to word order (SVO)	3	18	1240	-4.62***
	Omission of clitic pronoun (*SVO-wrong meaning*)	2	13	1201	-4.06***
Object Wh	Omission of object clitic (*subject Wh*)	10	34	1240	-3.84***
	Omission of Wh + object clitic (*main clause*)	7	30	885	-.01***
	Omission of Wh word (*CLD*)	2	24	1214	-4.4***
	Omission (fragment *structure*)	0	22	1170	-4.28***
Coordinate	Omission of coordinator (w) (*Two main clauses*)	4	32	1454	-6.01***

	Omission of coordinator (w) + one clause (*One main clause*)	1	19	1217	-4.44***
Complement	Omission resulting in one main clause	3	42	1474.5	-6.31***
	Omission (fragment *structure*)	0	4	1020	-2.64**
Level 3					
Subject relative	Omission of demonstrator *had* and relative *illi* (*main clause*)	3	86	1398	-5.99***
	Omission (fragment *structure*)	1	9	1095	-1.19***
Object relative	Omission of demonstrator had, relative *illi* and resumptive clitic(*main clause*)	6	67	1441.5	-6.14***
	Omission of demonstrator had and relative *illi* (*CLD*)	1	34	1401.5	-6.04***
	Omission of resumptive clitic (*subject relative*)	17	26	1073	-1.55**
	Omission (fragment *structure*)	0	12	1110	-3.7***
Conditional	Omission of conditional *iza* (*Two main clauses*)	13	43	1375	-4.72***
	Omission of conditional *iza* and one clause (*main clause*)	1	20	1337.5	-5.5***
	Omission (fragment *structure*)	0	11	1110	-3.67***
Subordinate	Omission of subordinate *ʕashan* (*Two main clauses*)	4	29	1281.5	-4.64***
	Omission of subordinate ʕashan + one clause (*main clause*)	2	12	1181	-4.07***
	Omission (fragment *structure*)	0	12	1080	-3.37***

Note. **DLD: Developmental** Language Disorder, **TD**: Typically Developing, **CLD**: Clitic Left Dislocation, * = $p < .05$, ** = $p < .01$, *** = $p < .001$

As for sentences with CLD, the TD and DLD groups omitted the clitic pronoun resulting in a sentence with canonical word order (see Example 4b). The frequency of this error was significantly higher in the DLD group compared to the TD group ($p < .001$)

4. (a) **Item 34**: *il-hadiya fatħatha il-binit* [sentence with CLD]
   The-gift open-PAST-3FS-CL-3FS the-girl
   "it is the gift the girl opened"

   (b) **DLD-10**: *fatħat il-hadiya* [SVO]
   Open-PAST-3FS the-gift
   "[She] opened it"

When repeating object Wh-questions, the TD and DLD groups demonstrated omissions of different elements which resulted in repeating object Wh-questions as subject Wh-question (5b), a sentence with CLD (5c), or a sentence with canonical word order (5d). An atypical pattern that only appeared in the DLD group was omitting several elements of the questions resulting in a fragmented structure (5e). These errors were significantly more frequent in the DLD group (for all comparisons, $p < .001$).

5. (a) **Item 28:** *ani bisse ħamlatha il-binit?* [object Wh- question]

   Which cat carry-PAST-3FS-CL-3FS the-girl?

   "Which cat did the girl carry?"

   (b) **DLD-2:** *ani: ħamlat il-bisse* …? [subject Wh-question]

   Which carry-PAST-3FS the-cat…?

   "Which carried the cat?"

   (c) **DLD-13:** …*binit ħamlatha bisse* [sentence with CLD]

   … girl carry-PAST-3FS-CL-3FS cat

   "it is the girl the cat carried"

   (d) **DLD-20:** …*ħamlat bisse…* [SVO]

   …carry-PAST-3FS cat…

   "[She] carried a cat"

   (e) **DLD-3:** *…ħamlat…* [Fragment]

   …carry-PAST-3FS …

   "[She] carried"

The DLD group showed atypical errors by which they either omitted the coordinator *w* or additional parts of sentences with coordination, resulting in two (6b) or one clause. Both groups showed omissions of several parts of the complement sentences which resulted in one clause. These errors rarely occurred in the TD group (all comparisons, $p < .001$). A further error that was unique to the DLD group only was the omission of several parts of the sentence resulted in a fragmented structure (6d).

6. (a) **Item 40:** *te:ta ʕimlat ʃa:j w baba akal basko:t* [sentence with coordination]

   Grandma make-PAST-3MS tea and dad eat-PAST-3MS biscuits

   "Grandma made tea and dad ate biscuits

(b) **DLD-1:** *mama ʕimlat ʃa:j baba akal ...* [two main clauses]

Mom make-PAST-3MS tea and dad eat-PAST-3MS …

"Mom made tea, dad ate"

(c) **DLD-24:** … *akal basko:t* baba [SVO]

… eat-PAST-3MS biscuits dad

"Dad ate biscuits"

(d) **DLD-26:** *tei:ta .. ʃa:j .. baba ..* [fragment]

Grandma .. tea .. dad ..

As for the repetition of subject and object relatives, the TD and DLD groups omitted the demonstrative *had* and relative noun *illi* (and the resumptive clitic pronoun of object relatives), resulting in a clause with canonical word order (7b). Atypical errors of the DLD group included: omission of demonstrative *had* and relative noun *illi* of object relatives, resulting in sentences with CLD (7c). All of these errors occurred at a significantly higher frequency in the DLD group relative to the TD group (for all comparisons, $p < .001$).

7. (a) **Item 65:** *ha:d il-ʕasi:r illi ʃirbo il-walad* [object relative]

This the-juice that drink-PAST-3MS-RES-3MS the-boy

"This is the juice that the boy drank"

(b) **DLD-3***: …il-walad… ʃirb il-ʕasi:r* [SVO]

…the-boy… drink-PAST-3MS the-juice

"The boy drank the juice"

(c) **DLD-12:** *…ʕasi:r ʃirbu walad* [sentence with CLD]

… juice drink-PAST-3MS-CL-3MS boy

"it is the juice the boy drank"

(d) **DLD-5:** *ha:d ʕasi:r ʃirb il-walad* [subject relative]

This the-juice drink-PAST-3MS the-boy

"This juice drank a the boy"

(e) **DLD-7**: .. *ʕasi:r .. walad ..* [fragment]

.. juice .. boy..

As for conditional sentences, the TD and DLD groups omitted the conditional *iza* which resulted in two main clauses (8b). In some cases, an additional omission of a clause resulted in only one main clause to be produced (8c). Both error types occurred more often in the DLD than the TD group ($p < .001$).

8. (a) **Item 62:** *iza il-walad byiʕmal il-wadʒib, raħ yruħ ʕala il- ħadi:qa* [conditional]

If the-boy do-PRES-3MS the-homework, will go-IMPER-3MS to the-park

"If the boy does the homework, he will go to the park"

(b) **DLD-7:** …*walad yiʕmal wadʒib,… yruħ ʕal ħadi:qa* [two main clauses]

…boy do-IMPER-3MS homework, …go-IMPER-3MS to the-park

"Boy do homework, go to park"

(c) **DLD-27:** …*ʕimil wadʒib,… ħadi:qa* [SVO]

...do-PAST-3MS homework …park

"[He] did homework, park"

(d) **DLD-12**: … ʕaħadi:qa [fragment]

…. Park

The most common error type in repeating sentences with subordination was the omission of the subordinator *ʕashan,* which resulted in two main clauses (10b). Sometimes this error was associated with an additional omission of either the main or subordinate clause, which resulted in only one clause (10c). Both errors occurred more often in the DLD group relative to the TD group (both comparisons, $p < .001$). The omission of several elements of sentences in level 3 resulted in fragmented sentences (see 8e, 9e, 10e) This was an atypical error specific to the DLD group (see Table 4.5).

9. (a) **Item 58***: il-walad ʕayyat ʕashan dayyaʕ il-luʕbeh* [subordinate]

The-boy cry-PAST-3MS because lose-PAST-3MS the-toy

"The boy cried because [he] lost the toy"

(b) **DLD-22:** *walad ʕayyat… dayyaʕ luʕbeh* [two main clauses]

Boy cry-PAST-3MS… lose-PAST-3MS toy

"Boy cried…[he] lost the toy"

(c) **DLD-14** *il-walad… dayyaʕ luʕbeh* [SVO]

The-boy …lose-PAST-3MS toy

"The boy lost a toy"

(d) **DLD-27**: *..luʕbeh ..walad..* [fragment]

.. toy..boy ..

4.3.3 Analysis 3: Diagnostic Accuracy

Sensitivity, specificity, and likelihood ratios were calculated for the final cutoff scores across the scoring methods are summarized in Table 4.6. For the binary scoring method, a cut-off score of 70.14 % or below correctly classified 93% of children with DLD (sensitivity) and 93% of TD children (specificity). A child with DLD was 14 times more likely to obtain a "fail" score (i.e., at or below the 70.14% cut-off) on the LITMUS-SR-PA-72 than a TD child, and only .07 times more likely to obtain a "pass" score (i.e., above the 70.14% cut-off) than a TD child.

With a cut-off score of 79.4 %, the error scoring method achieved a good sensitivity level of 93% and a good specificity level of 98%. A child with DLD was 54 times more likely to receive a "fail" score on the task than a TD child, and only .07 times more likely to obtain a "pass" score than a TD child. Similarly, the structural scoring method achieved a high level of diagnostic accuracy. A cut-off score at 90.97% correctly classified 97% of children with DLD (sensitivity) and 92% of TD children (specificity). A child with DLD was 11 times more likely to obtain a "fail" score (i.e., below 90.97% cut-off) on the LITMUS-SR-PA-72 compared to a TD child, and was only .07 more likely than a TD child to score above cut-off score. The diagnostic accuracy of the LITMUS-SR-PA-72 achieved using the binary scores did not differ from that achieved using the error (p = .09) or structural scores (p = .986). Similarly, there was no significant difference in the diagnostic accuracy of error scores compared to the structural scores (p = .986).

Table 4.6. *Diagnostic Accuracy metrics of the LITMUS-PA-SR-72*

Scoring method	Cut off % (raw)	Sensitivity (Correct DLD) [95% CI]	Specificity (Correct TD) [95% CI]	LR+ [95% CI]	LR- [95% CI]	AUC [95% CI]
Binary scoring	70.14 % (52/72)	**.93** (28/30) [.83 - 1]	**.93** (56/60) [.87 - .98]	**13.93** [5 .41- 36.26]	**.07** [5.41 - .27]	**.97** [.94 - 1]
Error scoring	79.4% (160/216)	**.93** (28/30) [.83 - 1]	**.98** (59/60) [.95 - 1]	**54.88** [8 - 329]	**.07** [02 - .26]	**.98** [.96 – 1]
Structural scoring	90.97% (65/72)	**.97** (29/30) [.83 – 1]	**.92** (55/60) [.83 – 1]	**11.27** [84.83 - 26.12]	**.07** [.02 - .27]	**.99** [.97 – 1]

Note. LR+: Positive Likelihood Ratio**. LR-:** Negative Likelihood Ratio. AUC: Area Under Curve

4.4 Discussion

The TD and DLD groups differed significantly in their scores on the LITMUS-SR-PA-72, showing that SR is a locus of difficulty for Arabic-speaking children with DLD. The pattern of grammatical errors in the TD and DLD groups were largely similar, with a higher frequency of grammatical errors in the DLD than in the TD group. The LITMUS-SR-PA-72 discriminated accurately between Arabic-speaking children with DLD and their age-matched TD peers.

4.4.1 Arabic-speaking children with DLD performed poorly on the LITMUS-SR-PA-72

Our first research question addressed how children with and without DLD differ on their performance accuracy on the LITMUS-SR-PA-72. As predicted, we found large and significant differences in the average performance of the TD and DLD groups using binary, error and grammatical structural scoring methods. These findings are in line with previous studies suggesting that SR is an area of weakness for children with DLD acquiring a variety of languages (e.g., Conti-Ramsden, 2003; Pham & Ebert, 2020; Thordardottir et al., 2011; Vang Christensen, 2019), including Arabic-speaking children with or at risk of DLD (Shaalan, 2010; Wallan, 2018). The average grammatical structural scores on the LITMUS-SR-PA-72 task improved with age, suggesting that the task captured grammatical developmental changes within the age span of 4 to 6 years. Although the repetition accuracy of the grammatical structures in level 1 remained stable against age, it increased significantly with age for structures within levels 2 and 3 for both groups. The grammatical structures tested within level 1 are acquired by 3 years of age in Arabic, except for Masculine Sound Plurals and Broken Plurals which are acquired gradually into school-age years (Abdalla et al., 2013; Aljenaie, 2001; Faquih, 2014; Omar, 1973; Ravid & Farah, 1999). This explains the limited variation between older and younger children in repetition accuracy of level 1 structures. No data is available in Arabic on the acquisition of structures within levels 2 (passive, clitic left dislocation, object Wh-questions, coordinates, complements) and 3 (subject relative, object relative, subordinates and conditionals). Evidence from other languages suggests that the acquisition of these structures extends into school-age (e.g., Friedmann & Novogrodsky, 2004; Leonard, 1989; Mastropavlou & Tsimpli, 2011; Stromswold, 1995). They could be emerging and not yet fully acquired by the children in our sample which may contribute to the observed age effect.

The SR accuracy decreased as sentence length increased. Sentences within levels 2 and 3 did not differ in length but were significantly longer than sentences within level 1. For both groups, the average SR accuracy scores for levels 2 and 3 were significantly lower than those for level 1, but no difference was observed in the average SR accuracy between levels 2 and 3. The reduction of SR accuracy with increasing length could point out the role of short-term memory in SR, with longer sentences placing greater demands on memory capacity than shorter sentences (e.g., Alloway & Gathercole, 2005). However, the observed decline in SR accuracy across levels occurred despite controlling for sentence length in the analysis. Furthermore, sentences within levels 2 and 3 were not only longer but also syntactically more complex than sentences in level 1. Hence, the increased difficulty with the repetition of sentences within level 2 and 3 relative to level 1 cannot be attributed solely to differences in short-term memory load but could also reflect differences in underlying syntactic representations in long-term memory (Frizelle et al., 2017).

In an attempt to disentangle the influence of length and grammatical complexity on repetition accuracy, we conducted an additional analysis in which we compared the performance of both groups on levels 2 and 3 which did not differ significantly in length, but rather in grammatical complexity. Before discussing the results, we would like to acknowledge that while the results of this analysis (repeated measures ANOVA) were largely similar to the original mixed-effects model we conducted, there are slight differences. The difference is likely to be due to the increased complexity of mixed-effects model relative to the repeated-measures model. Specifically, the inclusion of random effects structures may reduce the amount of variance that is attributed to fixed effects and their interactions. That is, if fixed effects or their interactions are small or weak, they could appear as being non-significant in mixed-models. This could explain the lack of group by level interaction in the mixed-model but the significance of this interaction in this follow-up analysis. Within the TD group, there was no significant difference in performance between the levels 2 and 3. In contrast, the DLD group showed a significantly lower repetition accuracy of sentences in level 3 compared to level 2. This finding suggests that syntactic complexity influenced the repetition accuracy in the DLD group but not the TD group. That is, the DLD group appeared to be more sensitive and found it more challenging to repeat syntactically complex sentences. The lack of this effect in the TD group could be attributed to the fact that their performance approached ceiling across all levels. This result extends previous evidence showing that when sentence length was constant, increasing syntactic complexity resulted in a greater number of errors in SR (Frizelle & Fletcher, 2014; Kidd et al., 2007; Riches et al., 2010). This conclusion is in line with accumulating evidence maintaining that SR is not a pure measure of memory, but rather requires interaction between linguistic representations and memory resources (e.g., Marinis & Armon-Lotem, 2015; Moll et al., 2015).

We further examined the accuracy of repetition of the target grammatical structures assessed by the task. The DLD group had lower accuracy scores compared to the TD group in producing past tense and present tense verbs. This finding confirms that the production of verb tense and subject-verb agreement morphology is a weakness for Arabic-speaking children with DLD (Abdallah & Crago, 2008; Fahim, 2017; Taha et al., 2021b). The children with DLD in our sample repeated noun plurals and possessive pronouns with high accuracy (> 90%), suggesting that these structures were not problematic for them. This contrasts with findings from previous studies which have used elicitation tasks (Abdallah et al., 2013; Faquih, 2014). Importantly, this finding is inconsistent with the results of the DLD group on the APNT in which they had an average score of 22%. As mentioned in the Methods section, the items used in the LITMUS-SR-PA-72 were limited to early acquired words. Hence these items had high frequency and were familiar to the children. These findings could be explained by referring to the Critical Mass Hypothesis, which assumes a relationship between lexical development and morpho-syntactic skills in children (Marchman & Bates, 1994; Windfuhr et al., 2002). It proposes that, once the children have acquired a critical mass of words (i.e., nouns), acquiring morphological properties (e.g., noun plurals and possessive pronouns) would be facilitated. As the children in our study had acquired all the nouns used in the LITMUS-SR-PA-72 task, they did not have much difficulty with the morphological properties of these nouns (i.e., forming plurals or possessive pronouns) as they would have acquired these nouns.

The DLD group had significantly lower scores than the TD group in repeating all language-independent structures within level 2 (sentences with Clitic Left Dislocations, passives and object Wh-questions) and level 3 (subject and object relatives, conditionals and sentences with subordination). This finding is not surprising as the production and comprehension of syntactic constructs that involve movement (e.g., sentences with Clitic Left Dislocation, passives, object relatives, object Wh-questions) have been identified to be cross-linguistically impaired in children with DLD (e.g., Arosio et al., 2009; Bedore & Leonard, 2001; Deevy & Leonard, 2004; Friedmann & Novogrodsky, 2011; Marinis & Saddy, 2013; Novogrodsky & Friedmann, 2006;

Prévost et al., 2014; Shaalan, 2010; Tuller et al., 2011). These non-canonical structures are derived via syntactic movement which involves building long-distance syntactic dependencies. Surface SVO word order corresponds to a canonical order of arguments (agent-action-theme) whereas non-canonical sentences do not. The DLD group's low scores when repeating movement-derived sentences could be attributed to a difficulty in the assignment of a thematic role to the moved element (Friedmann & Novogrodsky, 2011). The DLD group’s poor repetition ability of sentences with clausal embedding (e.g., sentences with complements, subject and object relatives, sentences with conditionals) has also been identified as an area of difficulty for children with DLD across languages (e.g., Arosio et al., 2009; Fleckstein et al., 2018; Frizelle & Fletcher, 2014; Gavarró, 2017; Owen & Leonard, 2006).

Our study provides an initial picture of the morpho-syntactic difficulties of Arabic-speaking children with DLD. We found significant group differences in the production of verbs, sentences with passives, clitic left dislocation, object Wh-questions, subject and object relative clauses, sentences with coordination, complements, subordination and conditionals. These structures appear to be sensitive to the language differences between children with and without DLD and could potentially support the identification of DLD in Arabic. A further investigation of these structures using other probes (e.g., elicitation tasks, language samples) is warranted to better establish their potential as clinical markers of DLD in Arabic-speaking children.

4.4.2 Grammatical errors

Our second research question focused on how children with and without DLD differ in terms of their morpho-syntactic errors on the LITMUS-SR-PA-72 task. The quantity, and in some instances the type of grammatical errors, differed between the TD and DLD groups. The proportion of errors in the DLD group was significantly higher than that of the TD group.

With regard to the language-specific structures, the DLD group showed either omission or substitution errors when repeating present or past tense verbs. Tense errors consisted of replacing the target tense with a non-finite form (imperative or imperfective).

Given the fusional properties of Arabic verb morphology, errors in tense were sometimes associated with errors in agreement. Main agreement errors were the use of 2nd person plural verbs in place of 3rd person verbs (in cases where the imperative was used), or the use of singular verbs instead of plural verbs. These errors were barely produced by the TD group, suggesting that they are age-inappropriate errors, whereas they have been observed in Arabic-speaking children with DLD and toddler TD children (Abdallah & Crago, 2008; Ouali, 2018; Qasem & Sircar, 2017; Taha et al., 2021b).

As for the language-independent structures, omissions were the dominant error type observed in both groups, with a higher proportion of errors in the DLD compared to the TD group. The omission errors primarily affected grammatical suffixes such as passive prefix *–in* from the passive verb, which resulted in producing an active voice sentence. Object clitic pronouns in sentences with clitic left dislocation, Wh object questions and object relative clauses were omitted, which resulted in a change of the target grammatical structure. Furthermore, omission errors affected function words such as the coordinator *w* "and", conditional *iza* "if", subordinate *ʕashan* "because", demonstrative *ha:d* "this", relative pronoun *illi* "that" and Wh-words such as *mi:n* "who" and *ani/u* "which". The omission of the grammatical suffixes or function words frequently co-occurred or was associated with word order changes. These errors occurred in the TD group as well, so could be described as typical. In general, the omission error patterns in the DLD group have been observed in other languages, which extends the evidence that the use of grammatical affixes is an area of weakness in children with DLD (e.g., Bedore & Leonard, 1998; Frizelle & Fletcher, 2014; Grüter, 2005; Hansson & Nettelbladt, 2006a; Novogrodsky & Friedmann, 2006; Seeff-Gabriel et al., 2010). One atypical error type that occurred exclusively in the DLD group was the omission of several elements of the target sentences which resulted in a fragmented structure. This particularly applied to sentences involving syntactic movement: passive sentences, sentences with CLD, object Wh-questions and object relatives. The repetition of structures involving movement and/or embedding as fragmented structures could indicate poor morpho-syntactic representations of these structures in the long-term memory of children with DLD (Frizelle et al., 2017), or that these structures have not yet been acquired.

4.4.3 LITMUS-SR-PA-72 could be a clinical marker of DLD in Arabic

Our third research question addressed whether the LITMUS-SR-PA-72 task can reliably distinguish children with and without DLD. ROC analyses were performed to obtain the best cut-off points for the binary, error and structural scoring methods. The AUC levels associated with the optimal cut-off scores ranged from .97 to .99 for the three scoring methods, suggesting that the LITMUS-SR-PA-72 yielded an excellent diagnostic accuracy. This finding is consistent with previous studies showing a good diagnostic accuracy of SR tasks in identifying DLD in many languages (e.g., Armon-Lotem & Meir, 2016; Leclercq et al., 2014; Pham & Ebert, 2020; Vang Christensen, 2019).

Overall, the binary, error and structural scoring methods showed sensitivity (proportion of children with DLD correctly identified) and specificity values (proportion of TD children correctly identified) were larger than 90%. Hence, scoring the LITMUS-SR-PA-72 test with any of these scoring methods yielded a good power in differentiating between children with and without DLD (Plante & Vance, 1994). Across the scoring methods, the positive likelihood ratios (LR+) were higher than 10 and the negative likelihood ratios (LR-) was less than .1. These values suggest that a child with DLD was more than 10 times more likely to obtain a "fail" score (i.e., at or below the specified cut-off) on the LITMUS-SR-PA-72 than a TD child, and only less than .1 times more likely to obtain a "pass" score (i.e., above the specified cut-off) than a TD child. Together, these findings indicate that a score above or below the specified cut-off point on the LITMUS-SR-PA-72 could be interpreted with strong confidence as indicative of the presence or absence of DLD (Dollaghan, 2007).

Despite the good levels of sensitivity, specificity and strong likelihood ratios, these values should be interpreted in consideration of the associated 95% CIs. The lower bound of the 95% CIs for the sensitivity of the binary and error scoring methods was 83% and for the specificity of the binary and structural scoring methods were 87% and 83%, respectively. These values fall below the 90% threshold which characterizes tests with good diagnostic accuracy; rather, they are only indicative of adequate diagnostic accuracy (Plante & Vance, 1994). Similarly, the lower bound of 95% CIs for the LR+ of the binary and error methods were ≥ 3 whereas the upper bound of 95% CIs for the LR- of all the scoring methods was ≤ .3. These values do not meet the criteria of Dollaghan (2007) for a clinically informative test i.e., of LR+ ≥ 10.0 and LR- ≤ 0.1; rather, they indicate that the test is suggestive but insufficient to rule in or rule out the disorder. Therefore, we refrain from suggesting that the LITMU-SR-PA is a strong indicator of the presence or absence of DLD. Instead, we propose that the LITMU-SR-PA test is suggestive of DLD and should be used in combination with other assessment tools to achieve accurate DLD diagnosis.

4.4.4 Clinical Implications

The LITMUS-SR-PA-72 task, a theoretically-based measure, forms the first step towards a more research-informed approach to DLD diagnosis in the Palestinian-Arabic context. Our study suggests that poor SR may characterize DLD in Arabic-speaking children. This leads us to recommend SLTs consider assessing SR as part of the diagnostic procedures of DLD. Particularly, the LITMUS-SR-PA-72 task could be used to conduct a systematic evaluation of the morpho-syntactic structures known to be problematic for children with DLD. These grammatical structures may be avoided by children with DLD in traditional elicitation tasks (i.e., spontaneous language samples), providing fewer opportunities to assess these structures. Our study emphasizes the potential diagnostic value of the LITMUS-SR-PA-72 task as an index with good diagnostic accuracy in differentiating 4 to 6-year-old Palestinian Arabic-speaking children with DLD from TD children. This good discriminatory power of the task was consistent across the binary, error and structural scoring methods, hence either of these scoring systems could be applied according to the purpose of the assessment. The binary scoring method is a simplified scoring system that is quick and easy to administer; it could be most useful when the LITMUS-SR-PA-72 is used to determine whether or not a child's linguistic abilities require further assessment. Error scoring is a more fine-grained method and could be used to determine the severity of a child's language production deficits. Grammatical structure scoring is a precise scoring system that could be used to build a profile of the child's grammatical strengths and weaknesses. Morpho-syntactic structures that the child fails to repeat could then be further assessed. The structural scoring method could be used to measure the effectiveness of language intervention and progress in the mastery of the target grammatical structures. Unlike informal language tasks, the LITMUS-SR-PA-72 provides clinicians with language scores that could be compared to different cut-off points according to the scoring method that is being used.

4.4.5 Limitations and future directions

The two-gate design alongside the stringent criteria employed for the inclusionary language measures could have resulted in a spectrum bias (Dollaghan & Horner, 2011; Pawłowska, 2014; Redmond et al., 2019). The children in our study were recruited from preselected samples (e.g., children with a prior DLD diagnosis versus children with typical language development). A confirmation of the DLD/TD status involved scoring above (for TD children) or at/below cut-off

-1.5 SD (for children with DLD) on at least three language measures assessing morpho-syntax and phonology. This could have resulted in two groups on the polarized ends of the spectrum of language abilities. The comparison of the DLD group with severe language deficits and the TD group with average language abilities could have led to an overestimation of the diagnostic accuracy levels (Pawłowska, 2014). Furthermore, our DLD group may not be representative of Palestinian children with DLD. Given that DLD diagnosis in Palestine is based on informal language assessments, children whose language difficulties are borderline and/or do not present with comorbid speech sound disorder are more likely to be undiagnosed and, consequently, not entitled to receive language intervention services. Children who receive a clinical diagnosis of DLD usually have more severe language deficits. To address these limitations, Pawłowska (2014) recommends employing one-gate designs in which all children are recruited from a single population (unselected sample) so that heterogeneous and representative samples of children with and without DLD are recruited.

The current study examined the clinical usefulness of the LITMUS-SR-PA-72 task in identifying DLD in 4 to 6-year-old children hence limiting generalizability of the results to older or younger children. Future work could examine the diagnostic value of the task in identifying DLD in a wider age range and also establishing norms for the acquisition of grammatical structures in Arabic. This information is imperative for the development of age-appropriate grammatical assessments, to inform clinicians and educators of the grammatical structures to watch out for when assessing children across different age groups.

The LITMUS-SR-PA-72 was administered live. This could have resulted in variations in the pitch, speed and loudness of the examiner when reading sentences to different children. For a more consistent task delivery, a computerized version of the task using audio-recorded sentences could be developed. Finally, the vocabulary used in the task was not controlled for frequency or imageability. More research is needed to establish psychometric properties of Arabic vocabulary.

4.4.6 Conclusion

The present study found that sentence repetition deficits could be a potential clinical marker of DLD in Arabic-speaking children. Compared to age-matched controls, the DLD group scored significantly lower on repetition of both language-specific and language-independent syntactically complex structures. The frequency of morpho-syntactic errors was significantly higher in the DLD than in the TD group. Some errors occurred exclusively in the DLD group, suggesting clinicians should consider the type and frequency of error patterns when assessing children's expressive grammar. The LITMUS-SR-PA-72 is moderately accurate in differentiating between Palestinian Arabic-speaking children with and without DLD. The task is only suggestive of the presence or absence of DLD and should be used alongside information from other sources to improve the accuracy of DLD diagnosis. The clinical utility of a refined version of the task should be confirmed in a more representative sample of Palestinian children via larger-scale population studies.

Acknowledgments

We thank all of the children and parents who took part of this study, not to mention the speech and language therapists and teachers, especially Nairouz Nijm Aldeen, who facilitated the recruitment of children for this study. We would also like to thank Theo Marinis for his helpful advice and feedback on the coding of error data.

Chapter 5: General Discussion

The primary goal of this thesis was to examine potential clinical markers that could differentiate Arabic-speaking children with DLD from TD children. Three studies were conducted, each exploring the performance of Arabic-speaking children with and without DLD on a clinical marker task. Study 1 examined verb morphology use, a measure of linguistic knowledge. Study 2 and Study 3 examined nonword repetition and sentence repetition, respectively, both being measures of linguistic processing. The following section provides a summary of the primary findings.

5.1 Summary of findings

Study 1 revealed that, while Arabic-speaking children with DLD exhibit a general difficulty with verb morphology production, not all verb forms are problematic for this group. Out of the seven tense and agreement forms investigated, children with DLD used three of them with comparable accuracy to TD children (e.g., masculine, singular and plural verbs). Two other forms were used by children with DLD with more than 90% accuracy (e.g., past tense and 3rd person verbs). The two forms that posed challenges for children with DLD were the present tense and feminine verbs. Children with and without DLD made nonfinite (e.g., used the imperative or the imperfective) and finite tense errors (e.g., used incorrect tense). Agreement errors resembled the use of structurally simpler (e.g., less marked) agreement categories in place of the target marked categories: masculine verbs were used in place of feminine verbs, singular verbs were used in place of plural verbs, and 2nd person verbs were used in place of 3rd person verbs. Verb morphology was not significantly affected in Arabic-speaking children with DLD, just as it is the case in other morphologically rich languages such as Hebrew (Dromi et al., 1993, 1999; L. B. Leonard, Dromi, et al., 2000), Italian (Bortolini et al., 1997; Leonard, Caselli, Bortolini, Karla, et al., 1992) and Spanish (Bedore & Leonard, 2001, 2005; Grinstead et al., 2013). The results of Study 1 are in contrast to the findings from English, where verb morphology use has been found to be seriously affected in English-speaking children with DLD (e.g., Rice & Wexler, 1996).

Nonword repetition difficulties are well-attested in children with DLD across many languages (Ahufinger et al., 2021; Armon-Lotem & Meir, 2016; Graf Estes et al., 2007; Kalnak et al., 2014; Thordardottir et al., 2011). Study 2 extends this evidence by showing that nonword repetition is also impaired in Arabic-speaking children with DLD compared to age-matched TD peers. In the DLD and TD groups, nonword repetition accuracy decreased significantly for three-syllable nonwords relative to one and two-syllable nonwords. Wordlikeness influenced the nonword repetition accuracy in children with and without DLD. Both groups showed a higher repetition accuracy of high word-like nonwords than low word-like nonwords. The phonological complexity of the nonwords differentially affected the TD and DLD groups. While the TD group repeated nonwords with one, two or no consonant clusters with similar accuracy, the DLD group exhibited pronounced difficulties with repeating nonwords containing two consonant clusters compared to nonwords with one or no consonant clusters. Overall, these findings support the view that nonword repetition is a multi-dimensional task that taps into several processes, including verbal short term memory (as indicated by the length effect), discrimination, encoding, processing or speech motor production of phonologically complex forms (e.g., as suggested by the phonological complexity effect) as well as access to linguistic representation in long-term memory (as indicated by the wordlikeness effect; Leonard, 2014a, p.279). Study 2 further revealed that poor nonword repetition performance could correctly classify 93% of children with DLD and 93% of TD children, signifying its potential as a clinical marker of DLD in Arabic.

Study 3 concluded that sentence repetition is a locus of difficulty for Arabic-speaking children with DLD, just as it is for children with DLD acquiring other languages (e.g., Conti-Ramsden, 2003; Pham & Ebert, 2020; Thordardottir et al., 2011; Vang Christensen, 2019). In the TD and DLD groups, a decrease in repetition accuracy with increased sentence length was observed. This suggests that verbal short-term memory plays a role in sentence repetition. Findings also revealed that syntactic complexity influenced sentence repetition accuracy in the DLD group but not the TD group. Together, these outcomes support the view that adequate sentence repetition requires an interaction between linguistic representations and memory resources (e.g., Marinis & Armon-Lotem, 2015; Moll et al., 2015).

Study 3 showed that Arabic-speaking children with DLD have difficulties with repeating language-specific grammatical structures (e.g., present and past tense verbs) as well as language-independent syntactically complex structures (for a review, see Leonard, 2014a). These language-independent structures include sentences with clitic left dislocation, passives, object Wh-questions, subject and object relatives, conditionals and sentences with subordination. Moreover, children with DLD made a significantly higher number of grammatical errors relative to their TD peers. Both groups showed omissions of grammatical inflections and function words. These errors frequently co-occurred or were associated with word order changes. Some errors were only observed in the DLD group, including the omission or substitution of verbal tense and agreement inflections, the omission of verbs, and the omission of several constituents of the sentences resulting in fragmented structures. Lastly, study 3 found that sentence repetition correctly identified more than 90% of children with DLD and more than 90% of TD children across the different scoring methods. This indicates the potential of sentence repetition as a clinical marker of DLD in Arabic.

Returning to the main aim of this thesis, study 1 showed that the poor production of present tense and feminine verbs could be potential grammatical markers of DLD in Arabic. However, this finding is tentative as it is only based on group differences between children with and without DLD in using these forms. An analysis of diagnostic accuracy (which considers individual differences) is required to determine how accurate poor verb morphology is in discriminating children with and without DLD. Study 2 and Study 3 supported the diagnostic accuracy of nonword repetition and sentence repetition as possible clinical markers of DLD in Arabic. Therefore, the results of this thesis pertaining to Arabic strengthen the potential of sentence repetition and nonword repetition as cross-linguistic markers of DLD.

Due to space limitations of the journal where these studies are published, there was no opportunity within chapters 2, 3 and 4 to discuss the theoretical implications of the results. The following section fills this gap. In chapter 1, different theoretical accounts of DLD were reviewed (see section 1.5). These theories view DLD as either a linguistic deficit (knowledge-based accounts) or as a processing limitation (processing-based accounts). Findings from the three studies will be discussed in relation to these accounts in the next section.

5.2 Theoretical implications

5.2.1 Explaining the verb morphology deficits

The Extended Optional Infinitive (EOI), a knowledge-based account, proposes that, like younger TD children, children with DLD will go through a stage where they show optional use of grammatical morphemes that carry verb tense and agreement features (Rice & Wexler, 1996; Rice et al., 1995; Rice et al., 1998). Due to a maturational delay, this stage lasts longer for children with DLD. When children fail to mark verb tense or agreement, they are expected to use a nonfinite form, i.e., the bare stem or the infinitive (as reported in English; Rice & Wexler 1996). In line with the EOI, Study 1 revealed that Arabic-speaking children with DLD produced the target tense and agreement forms sometimes correctly and sometimes incorrectly, indicating that they optionally marked tense and agreement in obligatory contexts.

Unlike English, Arabic does not have infinitive forms (Abdallah, 2002; Benmamoun, 2000). The tense errors of the DLD group were characterized by the use of the imperative or the imperfective, which are tenseless/nonfinite forms (Aljenaie, 2001; Benmamoun, 2000). Consistent with the EOI, when Arabic-speaking children with DLD failed to mark tense, they used nonfinite forms that have been previously observed in the language of Arabic-speaking TD toddlers (Aljenaie, 2000, 2001, 2010; Qasem & Sircar, 2017). The verb deficits in Arabic-speaking children with DLD may reflect an extension of a typical acquisition stage. In Study 1, the significant group difference between children with DLD and same-age TD peers suggests that the verb morphology difficulties in the DLD group may reflect a delayed development pattern. However, this conclusion may be misleading and requires further investigation by comparing the performance of Arabic-speaking children with DLD to that of a younger language-matched TD group. Abdallah and Crago (2008) found that Hijazi Arabic-speaking children with DLD were significantly less accurate than younger MLU-matched TD children in marking verb tense and agreement, suggesting that the verb morphology difficulties in Arabic-speaking children with DLD may reflect a disordered, rather than a delayed pattern of development.

In Study 1, Arabic-speaking children made finite tense and agreement errors (e.g., they used past tense for present tense or masculine present tense for feminine present tense). These error types have been observed in children with DLD acquiring languages such as Spanish, Italian, Greek and Inuktitut (Bedore & Leonard, 2001; Bortolini et al., 1997; Clahsen & Dalalakis, 1999; Crago & Allen, 2001). The EOI does not predict the occurrence of finite tense and agreement errors. Therefore, Study 1 supports Paradis and Crago (2001)'s proposal to refer to the EOI stage as the "extended optional default". The term "default" captures the cross-linguistic variations of verb tense and agreement errors of children with DLD. Overall, Arabic-speaking children with DLD appear to go through a stage similar to the EOI. Yet, its linguistic characteristics are different from what has been reported in English.

The verb morphology difficulties in Arabic-speaking children with DLD may be partially explained by the **Grammatical Agreement Deficit** (GAD) model, another knowledge-based account. Recall that the GAD posits that the optional phi-features of subject-verb agreement that do not have semantic interpretation (e.g., person, number and gender) can be seriously affected by DLD (Clahsen et al., 1997). The results of Study 1 appear to be consistent with this prediction. Arabic-speaking children with DLD were significantly less accurate than TD peers in marking subject-verb agreement. According to the GAD, all agreement categories are expected to be problematic for children with DLD. In contrast to this prediction, Arabic-speaking children with DLD did not have difficulties marking subject-verb agreement for number (singular and plural verbs) or person (3rd person verbs). Although subject-verb agreement for gender was problematic for children with DLD, these children were more proficient in using masculine verbs relative to feminine verbs. Besides failing to account for the variations in the subject-verb agreement accuracy, GAD does not predict the tense difficulties observed in Arabic-speaking children with DLD in Study 1.

The accounts that attribute DLD to a deficit in linguistic knowledge, namely, the EOI and GAD, cannot fully account for the verb morphology deficits in Arabic. Processing-based accounts offer an alternative perspective. According to the **Morphological Richness Account (MRA)**, children with DLD will direct their limited processing resources to the most dominant grammatical cues of the linguistic input. In languages with sparse morphology (e.g., English), these cues are carried by word order. In contrast, in richly inflected languages (e.g., Arabic), these cues are in the grammatical inflections (Lukács et al., 2009). The MRA, therefore, predicts that children with DLD acquiring Arabic will fare better than their English-speaking counterparts in using grammatical inflections. The results of Study 1 are compatible with this notion as verb morphology was not greatly affected in Arabic-speaking children with DLD. To illustrate, the Arabic-speaking children with DLD in Study 1 used past tense with 92% of accuracy. In contrast, same-age English-speaking counterparts in Conti-Ramsden (2003)'s study produced this form with only 20.5% accuracy.

The MRA predicts that children with DLD will make grammatical errors. This is due to the limitations in their processing resources combined with the demands of processing inflections that denote several grammatical functions. Given the assumption that the main obstacle for these children is their processing limitations rather than their knowledge of the grammatical function of the inflections, the resulting grammatical forms would minimally differ from the target forms, i.e. they will be near misses. The results from Study 1 support this prediction. The majority of the incorrect verb productions of the DLD group in Study 1 differed from the target by one feature (apart from the imperative errors where it reflected tense and person errors). However, the results of Study 1 pose challenges to the MRA account. Subject-verb agreement in Arabic is fusional (Abdallah, 2002), where a single prefix or a circumfix denotes several agreement categories. On the other hand, tense is either marked by zero inflections (for past tense) or by attaching the present progressive clitic *b-* to the imperfective verb stem (for present tense). According to the MRA, subject-verb agreement inflections would be expected to be more problematic for children with DLD than tense inflections, given that the former simultaneously carry several grammatical functions. In contrast, Study 1 showed that Arabic-speaking children with DLD were more accurate in marking subject-verb agreement relative to tense. This may be because a few of the target agreement categories (e.g., singular number agreement and masculine gender agreement in past tense) are unmarked by overt inflections. The MRA can account for the cross-linguistic differences in the severity of verb morphology between Arabic and English-speaking children with DLD. However, it does not explain the within-language difference between tense and agreement marking accuracy.

Another processing-based theoretical explanation is the **Surface Account** (SA), according to which children with DLD have difficulties in perceiving morphemes of low acoustic salience. In line with the SA, Study 1 showed that children with DLD produced present tense verbs with the unstressed prefix (e.g., ***byi****ʹlawwin* "he is colouring") less accurately than verbs with the stressed prefix (e.g., *ʹ**byi**ftaħ* "he is opening"). Furthermore, the past tense feminine agreement morpheme *-at* as in *ʹkatb**at*** "she wrote" has low acoustic salience as it always occurs as a word-final unstressed syllable. This suffix was challenging for the DLD group in Study 1 and it was often omitted from the past tense feminine verbs resulting in a masculine verb *katab* "he wrote". However, two findings of Study 1 cannot be accounted for by the SA. First, even when the present tense prefix was stressed, children with DLD produced present tense verbs with significantly lower accuracy than TD peers, suggesting that these children had difficulties with using inflections that are were acoustically salient. Second, the plural inflection *-u* as in *kaʹtab**u*** "they wrote" was also expected to be challenging for the DLD group since the inflection is a short vowel that always occurs in a word-final unstressed syllable. However, the children with DLD in Study 1 did not differ from their TD peers in using plural verbs. Study 1, therefore, indicates that Arabic-speaking children with DLD may show difficulties in perceiving grammatical inflections of low acoustic saliency. However, this is unlikely to be the only factor that underpins their expressive verb morphology deficits.

Overall, none of the above-mentioned accounts could fully explain the range of verb morphology difficulties in Arabic-speaking children with DLD. The findings of Study 1 provide evidence of limitations in linguistic knowledge, which explain the optional marking of tense and agreement by Arabic-speaking children with DLD. Study 1 findings also suggest limitations in processing resources which may result in incomplete/inadequate processing of tense and agreement inflections by children with DLD. Hence, a combination of both theoretical views could better explain the verb morphology production difficulties in Arabic-speaking children with DLD.

5.2.2 Explaining the nonword repetition deficits

In their processing-based view, Gathercole and Baddeley (1990) proposed that the root of the nonword repetition difficulties in children with DLD is a limitation in **their verbal short-term memory** capacity. Therefore, when the nonwords increase in length, the nonword repetition accuracy will be more adversely affected in children with DLD than TD children. This is because longer nonwords would exceed the capacity of the verbal short term memory of children with DLD leading to more repetition errors. As predicted, the nonword repetition accuracy of children with DLD in Study 2 decreased significantly for three-syllable nonwords relative to one and two-syllable nonwords. This effect was also evident in the TD group. Hence, in contrast to Gathercole and Baddeley (1990)'s view, nonword length influenced the performance of the DLD and TD groups to a similar degree. Furthermore, children with DLD were significantly less accurate than TD peers in repeating one and two-syllable nonwords. These nonwords are short and are not expected to overload the verbal short-term memory of children with DLD. This suggests that mechanisms or processes, other than limitations in verbal short-term memory, may underlie poor nonword repetition in Arabic-speaking children with DLD.

Study 2 found that phonological complexity differentially affected TD and DLD groups. The TD group repeated nonwords with one, two, or no consonant clusters with similar accuracy. In contrast, the DLD group exhibited pronounced difficulties with the repetition of nonwords containing two consonant clusters compared to nonwords with one or no consonant clusters. The **Computational Grammatical Complexity (CGC) account**, a linguistic model, may account for this finding. This framework posits that children with DLD have a deficit in the computation of complex grammar, including syntax, morphology and phonology (van der Lely, 2005). In terms of phonology, the CGC proposes that children with DLD have difficulties with marked phonological structures such as complex syllabic structures that contain consonant clusters (Briscoe et al., 2001; Gallon et al., 2007; Marshall & van der Lely, 2007). Gallon (2007) argues that the greater the number of marked structures (e.g., consonant clusters) a nonword contains, the more complex it is. According to Gallon (2007), complex syllabic structures are available to children with DLD, but the processing and/or the reproduction of these structures is treated as being optional. It is also suggested that the increased phonological complexity is also associated with increased articulatory complexity of the nonwords. The articulation of consonant clusters places higher demands (than singleton consonants) on the speech motor output processes since their production involves coordinating many articulatory movements within syllables (Archibald et al., 2013). In Study 2, the speech production abilities of the children were not measured, and therefore this conclusion can only be tentative.

In Study 2, the TD and DLD groups found nonwords with high wordlikeness ratings easier to repeat than nonwords with low wordlikeness ratings. Neither the verbal short-term memory account (Gathercole & Baddeley, 1990) nor the CGC account (van der Lely, 2005) can accommodate for this effect. The presence of the wordlikeness effect reflects the occurrence of a cognitive process called redintegration. During the repetition of nonwords with high wordlikeness ratings, when an incomplete trace of the nonword is created in verbal short-term memory, lexical and phonotactic information is retrieved from long-term memory to "fill in the gaps" of the phonological representation of the nonword, leading to a correct or near-correct repetition (Stokes et al., 2006). Hence, existing lexical knowledge may facilitate the repetition of nonwords rated to sound similar to real words in Arabic. The lower accuracy of the DLD group in repeating high word-like nonwords relative to their TD peers suggests that they were less efficient using redintegration to form new phonological representations. Arabic-speaking children with DLD are characterized by limitations in their expressive and receptive vocabulary (Balilah, 2017; Khater, 2016; Shaalan, 2010). Accordingly, less lexical information may be available to these children to aid nonword repetition. The vocabulary abilities of the children were not measured in this thesis. Hence this interpretation requires further investigation.

The DLD group also underperformed the TD group in repeating nonwords with low wordlikeness ratings where the facilitative effect from previous linguistic knowledge is limited. Some researchers (e.g., Gathercole, 1995) argue that the repetition of these items relies more heavily on verbal short-term capacity, which (as we found) is limited in Arabic-speaking children with DLD. This may explain the lower performance of the DLD group in repeating less word-like nonwords. This interpretation is, however, inconclusive. Indeed, Study 2 showed that nonword repetition not only taps into verbal short-term memory but also taps into representations of complex phonological structures. Examining the correlations between the performance on nonwords with low wordlikeness ratings and measures of verbal short-term memory (e.g., digit recall) or measures of articulatory complexity (production accuracy of consonant clusters) is necessary to determine the processes that underlie poor repetition of nonwords with low wordlikeness ratings.

To summarize, in contrast to Gathercole and Baddeley (1990)'s proposal, the findings of Study 2 show that poor nonword repetition in Arabic-speaking children with DLD reflects deficits that are over and above limitations in their verbal short-term memory. Consistent with the CGC account, Arabic-speaking children with DLD experienced more pronounced difficulties with the repetition of nonwords containing consonant clusters suggesting optionality in the representations/reproduction of these structures. An alternative explanation that requires further research is that difficulties with complex structures may reflect articulation or motor planning difficulties in children with DLD. The presence of a wordlikeness effect indicates that poor nonword repetition abilities in Arabic-speaking children with DLD may also reflect less efficient access to lexical knowledge to support the recreation of nonwords. Study 2 highlights that poor nonword repetition in Arabic-speaking children with DLD may result from a combination of limitations in verbal-short term memory, inefficient access to linguistic representations in long-term memory, and deficits in the representations/production of complex syllabic structures. This is in line with recent evidence showing that irrespective of speech or language abilities, nonword repetition is strongly predicted by verbal short-term memory followed by oromotor sequencing, word reading, and oromotor control (Pigdon et al., 2019). Further research into this area is needed to verify the mechanisms underlying poor nonword repetition in Arabic-speaking children with DLD.

5.2.3 Explaining the sentence repetition deficits

Study 3 shows that the Arabic-speaking children with DLD scored significantly below their age-matched TD peers on the sentence repetition task. Multiple components appear to drive sentence repetition performance, including linguistic representations in long-term memory, i.e., language knowledge (Klem et al., 2015; Lombardi & Potter, 1992; Polišenská et al., 2015; Riches, 2012; Tuller et al., 2018), verbal short-term memory and verbal working memory (Alloway & Gathercole, 2005; Poll et al., 2013, 2016; Riches, 2012; Willis & Gathercole, 2001) as well as non-verbal working memory (Ebert, 2014). Therefore, the lower sentence repetition scores of Arabic-speaking children with DLD may reflect deficits in any of the aforementioned mechanisms.

Study 3 revealed that, as the sentence length increased (as indexed by the number of syllables), the sentence repetition accuracy of children with and without DLD decreased. The adverse effect of length on repetition accuracy suggests that sentence repetition is constrained by the verbal short-term memory capacity, with longer sentences placing a greater load than shorter sentences on verbal short-term memory capacity (Alloway & Gathercole, 2005; Hesketh & Conti-Ramsden, 2013; Willis & Gathercole, 2001). The lower sentence repetition scores of Arabic-speaking children with DLD may be caused, in part, by a deficit in their verbal short-term memory (as documented in Study 2).

It is important not to confound the effect of length (a memory-related factor) with the effect of syntactic complexity (a language-related factor). In the sentence repetition task used in Study 3, longer sentences were also syntactically more complex. Hence, further analysis focused on sentences of equal length but with different complexity. It revealed that TD children repeated sentences of increasing syntactic complexity with similar accuracy. On the other hand, the repetition accuracy in children with DLD decreased significantly as the syntactic complexity of the sentences increased. This effect of syntactic complexity suggests that sentence repetition tasks tap into the syntactic representations in long-term memory (Frizelle et al., 2017; Klem et al., 2015; Riches et al., 2010). The finding that syntactic complexity affected TD and DLD groups differently points to a deficit in the linguistic representational knowledge of children with DLD. Consistent with this conclusion, the DLD group in Study 3 performed below TD peers on a set of background morphological production tasks (verb morphology and noun plurals production) and used significantly less complex morpho-syntactic structures and morphological inflections (as indexed by their lower mean morpheme per utterance scores). This indicates that the DLD group in Study 3 had impaired grammatical abilities.

The CGC account provides a linguistic explanation of the syntactic complexity effect in the DLD group. The CGC framework posits that the syntactic difficulties in children with DLD stem from an underlying deficit in the computation of syntactic dependencies at the clause level (van der Lely & Marshall, 2010; van der Lely, 1998, 2005). It is assumed that the representations and/or mechanisms responsible for building such dependencies are available in children with DLD. However, their implementation is optional. Therefore, the CGC predicts that the production of hierarchal structures involving syntactic dependency operations (such as movement and embedding) is impaired in children with DLD. The findings of Study 3 were consistent with the predictions of the CGC account. Recall that sentences within levels 2 and 3 were problematic for Arabic-speaking children with DLD. Level 2 contained movement-derived structures such as passives, object Wh-questions and sentences with clitic left dislocation and structures with embedding, including bi-clausal sentences with coordination or complementizers. Level 3 included object relatives, which require both embedding and movement, and structures with embedding, including conditionals, subject relatives and sentences with subordination. The increased difficulty that the DLD group experienced with sentences within level 3 may be due to the inclusion of object relatives in this level. Object relatives involved both movement and embedding, whereas sentences within level 2 contained movement-only or embedding-only structures. In the DLD group, the average repetition accuracy of object relatives was 34%. This was lower than their level 2 overall repetition accuracy of 47%. This finding is in line with the CGC account, which predicts that children with DLD will make more errors as the number of syntactic dependencies increase. Notably, the most challenging structures within level 3 were conditional sentences. Conditionals are not only syntactically complex (e.g., involve embedding) but may also be semantically complex (Duman et al., 2015).

The syntactic complexity effect may be alternatively explained following a processing-based approach. Several studies have established a link between the production and/or comprehension of syntactically complex sentences and working memory in children with DLD (Delage & Frauenfelder, 2020; Durrleman & Delage, 2016; Frizelle & Fletcher, 2015; Marinis & Saddy, 2013; Montgomery et al., 2018; Montgomery & Evans, 2009; Zebib et al., 2020). Delage and Frauenfelder (2020) found that a combination of age, verbal short-term memory measures and verbal working memory measures explained 51% and 58% of the variance in the repetition of complex sentences in TD children and children with DLD, respectively. Frizelle and Fletcher (2015) found a correlation between the repetition of complex relative clauses and measures of verbal working memory (listening span) in children with DLD but not in TD children. For the latter group, the repetition of simple and complex relative clauses correlated with their verbal short-term memory capacity (indexed by digit recall). Frizelle and Fletcher (2015) suggested that children with DLD rely on their verbal working memory to repeat syntactically complex structures. On the other hand, TD children who have sufficient knowledge of complex syntax rely on their passive verbal storage to repeat these structures. Similarly, Zebib et al. (2020) found that in bilingual TD children whose language is efficient, sentence repetition was predicted by language measures. In contrast, in bilingual children with DLD, whose language is deficient, sentence repetition was predicted by their general processing abilities, i.e., working memory. Generally, it appears that when language deficits are evident (as is the case in children with DLD), the repetition of complex syntax will rely more heavily on general processing abilities. As explained earlier, the presence of a syntactic complexity effect suggests that the sentence repetition task tapped into the syntactic knowledge of the children.

Given that the grammatical abilities of Arabic-speaking children with DLD in Study 3 are impaired (based on their performance on the background measures), it may be the case that they had to rely on their working memory to repeat the syntactically complex structures. No measures of working memory were obtained for the children in Study 3. However, previous studies have documented weaknesses in the verbal working memory (backwards digit recall) of Arabic-speaking children with DLD (Balilah, 2017).

According to Jakubowicz (2011), working memory deficits in children with DLD will lead to limitations in the processing of complex syntax. This, in turn, may explain the pronounced deficits of the DLD group in repeating syntactically complex sentences (e.g., subject and object relative clauses, passives, sentences with clitic left dislocation, Wh- object questions…). The number and nature of syntactic operations necessary for sentence processing are assumed to be associated with the cognitive load involved in processing complex syntax (Jakubowicz, 2011a). Therefore, sentences in level 3 may have been more taxing to the working memory of the DLD group compared to sentences within level 2. This may explain why the latter set of sentences were repeated more accurately by Arabic-speaking children with DLD.

The findings of Study 3 suggest that the sentence repetition deficits in Arabic-speaking children with DLD may stem from a combination of limitations in verbal short-term and working memory and linguistic knowledge. To better understand the locus of sentence repetition deficits in Arabic-speaking children with DLD, it will be essential to examine the correlations and predictive relationships between verbal and non-verbal working memory, verbal short-term memory, and measures of language knowledge and sentence repetition performance in children with and without DLD. It will also be essential to assess other aspects of syntactic complexity such as production as well as comprehension and examine how these aspects link to the other memory and language measures.

The findings of this thesis call into question the dichotomy of theoretical views on the nature of DLD. Neither linguistic nor processing models alone could fully explain the pattern of deficits in Arabic-speaking children with DLD. Instead of viewing processing and linguistic accounts as competing theories, it may be more productive to emphasize how these views complement one another. The language difficulties of Arabic-speaking children with DLD seem to be a consequence of a combination of limitations in processing abilities (i.e., verbal short-term and working memory) and linguistic knowledge. A significant short-coming of the existing DLD theoretical models is that none of them is adequately comprehensive to capture the heterogeneous nature of the disorder. Most of the available theories do not offer explanations of the non-linguistic deficits that are often associated with DLD. These include deficits in motor control/coordination, executive functioning, non-verbal memory and processing and production of rhythm. The presence of these deficits alongside the functional limitations associated with DLD supports the notion that a domain-general deficit may better suit the heterogeneity and multifactorial nature of DLD.

5.3 Clinical implications

The findings of this thesis form a stepping-stone into advancing the diagnostic procedures of DLD in the Palestinian context and other Arab countries where speech and language therapy remains a relatively underdeveloped field. This thesis extends previous research by refining the description of the profiles of Arabic-speaking children with DLD. Specifically, the findings of this work inform SLTs of the language areas that they should be considered during assessment and intervention when working with Arabic-speaking children with DLD. This, in turn, may facilitate the early and accurate identification of DLD in this population.

Study 1 provides a detailed description of the verb morphology difficulties in Arabic-speaking children with DLD. Study 1 suggests that, when assessing the use of verb inflections in Arabic-speaking children, SLTs should pay close attention to present tense and feminine verbs. These forms were especially challenging for Arabic-speaking children with DLD. SLTs may use the verb elicitation task developed within Study 1 for this purpose.

Study 2 shows that poor nonword repetition may be a possible clinical marker of 4 to 6-year-old Arabic-speaking children with DLD. The Arabic version of a Quasi-Universal LITMUS nonword repetition test (dos Santos et al., n.d) showed a good discriminatory power in distinguishing between Arabic-speaking children with and without DLD. Study 2 therefore highlights the potential usefulness of this task as an indicator/index of DLD in Arabic. Study 2 suggests that children who score lower than 24 out of 30 on the nonword repetition task may require further assessment of their language abilities. The nonword repetition task employed a binary scoring method which makes it a quick measure that SLTs can use as part of their language assessment procedures.

Study 3 provides SLTs with a detailed and systematic description of the morpho-syntactic structures that are problematic for Arabic-speaking children with DLD. The Palestinian Arabic LITMUS Sentence Repetition task (LITMUS-SR-PA-72) showed good accuracy in differentiating 4 to 6-year-old Arabic-speaking children with DLD from same-age TD peers. Hence, SLTs may consider including sentence repetition as part of in their language evaluation protocols. Study 3 established cut-off scores across three different scoring methods of the LITMUS-SR-PA-72. SLTs may use these cut-off scores as reference points to determine children who require further assessment. The binary scoring method is quick to administer and may be used in the first instance to determine whether or not the child's grammatical abilities require further investigation. Error scoring is a more fine-grained scoring method that could help SLTs determine the severity/extent of the grammatical difficulties of the children. The structural scoring method could help identify the grammatical weaknesses and strengths of children. This may help SLTs in identifying the grammatical structures to target in the intervention. The scoring methods of the sentence repetition task provide quantitative values of the level of morpho-syntactic production. Hence, the task may be re-administered once the child has started intervention. A change in scores would indicate the progress on grammatical acquisition in Arabic-speaking children with DLD. This will be important not only for SLTs but also for parents to better understand the status of their child's progress in therapy.

It is important to note that in Study 2 and Study 3, the confidence intervals associated with the diagnostic accuracy metrics of the nonword and sentence repetition tasks included values that are outside of the "good" diagnostic accuracy criteria. Hence, it is recommended to use the nonword and sentence repetition tasks in combination with other tasks (e.g., informal assessment asks) to identify Arabic-speaking children with DLD. Overall, the present thesis provides three theoretically-based measures that could increase the accuracy of diagnosing DLD in Arabic. Specifically, Study 2 and Study 3 provide clinicians with nonword and sentence repetition tasks with determined cut-off points. These reference scores could guide SLTs in determining which children may require further assessment, making these tasks clinically relevant and applicable.

5.4 Limitations and future directions

In Study 1, a diagnostic accuracy analysis was not feasible due to limitations related to sample characteristics. Cases (i.e., children with DLD) and controls (i.e., TD children) are often matched with respect to covariates that are known to be associated with the clinical marker and its classification accuracy (Janes & Pepe, 2008a, 2008b). Regarding Study 1, this covariate is age which is known to be positively correlated with verb morphology production (Guo et al., 2019; Rice et al., 1998). Although the TD and DLD groups did not differ significantly in their chronological age, they were not age-matched which may introduce potential biases due to differential sample characteristics. Moreover, there were only 14 children with DLD across a wide age range. The inadequate sample size and lack of age-matching may therefore result in invalid diagnostic accuracy estimations. Future research should employ carefully designed studies with a larger sample size and appropriate age-matching procedures to determine the diagnostic accuracy of verb morphology production in identifying DLD in Arabic.

Study 1 did not include a language-matched TD group. Studies have traditionally included age and language-matched TD control groups to address whether the language development of children with DLD is merely delayed (i.e., compared to age-matched TD children), or deviant (i.e., compared to language-matched TD children). Such comparisons, however, rely on group mean scores, and do not take into account the changes in developmental patterns over time (Karmiloff-Smith et al., 2004). The developmental trajectory approach is a more recent and alternative approach to group matching. The aim of this approach is to construct a function linking performance on a specific experimental task with chronological /language age, and then to assess whether this function differs between the TD group and the disorder group (Thomas et al., 2012). This approach is suitable for cross-sectional designs, and it takes advantage of the wide age and ability ranges shown by the children in the disordered group. Another advantage of this approach is that it provides richer descriptive vocabulary to distinguish the different types of delay (for details, see Thomas et al., 2009). Employing the developmental trajectory approach on Arabic data would provide much-needed information about the pattern of typical and impaired language acquisition, especially verb morphology acquisition, in Arabic-speaking children.

Study 1 only assessed the production of verb morphology. Yet, research suggests that comprehension (Leonard, Miller, et al., 2000) and processing (e.g., Blom et al., 2014; Redmond & Rice, 2001; Rice et al., 1999, 2009) of verb morphology are also adversely affected by DLD. Examining the parallels between production, comprehension and processing aspects may help uncover the underlying cause of the verb morphology deficits associated with DLD in Arabic. If verb morphology comprehension impairments are evident in Arabic-speaking children with DLD, their verb morphology difficulties may be attributed to a deficit in their knowledge of the grammatical functions of verb inflections. On the other hand, intact verb morphology comprehension may suggest that there are other factors that may constrain the use of verb inflections such as their articulatory complexity (Bishop et al., 1996). Therefore, it will also be vital to assess whether children with DLD can produce the consonant and syllabic structures required to inflect verbs. The perceptual saliency (Leonard et al., 1997b) of the verb inflections in Arabic should also be considered. In Study 1, Arabic-speaking children with DLD showed lower accuracy in using present tense verb inflections that were not acoustically salient (e.g., unstressed). Therefore, future studies should examine this aspect in more depth. Study 1 investigated the use of 3rd person verb inflections only. To enhance our understanding of the extent and severity of verb morphology deficits in Arabic-speaking children with DLD, future investigations should also examine 1st and 2nd person verb inflections.

Study 2 and Study 3 showed a good diagnostic accuracy of nonword repetition and sentence repetition in identifying DLD in Arabic, respectively. These results are only preliminary and need to be replicated. Study 2 and Study 3 followed a two-gate design. TD and DLD groups were recruited from pre-selected samples (e.g., children with a prior DLD diagnosis versus children with typical language development). This could have resulted in two groups on the polarized ends of the spectrum of language abilities, i.e., a spectrum bias (Pawłowska, 2014; Redmond et al., 2019). That is a group of children with DLD with significant language deficits and a group of TD children with average language abilities. Therefore, it is possible that the accuracy of nonword repetition and sentence repetition tasks in detecting children with DLD (i.e., sensitivity) and excluding TD children (i.e., specificity) were inflated (Willis, 2008). Future studies are recommended to employ a one-gate design (Pawłowska, 2014) in which TD and children with DLD are recruited from a single population (unselected sample). Such design will allow for the inclusion of heterogeneous and representative samples of children with and without DLD leading to a more accurate and valid estimation of the diagnostic accuracy of the sentence repetition and nonword repetition tasks.

The diagnostic accuracy of a clinical marker is traditionally estimated through comparison to the best available reference/gold standard that is used to diagnose the condition (Umemneku Chikere et al., 2021). Due to the lack of Arabic language assessments, Study 2 and Study 3 used the following standards to recruit children with DLD: a) a prior diagnosis of DLD and b) scores at or below -1.5 SD on three informal language measures. The sole reliance on these two reference standards may be problematic. First, children with DLD whose language deficits are mild may not be identified by the informal language assessments, and hence, may not have a DLD diagnosis. There is a chance that these children were under-represented in Study 2 and Study 3. Second, DLD diagnostic criteria and assessment procedures vary from one clinic to another, limiting the results' replication and generalizability. Finally, the informal language measures that were used to confirm DLD only assessed morpho-syntax and phonology. It is unclear whether the diagnostic accuracy of sentence repetition and nonword repetition may apply in identifying children with DLD whose language deficits are prominent in other languages domains (e.g., pragmatics, lexical-semantics). Future studies on the diagnostic accuracy of nonword repetition, sentence repetition or other clinical markers in Arabic should include additional reference standards to ensure representative samples of children with DLD. These may include but are not limited to composite scores on one, a combination of, or sub-sets of standardized language tests, receipt of services and parental reports (Redmond et al., 2019). The two recently published standardized Arabic-language tests: *The Arabic language: Evaluation of function* (Rakhlin et al., 2021), or *Évaluation du langage oral chez l'enfant libanais* "Oral language assessment in Lebanese children" (Zebib et al., 2019) may be used for this purpose.

The performance of Arabic-speaking children with and without DLD on the nonword repetition and sentence repetition tasks improved significantly with age. This suggests that the diagnostic accuracy of these tasks may vary across different developmental stages. Hence, establishing the age-specific cut-off points for optimal classification accuracy is necessary. Study 2 and Study 3 administered the sentence repetition and nonword repetition tasks to the same participants. However, the diagnostic accuracy of each task was examined separately. Examining the combined diagnostic power of sentence repetition, nonword repetition, and other potential clinical markers may be more clinically relevant. Establishing the diagnostic accuracy of different combination of clinical marker tasks (e.g., Bonifacci et al., 2020; Conti-Ramsden, 2003; Poll et al., 2010) could inform clinicians of the most efficient clinical protocols to follow in order incorporate the clinical markers tasks in language assessment. For instance, this would inform clinicians which clinical marker has the best diagnostic accuracy, hence, given the priority in assessments, or which combination of markers is the most beneficial in identifying DLD in Arabic.

The use of live voice to administer repetition tasks has been widely applied within speech and language therapy practices (Chiat & Roy, 2007), and shown to be effective in engaging children more readily in these tasks and maintaining their attention (Chiat & Roy, 2007; Frizelle et al., 2017; Kapalková et al., 2013). Although the same examiner administered the tests to all children, there may have been inevitable variations in pitch, loudness and rate of stimulus presentation. These variations are likely to be even greater if different people administer the tests. Accordingly, the results of children on the tests will depend, in part, to the character tics of the person delivering the test. This may affect the comparability or replicability of the results if different researchers/clinicians administer the test. To ensure a homogenous and consistent delivery of the nonword repetition and sentence repetition tasks, future studies are recommended to create computerized versions of the tasks (in the form of a game) that utilize audio-recorded material. This approach will ensure consistent and homogenous delivery of the tasks and be equally engaging for the children.

As discussed in sections 5.2.1, 5.2.2 and 5.2.3, difficulties in verb morphology, nonword repetition and sentence repetition may stem from a combination of deficits in processing abilities and linguistic knowledge. It will be vital to assess the performance of Arabic-speaking children with DLD on different measures of verbal short-term memory (e.g., digit recall, word recall), verbal working memory (e.g., listening span, backward digit recall), visuospatial short-term and working memory, and long-term linguistic measures (e.g., production and comprehension of syntax, morphology and vocabulary). Investigating these measures simultaneously would allow for defining a comprehensive profile of the linguistic and non-linguistic weaknesses and strengths of Arabic-speaking children with DLD. More importantly, examining how these measures correlate with/predict verb morphology production, sentence repetition, and nonword repetition will help uncover mechanisms that underly the difficulties of Arabic-speaking children with DLD in these aspects.

Research on DLD in Arabic is just beginning to emerge and several research areas are yet to be investigated in Arabic-speaking contexts. Most of the studies available to date, including this thesis, have only focused on exploring the linguistic abilities of Arabic-speaking children with DLD. These studies (Abdallah & Crago, 2008; Fahim, 2017; Morsi, 2009; Faquih, 2014; Shaalan, 2010; Wallan, 2018) commonly have identified morpho-syntactic production as a core area of difficulty for children with DLD speaking different Arabic dialects. Once the underlying causes of these deficits are identified (as described above), the next step for future research is to consider designing intervention programs that target the grammatical deficits of Arabic-speaking children with DLD. It is well-documented that DLD is associated with long-standing functional limitations that affect different functional domains (e.g., Conti-Ramsden et al., 2019; Durkin & Conti-Ramsden, 2007; Lindsay & Dockrell, 2012; St Clair et al., 2011). An essential step for future research is to look beyond language and assess how DLD impacts the academic performance, social interactions and peer relations, and social-emotional well-being of Arabic-speaking children with DLD. Such information will raise the awareness of parents, educators and clinicians of the breadth of difficulties that children with DLD experience. This may also help facilitate access of Arabic-speaking children with DLD to specialist interventions to address these functional limitations. DLD is likely to persist into adolescence (e.g., Johnson et al., 1999) and adulthood (Botting, 2020; Clegg et al., 2005). To date, no study has examined the language abilities of Arabic-speaking adolescents or adults with language disorders. Data from this population could be used to characterize the trajectories of atypical language development in Arabic. Identifying persisting deficits in language or other functional areas may advance the support and services that adolescents and adults with DLD need.

5.5 Conclusion

This thesis investigated the potential of verb morphology use, nonword repetition and sentence repetition as possible clinical markers of DLD in Arabic. Arabic-speaking children with DLD showed difficulties with verb morphology relative to same-age TD peers. Their difficulties were particularly evident in producing present tense and feminine verbs; both of these forms could be possible grammatical markers of DLD in Arabic. Nonword and sentence repetition were also problematic for Arabic-speaking children with DLD compared to their age-matched TD peers. These tasks could also be potential clinical markers of Arabic-speaking children with DLD.

None of the prominent DLD theoretical accounts could fully cover the expressive verb morphology deficits, the nonword repetition difficulties and the poor sentence repetition abilities in Arabic-speaking children with DLD. The findings of this thesis suggest that the language difficulties of Arabic-speaking children with DLD stem from combined deficits in linguistic knowledge and processing abilities (verbal short-term and working memory). Therefore, a domain-general view of DLD appears to be more compatible with the data from Arabic and the multifactorial nature of DLD.

The results of this thesis contribute to advancing the diagnostic practices of DLD in Arabic-speaking children by providing SLTs with information about the language deficits that characterize the profiles of Arabic-speaking children with DLD. Further research is necessary to establish and replicate the clinical viability of the verb morphology, nonword repetition and sentence repetition. Future research efforts should identify the linguistic, non-linguistic and functional limitations in Arabic-speaking children with DLD.

5.6 Beyond this thesis: The Arabic DLD platform

During my work on this thesis, I had the chance to work with Zakiyah Al-Siddiqi, an SLT and a PhD student at Reading University and Dr Aseel Alkadhi, an SLT and assistant professor in King Saud University, both of whom are from Saudi Arabia. The discussions of our clinical experiences in working with Arabic-speaking children with DLD were eye-opening to several issues. First, there is an inconsistency of the terminology that is used to describe Arabic-speaking children with DLD. For instance, the term “Language Delay” is continuously used by Arabic-speaking SLTs to refer to children who have persistent and unexplained language problems. It also is evident that, unlike Autism or hearing loss, DLD is much less known among the Arabic-speaking community. By chatting with the parents of children with DLD who participated in this PhD project, many said they expressed concerns that their child’s language abilities were different from same-age peers. However, they were advised by relatives and/or other healthcare professionals that their child would catch up. This “wait and see” culture may risk delaying the identification and the provision for intervention for Arabic-speaking children with DLD. Inspired by recent campaigns in English-speaking countries that focus on DLD, such as Raise Awareness of Developmental Language Disorder (radld.org), the DLD project (theDLDproject.com), engage with DLD (engage-dld.com) among others, we created the Arabic DLD platform (Twitter: DLDisorder, Website: dldisorderar.com). Through this platform, we aim to raise awareness and educate the Arabic-speaking community about DLD. We started doing so by creating resources in Arabic with information about DLD, its symptoms and features, and its long-term impact on other areas of functioning. The resources target the misconceptions/myths that are often associated with DLD. This is to inform parents, teachers, and other professionals working with children of the signs of DLD and to draw their attention to the importance of early identification and intervention of the disorder. We are currently creating resources for parents and teachers with information about strategies to support children with DLD at home or in educational settings. Another goal of the Arabic DLD platform is to summarize and share the latest research findings of the DLD literature, especially in Arabic, and create a database of the available published and unpublished Arabic language assessments. This is to increase access of Arabic-speaking SLTs to the up-to-date research and enhance the inclusion of evidence-based language measures as part of diagnosing DLD in Arabic-speaking children with DLD.

References

Abdalla, F., Aljenaie, K., & Mahfoudhi, A. (2013). Plural noun inflection in Kuwaiti Arabic-speaking children with and without Specific Language Impairment. *Journal of Child Language*, *40*(1), 139–168. https://doi.org/10.1017/S0305000912000499

Abdallah, F. (2002). *Specific Language Impairment in Arabic-speaking children: Deficits in morphosyntax. (Unpublished Thesis)*. McGill University.

Abdallah, F., & Crago, M. (2008). Verb morphology deficits in Arabic-speaking children with specific language impairment. *Applied Psycholinguistics*, *29*(2), 315–340. https://doi.org/10.1017/S0142716408080156

Abdu, D., & Abdu, S. (1986). *Dirasa fi mufradat tiflayn [A study on child vocabulary]*. Kuwait: Dhaat al-Salaasil Publication.

Abi-Aad, K., & Atallah, C. (2020). L'épreuve répétition de non-mots: LITMUS-NWR-LIBAN [nonword repetition: LITMUS-NWR-Lebanon]. In R. Zebib, P. Prévost, L. Tuller, & G. Henry (Eds.), *Plurilinguisme et Troubles Spécifiques du Langage au Liban [Multilingualism and Specific Language Disorders in Lebanon]* (pp. 79–92). Beyrouth : Presses universitaires de l'Université Saint-Joseph.

Abu-Ghazaleh, I. N. (1983). *Theme and the function of the verb in Palestinian Arabic narrative discourse.(Unpublished Thesis)*. University of Florida.

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research*, *38*(2), 403–414. https://doi.org/10.1044/JSHR.3802.403

Adams, C., & Bishop, D. V. M. (1989). Conversational characteristics of children with semantic-pragmatic disorder. I: Exchange structure, turntaking, repairs and cohesion. *International Journal of Language & Communication Disorders*, *24*(3), 211–239. https://doi.org/10.3109/13682828909019889

Adams, C., Clarke, E., & Haynes, R. (2009). Inference and sentence comprehension in children with specific or pragmatic language impairments. *International Journal of Language and Communication Disorders*, *44*(3), 301–318. https://doi.org/10.1080/13682820802051788

Adani, F., Stegenwallner-Schütz, M., Haendler, Y., & Zukowski, A. (2016). Elicited production of relative clauses in German: Evidence from typically developing children and children with specific language impairment: *First Language* , *36*(3), 203–227. https://doi.org/10.1177/0142723716648842

Adlof, S. M. (2017). Understanding Word Reading Difficulties in Children With SLI. *Perspectives of the ASHA Special Interest Groups*, *2*(1), 71–77. https://doi.org/10.1044/persp2.SIG1.71

Aguilar-Mediavilla, E. (2013). Comparative analysis of the acquisition of syllabic structure and errors in preschool children with SLI. *Anuario de Psicología/The UB Journal of Psychology*, *43*(2), 237–252. https://raco.cat/index.php/AnuarioPsicologia/article/view/271076

Aguilar-Mediavilla, E. M., Sanz-Torrent, M., & Serra-Raventos, M. S. (2002). A comparative study of the phonology of pre-school children with specific language impairment (SLI), language delay (LD) and normal acquisition. *Clinical Linguistics and Phonetics*, *16*(8), 573–596. https://doi.org/10.1080/02699200210148394

Ahufinger, N., Berglund-Barraza, A., Cruz-Santos, A., Ferinu, L., Andreu, L., Sanz-Torrent, M., & Evans, J. L. (2021). Consistency of a Nonword Repetition Task to Discriminate Children with and without Developmental Language Disorder in Catalan–Spanish and European Portuguese Speaking Children. *Children*, *8*(2), 85. https://doi.org/10.3390/children8020085

Al-Akeel, A. (1998). *The acquisition of Arabic language comprehension by Saudi children.(Unpublished Thesis)*. Newcastle University.

Al-Aqarbeh. (2011). *Finiteness in Jordanian Arabic: a semantic and morphosyntactic approach.(Unpublished Thesis)*. University of Kansas.

Aljenaie, K. (2000). The emergence of tense and agreement in Kuwaiti children speaking Arabic. *Reading Working Papers in Linguistics*, *4*(1), 1–24. http://www.personal.reading.ac.uk/~llsroach/wp4.pdf

Aljenaie, K. (2001). *The emergence of tense and agreement in Kuwaiti Arabic children.(Unpublished Thesis)*. University of Reading.

Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, *37*(4), 841–863. https://doi.org/10.1017/S0305000909990031

Alloway, T. P., & Gathercole, S. E. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology*, *17*(2), 207–220. https://doi.org/10.1080/09541440440000005

Almeida, L. de, Ferré, S., Morin, E., Prévost, P., Santos, C. dos, Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with Specific Language Impairment in

France. *Linguistic Approaches to Bilingualism*, *7*(3–4), 331–358. https://doi.org/10.1075/LAB.15019.ALM

Alsiddiqi, Z. A., Stojanovik, V., & Pagnamenta, E. (2021). Emergent literacy skills of Saudi Arabic speaking children with and without developmental language disorder. *Clinical Linguistics and Phonetics*, 1–18. https://doi.org/10.1080/02699206.2021.1955299

Alt, M., & Spaulding, T. (2011). The effect of time on word learning: An examination of decay of the memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of Communication Disorders*, *44*(6), 640–654. https://doi.org/10.1016/j.jcomdis.2011.07.001

Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, *41*(3), 642–653. https://doi.org/10.1044/jslhr.4103.642

Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Villanueva, L. (2016). Social cognition makes an independent contribution to peer relations in children with Specific Language Impairment. *Research in Developmental Disabilities*, *49–50*, 277–290. https://doi.org/10.1016/j.ridd.2015.12.015

Andrés-Roqueta, C., & Katsos, N. (2020). A distinction between linguistic and social pragmatics helps the precise characterization of pragmatic challenges in children with autism spectrum disorders and developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1494–1508. https://doi.org/10.1044/2020_JSLHR-19-00263

Archibald, L. M. (2016). Working memory and language learning: A review. *Child Language Teaching and Therapy*, *33*(1), 5–17. https://doi.org/10.1177/0265659016654206

Archibald, L. M. (2018). The Reciprocal Influences of Working Memory and Linguistic Knowledge on Language Performance: Considerations for the Assessment of Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *49*(3), 424–433. https://doi.org/10.1044/2018_LSHSS-17-0094

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language and Communication Disorders*, *41*(6), 675–693. https://doi.org/10.1080/13682820500442602

Archibald, Lisa M. D., & Griebeling, K. H. (2016). Rethinking the connection between working memory and language impairment. *International Journal of Language & Communication Disorders*, *51*(3), 252–264. https://doi.org/10.1111/1460-6984.12202

Archibald, Lisa M.D., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Archibald, Lisa M.D., Joanisse, M. F., & Munson, B. (2013). Motor control and nonword repetition in specific working memory impairment and SLI. *Topics in Language Disorders*, *33*(3), 255–267. https://doi.org/10.1097/TLD.0b013e31829cf5e7

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters.

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of Specific Language Impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language and Communication Disorders*, *51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clauses by children. In J. Brucart, J. M., Gavarró, A., & Solà (Ed.), *Merging Features: Computation, Interpretation, and Acquisition*. Oxford University Press.

Arosio, F., Branchini, C., Barbieri, L., & Guasti, M. T. (2014). Failure to produce direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children. *Clinical Linguistics & Phonetics* , *28*(9), 639–663. https://doi.org/10.3109/02699206.2013.877081

Arosio, F., & Guasti, M. T. (2019). The production of wh-questions in Italian-speaking children with SLI. *Clinical Linguistics and Phonetics*, *33*(4), 349–375. https://doi.org/10.1080/02699206.2018.1517191

Ash, A. C., & Redmond, S. M. (2014). Using finiteness as a clinical marker to identify language impairment. *Perspectives on Language Learning and Education*, *21*(4), 148–158. https://doi.org/10.1044/lle21.4.148

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. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/ANNUREV-PSYCH-120710-100422

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation -*

Advances in Research and Theory, *8*(C), 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Gathercole, S. E., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *American Psychological Association*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bakopoulou, I., & Dockrell, J. E. (2016). The role of social cognition and prosocial behaviour in relation to the socio-emotional functioning of primary aged children with specific language impairment. *Research in Developmental Disabilities*, *49–50*, 354–370. https://doi.org/10.1016/j.ridd.2015.12.013

Balilah, A. M. (2017). *Linguistic and Cognitive Measures in Arabic-Speaking English Language Learners (ELLs) and monolingual children with and without Developmental Language Disorder (DLD) (Unpublished Doctoral Thesis)*. The University of Western Ontario.

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. https://doi.org/10.1016/j.jml.2012.11.001

Basaffar, F., & Safi, S. (2012). The acquisition of verb inflections in Hijazi Arabic. *Arab World English Journal*, *3*, 266–304. https://awej.org/images/AllIssues/Volume3/Volume3Number2June2012/11.pdf

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children With specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*(4), 905–924. https://doi.org/10.1044/1092-4388(2001/072)

Bedore, L. M., & Leonard, L. B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics*, *26*(2), 195–225. https://doi.org/10.1017/s0142716405050149

Beers, W. (1995). *The Phonology of Normally Developing and Language-Impaired Children (Unpublished Thesis)*. Universiteit van Amsterdam.

Benmamoun, E. (1999). Arabic morphology: The central role of the imperfective. *Lingua*, *108*(2–3), 175–201. https://doi.org/10.1016/S0024-3841(98)00045-X

Benmamoun, E. (2000). The Tense Systems of Egyptian Arabic, Moroccan Arabic, and Standard Arabic. In *The feature structure of functional categories: A comparative study of Arabic dialects* (pp. 19–36). New York & Oxford: Oxford University Press.

Bishop, D. V. M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In Dotothy V.M Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: causes, characteristics ,intervention and outcome* (pp. 99–113). Philadelphia: PA: Psychology Press Ltd. https://doi.org/10.4324/9781315784878-12

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, *49*(4), 381–415. https://doi.org/10.1111/1460-6984.12101

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behavior*, *5*(2), 158–169. https://doi.org/10.1111/j.1601-183X.2005.00148.x

Bishop, D. V. M., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology*, *3*(1), 1–12. https://doi.org/10.1186/S40359-015-0062-7

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword Repetition as a Behavioural Marker for Inherited Language Impairment: Evidence From a Twin Study. *Journal of Child Psychology and Psychiatry*, *37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., … Whitehouse, A. (2016). CATALISE: A multinational and multidisciplinary delphi consensus study. Identifying language impairments in children. *PLoS ONE*, *11*(7), 1–27. https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., … Whitehouse, A. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V.M., & Adams, C. (1991). What do referential communication tasks measure? A study of children with specific language impairment. *Applied Psycholinguistics*, *12*(2), 199–215. https://doi.org/10.1017/S0142716400009140

Bishop, D. V.M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research*, *35*(1), 119–129. https://doi.org/10.1044/jshr.3501.119

Bishop, D. V.M., Chan, J., Adams, C., Hartley, J., & Weir, F. (2000). Conversational responsiveness in specific language impairment: Evidence of disproportionate pragmatic difficulties in a subset of children. *Development and Psychopathology*, *12*(2), 177–199. https://doi.org/10.1017/S0954579400002042

Blom, E., Vasić, N., & De Jong, J. (2014). Production and Processing of Subject-Verb Agreement in Monolingual Dutch Children With Specific Language Impairment. *Journal of Speech, Language & Hearing Research*, *57*, 952–965. https://doi.org/10.1044/2014_JSLHR-L-13-0104

Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, M., & Mari, R. (2020). Which Measures Better Discriminate Language Minority Bilingual Children With and Without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1898–1915. https://doi.org/10.1044/2020_JSLHR-19-00100

Bortolini, U., Arfé, B., Caselli, M. C., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *International Journal of Language and Communication Disorders*, *41*(6), 695–712. https://doi.org/10.1080/13682820600570831

Bortolini, U., Caselli, M. C., & Leonard, L. B. (1997). Grammatical Deficits in Italian-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *40*(4), 809–820. https://doi.org/10.1044/jslhr.4004.809

Bortolini, U., & Leonard, L. B. (2000). Phonology and children with specific language impairment: Status of structural constraints in two languages. *Journal of Communication*

Disorders, *33*(2), 131–150. https://doi.org/10.1016/s0021-9924(99)00028-3

Botting, N. (2020). Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, *55*(2), 255–265. https://doi.org/10.1111/1460-6984.12518

Botting, N., & Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *International Journal of Language and Communication Disorders*, *40*(1), 49–66. https://doi.org/10.1080/13682820410001723390

Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated Reading Skills in Children with a History of Specific Language Impairment (SLI). *Reading and Writing*, *19*(1), 77–98. https://doi.org/10.1007/S11145-005-4322-4

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, *22*(3), 441–469. https://doi.org/10.1017/S0142716401003083

Bowyer-Crane, C., Snowling, M. J., Duff, F., & Hulme, C. (2011). Response to Early Intervention of Children with Specific and General Language Impairment. *Learning Disabilities: A Contemporary Journal*, *9*(2), 107–121. https://eric.ed.gov/?id=EJ954697

Brewer, C. C., Zalewski, C. K., King, K. A., Zobay, O., Riley, A., Ferguson, M. A., Bird, J. E., McCabe, M. M., Hood, L. J., Drayna, D., Griffith, A. J., Morell, R. J., Friedman, T. B., & Moore, D. R. (2016). Heritability of non-speech auditory processing skills. *European Journal of Human Genetics*, *24*(8), 1137–1144. https://doi.org/10.1038/ejhg.2015.277

Brinton, B., Fujiki, M., & Powell, J. M. (1997). The ability of children with language impairment to manipulate topic in a structured task. *Language, Speech, and Hearing Services in Schools*, *28*(1), 3–11. https://doi.org/10.1044/0161-1461.2801.03

Briscoe, J., Bishop, D. V., & Norbury, C. F. (2001). Phonological Processing, Language, and Literacy: A Comparison of Children with Mild-to-moderate Sensorineural Hearing Loss and Those with Specific Language Impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(3), 329–340. https://doi.org/10.1017/S0021963001007041

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Bühler, D., Perovic, A., & Pouscoulous, N. (2018). Comprehension of novel metaphor in young children with Developmental Language Disorder. *Autism & Developmental Language Impairments*, *3*, 1–11. https://doi.org/10.1177/2396941518817229

Campbell, W. N., & Skarakis-Doyle, E. (2011). Innovations in measuring peer conflict resolution knowledge in children with LI: Exploring the accessibility of a visual analogue rating scale. *Journal of Communication Disorders*, *44*(2), 207–217. https://doi.org/10.1016/j.jcomdis.2010.10.001

Carter, J. V., Pan, J., Rai, S. N., & Galandiuk, S. (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery (United States)*, *159*(6), 1638–1645. https://doi.org/10.1016/j.surg.2015.12.029

Castilla-Earls, A., Auza, A., Pérez-Leroux, A. T., Fulcher-Rood, K., & Barr, C. (2020). Morphological Errors in Monolingual Spanish-Speaking Children With and Without Developmental Language Disorders. *Language, Speech, and Hearing Services in Schools*, *51*(2), 270. https://doi.org/10.1044/2019_LSHSS-19-00022

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Chen, J., Justice, L. M., Rhoad-Drogalis, A., Lin, T. J., & Sawyer, B. (2020). Social Networks of Children With Developmental Language Disorder in Inclusive Preschool Programs. *Child Development*, *91*(2), 471–487. https://doi.org/10.1111/cdev.13183

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 125–150). Bristol:Multilingual Matters.

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 429–443. https://doi.org/10.1044/1092-4388(2007/030)

Claessen, M., Leitão, S., Kane, R., & Williams, C. (2013). Phonological processing skills in specific language impairment. *International Journal of Speech-Language Pathology*, *15*(5), 471–483. https://doi.org/10.3109/17549507.2012.753110

Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. *De Gruyter Mouton*, *27*(5), 897–920. https://doi.org/10.1515/LING.1989.27.5.897

Clahsen, H. (1991). *Child Language and Developmental Dysphasia: Linguistic studies of the Acquisition of German*. Amersterdam: John Benjamins Publishing.

Clahsen, H., Bartke, S., & Göllner, S. (1997). Formal features in impaired grammars: A comparison of English and German SLI children. *Journal of Neurolinguistics*, *10*(2–3), 151–171. https://doi.org/10.1016/S0911-6044(97)00006-7

Clahsen, H., & Dalalakis, J. (1999). Tense and Agreement in Greek SLI: A Case Study. *Essex Research Reports in Linguistics*, *24*, 1–25. http://repository.essex.ac.uk/id/eprint/196

Clahsen, H., & Hansen, D. (1997). The Grammatical Agreement Deficit in Specific Language Impairment: Evidence from Therapy Experiments. In M. Gopnik (Ed.), *The Inheritance and Innateness of Grammars* (pp. 141–160). New York, NY: Oxford University Press.

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders - A follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *46*(2), 128–149. https://doi.org/10.1111/j.1469-7610.2004.00342.x

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, *43*(1), 1–40. https://doi.org/10.1080/13682820601116485

Coady, J., Evans, J. L., & Kluender, K. R. (2010). The role of phonotactic frequency in sentence repetition by children with specific language impairment. *International Journal of Language and Communication Disorders*, *45*(4), 494–509.

Cole, K. N., Schwartz, I. S., Notari, A. R., Dale, P. S., & Mills, P. E. (1995). Examination of the stability of two methods of defining specific language impairment. *Applied Psycholinguistics*, *16*(1), 103–124. https://doi.org/10.1017/S0142716400006445

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language and Hearing Research*, *46*, 1029–1037. https://doi.org/10.1044/1092-4388(2003/082)

Conti-Ramsden, G., & Botting, N. (2004). Social Difficulties and Victimization in Children with SLI at 11 Years of Age. *Journal of Speech, Language, and Hearing Research*, *47*(1), 145–161. https://doi.org/10.1044/1092-4388(2004/013)

Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *49*(5), 516–525. https://doi.org/10.1111/J.1469-7610.2007.01858.X

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Conti-Ramsden, G., & Durkin, K. (2008). Language and Independence in Adolescents With and

Without a History of Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *51*(1), 70–83. https://doi.org/10.1044/1092-4388(2008/005)

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, *53*(2), 237–255. https://doi.org/10.1111/1460-6984.12338

Conti-Ramsden, G., Mok, P., Durkin, K., Pickles, A., Toseeb, U., & Botting, N. (2019). Do emotional difficulties and peer problems occur together from childhood to adolescence? The case of children with a history of developmental language disorder (DLD). *European Child and Adolescent Psychiatry*, *28*(7), 993–1004. https://doi.org/10.1007/s00787-018-1261-6

Conti-Ramsden, G., Mok, P. L. H., Pickles, A., & Durkin, K. (2013). Adolescents with a history of specific language impairment (SLI): Strengths and difficulties in social, emotional and behavioral functioning. *Research in Developmental Disabilities*, *34*(11). https://doi.org/10.1016/J.RIDD.2013.08.043

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.

Crago, M. B., & Allen, S. E. M. (2001). Early Finiteness in Inuktitut: The Role of Language Structure and Input. *Language Acquisition*, *9*(1), 59–111. https://doi.org/10.1207/S15327817LA0901_02

de Bree, E., Rispens, J., & Gerrits, E. (2007). Non-word repetition in Dutch children with (a risk of) dyslexia and SLI. *Clinical Linguistics and Phonetics*, *21*(11–12), 935–944. https://doi.org/10.1080/02699200701576892

De López, K. J., Olsen, L. S., & Chondrogianni, V. (2014). Annoying Danish relatives: Comprehension and production of relative clauses by Danish children with and without SLI*. *Journal of Child Language*, *41*(1), 51–83. https://doi.org/10.1017/S0305000912000517

Deevy, P., & Leonard, L. B. (2004). The comprehension of Wh- questions in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *47*(August), 802–816. https://doi.org/10.1044/1092-4388(2004/060)

Deevy, P., Weil, L. W., Leonard, L. B., & Goffman, L. (2010). Extending use of the NRT to preschool-age children with and without specific language impairment. *Language,* *Speech, and Hearing Services in Schools*, *41*(3), 277–288. https://doi.org/0.1044/0161-1461(2009/08-0096)

Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with Developmental Language Disorder. *Journal of Child Language*, *47*, 600–632. https://doi.org/10.1017/S0305000919000722

Delage, H., Stanford, E., & Durrleman, S. (2021). Working memory training enhances complex syntax in children with Developmental Language Disorder. *Applied Psycholinguistics*, *42*(5), 1341–1375. https://doi.org/10.1017/S0142716421000369

Devescovi, A., & Caselli, M. C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language and Communication Disorders*, *42*(2), 187–208. https://doi.org/10.1080/13682820601030686

Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-Word and Nonword Repetition in Italian-Speaking Children with Specific Language Impairment: A Study of Diagnostic Accuracy. *Journal of Speech, Language, and Hearing Research*, *56*(February), 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics*, *19*(2), 193–207. https://doi.org/10.1017/s0142716400010031

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146. https://doi.org/doi.org/10.1044/jslhr.4105.1136

Dollaghan, C A, & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech Language & Hearing Research*, *54*, 1077–1088. https://doi.org/10.1044/1092-4388(2010/10-0093)

Dollaghan, Chris A. (2007). *The handbook for evidence-based practice in communication disorders*. Brookes.

dos Santos, C., & Ferré, S. (2018). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Taylor & Francis*, *25*(1), 58–71.

dos Santos, C., Ferré, S., Zebib, R., & Abou Melhem, N. (n.d.). The Non-word repetition task—Lebanese [Unpublished material]. Adapted from the LITMUS-NWR-FRENCH task.

Dromi, E., & Berman, R. A. (1982). A morphemic measure of early language development: Data from modern Hebrew. *Journal of Child Language*, *9*(2), 403–424.

https://doi.org/10.1017/S0305000900004785

Dromi, E., Leonard, L. B., Adam, G., & Zadunaisky-Ehrlich, S. (1999). Verb Agreement Morphology in Hebrew-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1414–1431. https://doi.org/10.1044/jslhr.4206.1414

Dromi, E., Leonard, L. B., & Shteiman, M. (1993). The Grammatical Morphology of Hebrew-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *36*(4), 760–771. https://doi.org/10.1044/jshr.3604.760

Dubois, P., St-Pierre, M.-C., Desmarais, C., & Guay, F. (2020). *Young Adults With Developmental Language Disorder: A Systematic Review of Education, Employment, and Independent Living Outcomes*. *63*(11), 3786–3800. https://doi.org/10.1044/2020_JSLHR-20-00127

Duinmeijer, I., Jong, J. de, & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders*, *47*(5), 542–555. https://doi.org/10.1111/J.1460-6984.2012.00164.X

Duman, T. Y., Blom, E., & Topbaş, S. (2015). *At the Intersection of Cognition and Grammar: Deficits Comprehending Counterfactuals in Turkish Children With Specific Language Impairment*. *58*(2), 410–421. https://doi.org/10.1044/2015_JSLHR-L-14-0054

Durkin, K., & Conti-Ramsden, G. (2007). Language, Social Behavior, and the Quality of Friendships in Adolescents With and Without a History of Specific Language Impairment. *Child Development*, *78*(5), 1441–1457. https://doi.org/10.1111/J.1467-8624.2007.01076.X

Durkin, K., Toseeb, U., Botting, N., Pickles, A., & Conti-Ramsden, G. (2017). Social Confidence in Early Adulthood Among Young People With and Without a History of Language Impairment. *Journal of Speech, Language and Hearing Research*, *60*(6), 1635–1647. https://doi.org/10.1044/2017_JSLHR-L-16-0256

Durrleman, S., & Delage, H. (2016). Autism Spectrum Disorder and Specific Language Impairment: Overlaps in Syntactic Profiles. *Language Acquisition*, *23*(4), 361–386. https://doi.org/10.1080/10489223.2016.1179741

Eadie, P., Conway, L., Hallenstein, B., Mensah, F., McKean, C., & Reilly, S. (2018). Quality of life in children with developmental language disorder. *International Journal of Language & Communication Disorders*, *53*(4), 799–810. https://doi.org/10.1111/1460-6984.12385

Ebbels, S. (2014a). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, *30*(1), 7–40. https://doi.org/10.1177/0265659013512321

Ebbels, S. (2014b). Introducing the SLI debate. *International Journal of Language & Communication Disorders*, *49*(4), 377–380. https://doi.org/10.1111/1460-6984.12119

Ebbels, S., & van der Lely, H. (2001). Meta-syntactic therapy using visual coding for children with severe persistent SLI. *International Journal of Language and Communication Disorders*, *36*, 345–350. https://doi.org/10.3109/13682820109177909

Ebert, K. D. (2014). Role of auditory non-verbal working memory in sentence repetition for bilingual children with primary language impairment. *International Journal of Language & Communication Disorders*, *49*(5), 631–636. https://doi.org/10.1111/1460-6984.12090

Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics*, *19*(2), 279–309. https://doi.org/doi.org/10.1017/S0142716400010079

Eisenbeiss, S., Bartke, S., & Clahsen, H. (2006). Language Acquisition Structural and Lexical Case in Child German: Evidence From Language-Impaired and Typically Developing Children Structural and Lexical Case in Child German: Evidence From Language-Impaired and Typically Developing Children. *Language*, *13*(1), 3–32. https://doi.org/10.1207/s15327817la1301_2

Eisenberg, S. (2004). Production of infinitives by 5-year-old children with language-impairment on an elicitation task. *First Language*, *24*(3), 305–321. https://doi.org/10.1177/0142723704045679

Fahim, D. (2005). *Developmental language impairment in Egyptian Arabic.(Unpublished Thesis)*. Birkbeck, University of London.

Fahim, D. (2017). Verb Morphology in Egyptian Arabic Developmental Language Impairment. *Arab Journal of Applied Linguistics*, *2*(1), 49–73. https://eric.ed.gov/?id=EJ1207995

Faquih, N. O. (2014). *Production of Arabic bound pronouns by typically developing children and by children with language disorders (Unpublished Thesis)*. Horward University.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1301–1318. https://doi.org/10.1044/1092-4388(2004/098)

Field, A. (2009). *Discovering Statistics using IBM SPSS Statistics* (Third edit). London: SAGE Publications Ltd.

Flapper, B. C. T., & Schoemaker, M. M. (2013). Developmental Coordination Disorder in children with specific language impairment: Co-morbidity and impact on quality of life. *Research in Developmental Disabilities*, *34*(2), 756–763. https://doi.org/10.1016/j.ridd.2012.10.014

Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., & Zebib, R. (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, *25*(1), 85–101. https://doi.org/10.1080/10489223.2016.1192635

Ford, J. A., & Milosky, L. M. (2008). Inference generation during discourse and its relation to social competence: An online investigation of abilities of children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *51*(2), 367–380. https://doi.org/10.1044/1092-4388(2008/027)

Freed, J., Lockton, E., & Adams, C. (2012). Short-term and working memory skills in primary school-aged children with specific language impairment and children with pragmatic language impairment: phonological, linguistic and visuo-spatial aspects. *International Journal of Language & Communication Disorders*, *47*(4), 457–466. https://doi.org/10.1111/J.1460-6984.2012.00148.X

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. *Journal of Child Language*, *31*(3), 661–681. https://doi.org/10.1017/s0305000904006269

Friedmann, N, & Novogrodsky, R. (2008). Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI. In A. Gavarró & M. João Freitas (Eds.), *Language acquisition and development* (pp. 205–217). Cambridge Scholars Press. https://doi.org/10.4324/9780203768990-9

Friedmann, Naama, & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, *121*(3), 367–382. https://doi.org/10.1016/j.lingua.2010.10.004

Frizelle, P., & Fletcher, P. (2014). Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders*, *49*(2), 255–264. https://doi.org/10.1111/1460-6984.12070

Frizelle, P., & Fletcher, P. (2015). The Role of Memory in Processing Relative Clauses in Children With Specific Language Impairment. *American Journal of Speech-Language Pathology*, *24*(1), 47–59. https://doi.org/10.1044/2014_AJSLP-13-0153

Frizelle, P., O'neill, C., & Bishop, D. V. M. (2017). Assessing understanding of relative clauses: a comparison of multiple-choice comprehension versus sentence repetition. *Journal of Child Language*, *44*, 1435–1457. https://doi.org/10.1017/S0305000916000635

Fujiki, M., Brinton, B., & Clarke, D. (2002). Emotion regulation in children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *33*(2), 102–111. https://doi.org/10.1044/0161-1461(2002/008)

Fujiki, M., Brinton, B., McCleave, C. P., Anderson, V. W., & Chamberlaind, J. P. (2013). A social communication intervention to increase validating comments by children with language impairment. *Language, Speech, and Hearing Services in Schools*, *44*(1), 3–19. https://doi.org/10.1044/0161-1461(2012/11-103)

Fujiki, M., Brinton, B., Morgan, M., & Hart, C. H. (1999). Withdrawn and sociable behavior of children with language impairment. *Language, Speech, and Hearing Services in Schools*, *30*(2), 183–195. https://doi.org/10.1044/0161-1461.3002.183

Fujiki, M., Brinton, B., & Todd, C. M. (1996). Social skills of children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *27*(3), 195–201. https://doi.org/10.1044/0161-1461.2703.195

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1363–1382. https://doi.org/10.1044/2014_JSLHR-L-12-0363

Gallon, N., Harris, J., & van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with Grammatical SLI. *Clinical Linguistics & Phonetics*, *21*(6), 435–455. https://doi.org/10.1080/02699200701299982

Gardner, H., Froud, K., McClelland, A., & Van Der Lely, H. K. J. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language and Communication Disorders*, *41*(5), 513–540. https://doi.org/10.1080/13682820500442644

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94. https://doi.org/10.3758/bf03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (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. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966–979. https://doi.org/10.1037/0012-1649.33.6.966

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children’s Test of Nonword Repetition: A Test of Phonological Working Memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Gavarró, A. (2017). A sentence repetition task for Catalan-speaking typically-developing children and children with specific language impairment. *Frontiers in Psychology*, *8*(OCT), 1–13. https://doi.org/10.3389/fpsyg.2017.01865

Gibson, J., Adams, C., Lockton, E., & Green, J. (2013). Social communication disorder outside autism? A diagnostic classification approach to delineating pragmatic language impairment, high functioning autism and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(11), 1186–1197. https://doi.org/10.1111/jcpp.12079

Girbau, D. (2016). The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children. *First Language*, *36*(1), 30–49.

Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language and Communication Disorders*, *42*(1), 59–75.

Gladfelter, A., & Leonard, L. B. (2012). Alternative Tense and Agreement Morpheme Measures for Assessing Grammatical Deficits During the Preschool Period. *Journal of Speech, Language, and Hearing Research*, *56*(2), 542–552. https://doi.org/10.1044/1092-4388(2012/12-0100)

Gonçalves, L., Subtil, A., Oliveira, M. R., & de Zea Bermudez, P. (2014). ROC curve estimation: An overview. *REVSTAT-Statistical Journal*, *12*(1), 1–20. https://doi.org/https://www.ine.pt/revstat/pdf/rs140101.pdf

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gray, S. (2004). Word Learning by Preschoolers With Specific Language Impairment: Predictors and Poor Learners. *Journal of Speech, Language and Hearing Research*, *47*, 1117–1132. https://doi.org/10.1044/1092-4388(2004/083)

Gray, Shelley. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*(2), 129–151. https://doi.org/10.1016/s0021-9924(03)00003-0

Gray, Shelley, & Brinkley, S. (2011). Fast mapping and word learning by preschoolers with specific language impairment in a supported learning context: Effect of encoding cues, phonotactic probability, and object familiarity. *Journal of Speech, Language, and Hearing Research*, *54*(3), 870–884. https://doi.org/10.1044/1092-4388(2010/09-0285)

Gray, Shelley, Brinkley, S., & Svetina, D. (2012). Word Learning by Preschoolers With SLI: Effect of Phonotactic Probability and Object Familiarity. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1289–1300. https://doi.org/10.1044/1092-4388(2012/11-0095)

Gray, Shelley, Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Gray, Shelley, Lancaster, H., Alt, M., Hogan, T. P., Green, S., Levy, R., & Cowan, N. (2020). The structure of word learning in young school-age children. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1446–1466. https://doi.org/10.1044/2020_JSLHR-19-00186

Grinstead, J., Baron, A., Vega-Mendoza, M., De la Mora, J., Cantú-Sánchez, M., & Flores, B.

(2013). Tense Marking and Spontaneous Speech Measures in Spanish Specific Language Impairment: A Discriminant Function Analysis. *Journal of Speech, Language, and Hearing Research*, *56*(1), 352–363. https://doi.org/10.1044/1092-4388(2012/11-0289)

Grüter, T. (2005). Comprehension and production of French object clitics by child second language learners and children with specific language impairment. *Applied Psycholinguistics*, *26*(3), 363–391. https://doi.org/10.1017/S0142716405050216

Guasti, M. T., Palma, S., Genovese, E., Stagi, P., Saladini, G., & Arosio, F. (2016). The production of direct object clitics in pre-school– and primary school–aged children with specific language impairments. *Clinical Linguistics & Phonetics*, *30*(9), 663–678. https://doi.org/10.3109/02699206.2016.1173100

Guiberson, M., & Rodríguez, B. L. (2013). Classification accuracy of nonword repetition when used with preschool-age spanish-speaking children. *Language, Speech, and Hearing Services in Schools*, *44*(2), 121–132.

Guo, L.-Y., Eisenberg, S., Schneider, P., & Spencer, L. (2019). Finite Verb Morphology Composite Between Age 4 and Age 9 for the Edmonton Narrative Norms Instrument: Reference Data and Psychometric Properties. *Language, Speech, and Hearing Services in Schools*, *51*(1), 128–143. https://doi.org/10.1044/2019_LSHSS-19-0028

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English-Speaking Children: Does the Language of Assessment Matter? *Learning Disabilities Research & Practice*, *25*(1), 48–58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Hadley, P. A., & Rice, M. L. (1991). Conversational responsiveness of speech- and language-impaired preschoolers. *Journal of Speech, Language and Hearing Research*, *34*(6), 1308–1317. https://doi.org/10.1044/jshr.3406.1308

Hadley, P. A., & Rice, M. L. (1996). Emergent Uses of BE and DO: Evidence From Children With Specific Language Impairment. *Language Acquisition*, *5*(3), 209–243. https://doi.org/10.1207/S15327817LA0503_2

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with development language disorder ii: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research*, *62*(4), 944–964. https://doi.org/10.1044/2018_JSLHR-L-18-0071

Haeri, N. (2000). Form and Ideology : Arabic Sociolinguistics and Beyond. *Annual Review of Anthropology*, *29*(1), 61–87. https://doi.org/doi.org/10.1146/annurev.anthro.29.1.61

Hamann, C. (2006). Speculations about early syntax: The production of wh-questions by normally developing French children and French children with SLI. *Catalan Journal of Linguistics*, *5*(1), 143–189. https://doi.org/10.5565/rev/catjl.82

Hansson, K., & Leonard, L. B. (2003). The use and productivity of verb morphology in specific language impairment: an examination of Swedish. *Journal of Speech, Language, and Hearing Research*, *44*(384), 351–379. https://doi.org/10.1515/ling.2003.012

Hansson, K., & Nettelbladt, U. (2006). Wh-questions in Swedish children with SLI. *International Journal of Speech-Language Pathology*, *8*(4), 376–383. https://doi.org/doi.org/10.1080/14417040600880722

Hart, K. I., Fujiki, M., Brinton, B., & Hart, C. H. (2004). The Relationship Between Social Behavior and Severity of Language Impairment. *Journal of Speech, Language, and Hearing Research*, *47*(3), 647–662. https://doi.org/10.1044/1092-4388(2004/050)

Henry, L. A., & Botting, N. (2016). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, *33*(1), 19–32. https://doi.org/10.1177/0265659016655378

Hesketh, A., & Conti-Ramsden, G. (2013). Memory and Language in Middle Childhood in Individuals with a History of Specific Language Impairment. *PLoS ONE*, *8*(2), e56314. https://doi.org/10.1371/journal.pone.0056314

Hestvik, A., Schwartz, R. G., & Tornyova, L. (2010). Relative Clause Gap-Filling in Children with Specific Language Impairment. *Journal of Psycholinguistic Research* , *39*(5), 443–456. https://doi.org/10.1007/S10936-010-9151-1

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing Limitations in Children With Specific Language Impairment: The Role of Executive Function. *Child Development*, *77*(6), 1822–1841. https://doi.org/10.1111/j.1467-8624.2006.00976.x

Jackson-Maldonado, D., & Maldonado, R. (2017). Grammaticality differences between Spanish-speaking children with specific language impairment and their typically developing peers. *International Journal of Language and Communication Disorders*, *52*(6), 750–765. https://doi.org/10.1111/1460-6984.12312

Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language and Communication Disorders*, *51*(1), 61–73. https://doi.org/10.1111/1460-6984.12185

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2020). Working, Declarative, and Procedural Memory in Children With Developmental Language Disorder. *Journal of Speech, Language and Hearing Research*, *63*(12), 4162–4178. https://doi.org/10.1044/2020_JSLHR-20-00135

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2021). Word learning and verbal working memory in children with developmental language disorder. *Autism and Developmental Language Impairments*, *6*. https://doi.org/10.1177/23969415211004109

Jakubowicz, C. (2011a). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua*, *121*(3), 339–351. https://doi.org/10.1016/j.lingua.2010.10.006

Jakubowicz, C. (2011b). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua*, *121*(3), 339–351. https://doi.org/10.1016/J.LINGUA.2010.10.006

Janes, H., & Pepe, M. S. (2008a). Matching in Studies of Classification Accuracy: Implications for Analysis, Efficiency, and Assessment of Incremental Value. *Biometrics*, *64*(1), 1–9. https://www.jstor.org/stable/25502015

Janes, H., & Pepe, M. S. (2008b). Adjusting for Covariates in Studies of Diagnostic, Screening, or Prognostic Markers: An Old Concept in a New Setting. *American Journal of Epidemiology*, *168*(1), 89–97. https://doi.org/10.1093/AJE/KWN099

Jarrar, M., Habash, N., Akra, D., & Zalmout, N. (2014). Building a Corpus for Palestinian Arabic: a Preliminary Study. *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Langauge Processing (ANLP)*, 18–27. https://doi.org/10.3115/v1/W14-3603

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., Lam, I., & Wang, M. (1999). Fourteen-Year Follow-Up of Children With and Without Speech/Language Impairments. *Journal of Speech, Language and Hearing Research*, *42*(3), 744–760. https://doi.org/10.1044/JSLHR.4203.744

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and Frequency in Specific Language Impairment: Accuracy and Error Data from Two Nonword Repetition Tests. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1642–1655. https://doi.org/10.1044/1092-4388

Kail, R., & Leonard, L. B. (1986). Word-Finding Abilities in Language-Impaired Children. *ASHA Monographs*, *25*, 5–47. https://www.asha.org/siteassets/publications/monographs25.pdf

Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Birgitta, S. (2014). Nonword Repetition – A Clinical Marker for Specific Language Impairment in Swedish Associated with Parents' Language-Related Problems. *PLoS ONE*, *9*(2), e89544. https://doi.org/10.1371/journal.pone.0089544

Kapalková, S., Polišenská, K., & Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: Novel scoring methods. *International Journal of Language and Communication Disorders*, *48*(1), 78–89. https://doi.org/10.1111/j.1460-6984.2012.00189.x

Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N., van Duuren, M., Pike, G., Grice, S., & Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(7), 1258–1274. https://doi.org/10.1111/j.1469-7610.2004.00322.x

Katsos, N., Roqueta, C. A., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, *119*(1), 43–57. https://doi.org/10.1016/j.cognition.2010.12.004

Kazemi, Y., & Saeednia, S. (2017). The clinical examination of non-word repetition tasks in identifying Persian-speaking children with primary language impairment. *International Journal of Pediatric Otorhinolaryngology*, *93*, 7–12.

Kelso, K., Fletcher, J., & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language and Communication Disorders*, *42*(1), 39–57. https://doi.org/10.1080/13682820600693013

Khater, M. (2016). *The relationship between nonword repetition, root and pattern effects, and vocabulary in Gulf Arabic speaking children.(Unpublished thesis)*. City, University of London.

Khoja, M. A. (2017). A survey of formal and informal assessment procedures used by speech-language pathologists in Saudi Arabia. *Speech, Language and Hearing*, *22*(2), 91–99. https://doi.org/10.1080/2050571X.2017.1407620

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860–897. https://doi.org/10.1080/01690960601155284

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. https://doi.org/10.1111/desc.12202

Komeili, M., & Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical Linguistics & Phonetics*, *27*(2), 152–162. https://doi.org/10.3109/02699206.2012.751625

Krok, W. C., & Leonard, L. B. (2015). Past Tense Production in Children With and Without Specific Language Impairment Across Germanic Languages: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *58*, 1326–1340. https://doi.org/10.1044/2015_JSLHR-L-14-0348

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(5), e1528. https://doi.org/10.1002/wcs.1528

Lalami, L. (1996). Clitic left dislocation in Moroccan Arabic. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, *4*, 115–130.

Law, J., Rush, R., Parsons, S., & Schoon, I. (2013). The relationship between gender, receptive vocabulary, and literacy from school entry through to adulthood. *International Journal of Speech-Language Pathology* , *15*(4), 407–415. https://doi.org/10.3109/17549507.2012.721897

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling Developmental Language Difficulties From School Entry Into Adulthood: Literacy, Mental Health, and Employment Outcomes. *Journal of Speech, Language and Hearing Research*, *52*(6), 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Leclercq, A. L., Maillart, C., & Majerus, S. (2013). Nonword repetition problems in children with specific language impairment: A deficit in accessing long-term linguistic representations? *Topics in Language Disorders*, *33*(3), 238–254. https://doi.org/10.1097/TLD.0b013e31829dd8c9

Leclercq, A. L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities*, *35*(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Leech, G. (2006). *A Glossary of English Grammar*. Edinburgh: Edinburgh University Press.

Leitãto, S., Hogben, J., & Fletcher, J. (1997). Phonological processing skills in speech and language impaired children. *International Journal of Language and Communication Disorders*, *32*(2s), 91–111. https://doi.org/10.1111/j.1460-6984.1997.tb01626.x

Lenth, R. (2020). *Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.6*. https://cran.r-project.org/package=emmeans

Leonard, L. B. (1981). An invited article Facilitating linguistic skills in children with specific language impairment. *Applied Psycholinguistics*, *2*(2), 89–118. https://doi.org/10.1017/S0142716400000886

Leonard, L. B. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, *10*(2), 179–202. https://doi.org/10.1017/S0142716400008511

Leonard, L. B. (2000). Children with Specific Language Impairment. In *Children with Specific Language Impairment* (2nd ed.). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Leonard, L. B. (2014a). Children with Specific Language Impairment. In *Children with Specific Language Impairment* (2nd Editio). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Leonard, L. B. (2014b). Specific Language Impairment Across Languages. *Child Development Perspectives*, *8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Caselli, M. C., Bortolini, U., Karla, K., & Sabbadini, L. (1992). Morphological Deficits in Children With Specific Language Impairment : The Status of Features in the Underlying Grammar. *Journal of Speech, Language and Hearing Research*, *2*(2), 151–179. https://doi.org/10.1207/s15327817la0202_2

Leonard, L. B., Caselli, M. C., Bortolini, U., Sabbadini, L., & McGregor, K. K. (1992). Morphological Deficits in Children With Specific Language Impairment: The Status of Features in the Underlying Grammar. *Journal of Speech, Language and Hearing Research*, *2*(2), 151–179. https://doi.org/10.1207/s15327817la0202_2

Leonard, L. B., Deevy, P., Miller, C. A., Rauf, L., Charest, M., & Robert, K. (2003). Surface Forms and Grammatical Functions : Past Tense and Passive Participle Use by Children with

Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *46*(1), 43–55. https://doi.org/10.1044/1092-4388(2003/004)

Leonard, L. B., & Dispaldro, M. (2013). *The Effects of Production Demands on Grammatical Weaknesses in Specific Language Impairment: The Case of Clitic Pronouns in Italian*. *56*(4), 1272–1286. https://doi.org/10.1044/1092-4388(2012/12-0295)

Leonard, L. B., & Dromi, E. (1994). The use of Hebrew verb morphology by children with specific language impairment and children developing language normally. *First Language*, *14*(42–43), 283–304. https://doi.org/10.1177/014272379401404219

Leonard, L. B., Dromi, E., Adam, G., & Zadunaisky-Ehrlich, S. (2000). Tense and finiteness in the speech of children with specific language impairment acquiring Hebrew. *International Journal of Language and Communication Disorders*, *35*(3), 319–335. https://doi.org/10.1080/136828200410609

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997a). Three Accounts of the Grammatical Morpheme Difficulties of English-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *40*(4), 741–753. https://doi.org/10.1044/jslhr.4004.741

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997b). Three accounts of the grammatical morpheme difficulties of english speaking children specific language impairment. *Journal of Speech, Language, and Hearing Research*, *40*(4), 741–753. https://doi.org/10.1044/JSLHR.4004.741

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J. B., & Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder i: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research*, *62*(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Leonard, L. B., & Kueser, J. B. (2019). Five overarching factors central to grammatical learning and treatment in children with developmental language disorder. *International Journal of Language and Communication Disorders*, *54*(3), 347–361. https://doi.org/10.1111/1460-6984.12456

Leonard, L. B., McGregor, K. K., & Allen, G. D. (1992). Grammatical morphology and speech perception in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *35*(5), 1076–1085. https://doi.org/10.1044/JSHR.3505.1076

Leonard, L. B., Miller, C. A., & Owen, A. J. (2000). The comprehension of verb agreement morphology by English-speaking children with specific language impairment. *Clinical Linguistics and Phonetics*, *14*(6), 465–481. https://doi.org/10.1080/026992000415886

Leonard, L. B., Nippold, M. A., Kail, R., & Hale, C. A. (1983). Picture naming in language-impaired children. *Journal of Speech, Language and Hearing Research*, *26*(4), 609–615. https://doi.org/10.1044/jshr.2604.609

Leonard, L. B., Sabbadini, L., Leonard, J. S., & Volterra, V. (1987). Specific language impairment in children: A cross-linguistic study. *Brain and Language*, *32*(2), 233–252. https://doi.org/10.1016/0093-934X(87)90126-X

Leonard, L. B., Wong, A. M.-Y., Deevy, P., Stokes, S. F., & Fletcher, P. (2006). The production of passives by children with specific language impairment: Acquiring English or Cantonese. *Applied Psycholinguistics*, *27*(2), 267–299. https://doi.org/10.1017/S0142716406060280

Leonard, Laurence B. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, *10*(2), 179–202. https://doi.org/10.1017/S0142716400008511

Leonard, Laurence B., & Dispaldro, M. (2013). *The Effects of Production Demands on Grammatical Weaknesses in Specific Language Impairment: The Case of Clitic Pronouns in Italian*. *56*(4), 1272–1286. https://doi.org/10.1044/1092-4388(2012/12-0295)

Liiva, C. A., & Cleave, P. L. (2005). Roles of initiation and responsiveness in access and participation for children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(4), 868–883. https://doi.org/10.1044/1092-4388(2005/060)

Lindsay, K., & Dockrell, J. E. (2012). Longitudinal patterns of behavioral, emotional, and social difficulties and self-concepts in adolescents with a history of specific language impairment. *Language, Speech and Hearing Services Il Schools*, *43*(4), 445–460. https://doi.org/10.1044/0161-1461(2012/11-0069)

Lloyd-Esenkaya, V., Russell, A. J., & St Clair, M. C. (2020). What are the peer interaction strengths and difficulties in children with developmental language disorder? A systematic review. *International Journal of Environmental Research and Public Health*, *17*(9), 3140. https://doi.org/10.3390/ijerph17093140

Lombardi, L., & Potter, M. C. (1992). The generation of syntax in short term memory. *Memory and Languagery and Language*, *31*, 713–733. https://doi.org/10.1016/0749-596X(92)90036-W

Lucas, R., & Norbury, C. F. (2015). Making inferences from text: It's vocabulary that matters. *Journal of Speech, Language, and Hearing Research*, *58*(4), 1224–1232. https://doi.org/10.1044/2015_JSLHR-L-14-0330

Lukács, Á., Leonard, L. B., Kas, B., & Pléh, C. (2009). The Use of Tense and Agreement by Hungarian-Speaking Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*(1), 98–117. https://doi.org/10.1044/1092-4388(2008/07-0183)

Łuniewska, M., Wodniecka, Z., Miller, C. A., Smolík, F., Butcher, M., Chondrogianni, V., Hreich, E. K., Messarra, C., Razak, R. A., Treffers-Daller, J., Yap, N. T., Abboud, L., Talebi, A., Gureghian, M., Tuller, L., & Haman, E. (2019). Age of acquisition of 299 words in seven languages: American English, Czech, Gaelic, Lebanese Arabic, Malay, Persian and Western Armenian. In *PLoS ONE* (Vol. 14, Issue 8, p. e0220611). Public Library of Science. https://doi.org/10.1371/journal.pone.0220611

Maddieson, I. (2006). Correlating phonological complexity: Data and validation. Linguistic Typology, 10(1), 106–123. https://doi.org/10.1515/LINGTY.2006.017.

Maddieson, I., Sébastien, F., Egidio, M., & François, P. (2011). LAPSyD: Lyon-Albuquerque Phonological Systems Databases (Version 1.0). https://lapsyd.huma-num.fr/index.php

Maggio, V., Grañana, N. E., Richaudeau, A., Torres, S., Giannotti, A., & Suburo, A. M. (2014). Behavior Problems in Children With Specific Language Impairment: *Journal of Child Neurology*, *29*(2), 194–202. https://doi.org/10.1177/0883073813509886

Maillart, C., & Parisse, C. (2006). Phonological deficits in French speaking children with SLI. *International Journal of Language and Communication Disorders*, *41*(3), 253–274.

Mainela-Arnold, E., & Evans, J. L. (2005). Beyond Capacity Limitations: Determinants of Word Recall Performance on Verbal Working Memory Span Tasks in Children With SLI. *Journal of Speech, Language, and Hearing Research*, *48*, 897–909. https://doi.org/10.1044/1092-4388(2005/062

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language*, *21*(2), 339–366. https://doi.org/10.1017/s0305000900009302

Marini, A., Gentili, C., Molteni, M., & Fabbro, F. (2014). Differential verbal working memory effects on linguistic production in children with Specific Language Impairment. *Research in Developmental Disabilities*, *35*(12), 3534–3542.

https://doi.org/10.1016/J.RIDD.2014.08.031

Marinis, T. (2011). On the nature and cause of Specific Language Impairment: A view from sentence processing and infant research. *Lingua*, *121*(3), 463–475. https://doi.org/10.1016/J.LINGUA.2010.10.010

Marinis, T., & Armon-Lotem, S. (2015). Sentence Repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from Language impairment* (pp. 95–124). Bristol:Multilingual Matters.

Marinis, T., & Saddy, D. (2013). Parsing the passive: Comparing children with specific language impairment to sequential bilingual children. *Language Acquisition*, *20*(2), 155–179. https://doi.org/10.1080/10489223.2013.766743

Marinis, T., & Van Der Lely, H. K. J. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *International Journal of Language and Communication Disorders*, *42*(5), 557–582. https://doi.org/10.1080/13682820601058190

Marshall, C. (2006). The Morpho-Phonological Interface in Specific Language Impairment. *Language Acquisition*, *13*(4), 373–375. https://doi.org/10.1207/S15327817LA1304_7

Marshall, C. C. R., & van der Lely, H. K. J. (2006). A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition*, *100*(2), 302–320. https://doi.org/10.1016/J.COGNITION.2005.06.001

Marshall, C. R., & van der Lely, H. K. J. (2007). The impact of phonological complexity on past tense inflection in children with Grammatical-SLI. *Advances in Speech Language Pathology*, *9*(3), 191–203. https://doi.org/10.1080/14417040701261509

Mastropavlou, M., & Tsimpli, I. M. (2011). Complementizers and subordination in typical language acquisition and SLI. *Lingua*, *121*(3), 442–462. https://doi.org/10.1016/j.lingua.2010.10.009

Mayer, M. (1969). *Frog, where are you?* New York, New York: Dial books for Young Readers.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “Specifics” of Specific Reading Disability and Specific Language Impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(7), 869–874. https://doi.org/10.1111/1469-7610.00674

McCarthy, J. J., & Prince, A. (1988). -Quantitative Transfer in Reduplicative Templatic Morphology and John McCarthy and Alan Prince. *Linguistics in the Morning Calm*, *2*(99).

McGregor, K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism and Developmental Language Impairments*, *5*, 1–19. https://doi.org/10.1177/2396941519899311

McGregor, K. K., Goffman, L., Van Horne, A. O., Hogan, T. P., & Finestack, L. H. (2020). Developmental Language Disorder: Applications for Advocacy, Research, and Clinical Service. *Perspectives of the ASHA Special Interest Groups*, *5*(1), 38–46. https://doi.org/10.1044/2019_persp-19-00083

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *45*(5), 998–1014. https://doi.org/10.1044/1092-4388(2002/081)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders*, *48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

McGregor, K. K., & Waxman, S. R. (1998). Object naming at multiple hierarchical levels: A comparison of preschoolers with and without word-finding deficits. *Journal of Child Language*, *25*(2), 419–430. https://doi.org/10.1017/S030500099800347X

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.

Menyuk, P. (1993). Children with Specific Language Impairment, Developmental Dysphasia: Linguistics aspects. In G. Blanken, H. Dittmann, H. Grimm, J. Mrashall, & C. W. Wallesch (Eds.), *Linguistics disorders and pathologies* (pp. 606–625). Berlin: Walter de Gruyter.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*(January), 104092.

https://doi.org/10.1016/j.jml.2020.104092

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. In *Journal of Educational Psychology* (Vol. 91, Issue 1, pp. 3–19).

Mirak, J., & Rescorla, L. (1998). Phonetic Skills and Vocabulary Size in Late Talkers: Concurrent and Predictive Relationships. *Applied Psycholinguistics*, *19*, 1–17. https://doi.org/10.1017/S0142716400010559

Moawad, R. A. (2006). *The acqusition of the Arabic gender and number systems. (Unpublished Thesis)*. University of Wales.

Moll, K., Hulme, C., Nag, S., & Snowling, M. J. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics*, *36*(2), 203–221. https://doi.org/10.1017/S0142716413000209

Montgomery, J. W. (1995). Sentence Comprehension in Children With Specific Language Impairment: The Role of Phonological Working Memory. *Journal of Speech, Language and Hearing Research*, *38*(1), 187–199. https://doi.org/10.1044/JSHR.3801.187

Montgomery, J. W. (2000). Verbal Working Memory and Sentence Comprehension in Children With Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, *43*(2), 293–308. https://doi.org/10.1044/JSLHR.4302.293

Montgomery, J. W. (2002). Understanding the Language Difficulties of Children With Specific Language Impairments. *American Journal of Speech-Language Pathology*, *11*(1), 77–91. https://doi.org/10.1044/1058-0360(2002/009)

Montgomery, J. W., & Evans, J. L. (2009). Complex Sentence Comprehension and Working Memory in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 269–288. https://doi.org/10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2018). Memory and language in children with SLI. In T. P. Alloway (Ed.), *Working Memory and Clinical Developmental Disorders* (1st Editio, pp. 22–37). London: Routledge.

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2016). *Syntactic Versus Memory Accounts of the Sentence Comprehension Deficits of Specific Language Impairment: Looking Back, Looking Ahead*. *59*(6), 1491–1504. https://doi.org/10.1044/2016_JSLHR-L-15-0325

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). *Working Memory and Specific*

Language Impairment: An Update on the Relation and Perspectives on Assessment and Treatment. *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Morsi, R. (2009). *Specific language impairment in Egyptian Arabic: A preliminary investingation. (Unpublished Thesis)*. University of Reading.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The Influence of Vocabulary Size, Phonotactic Probability, and Wordlikeness on Nonword Repetitions of Children With and Without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, *48*(2), 439–458. https://doi.org/10.1044/1092-4388(2005/030)

Newcomer, P., & Hammill, D. (2008). *Test of Language Development- Primary (4th ed.)*. Austin, TX: PRO-ED.

Newton, E. J., Roberts, M. J., & Donlan, C. (2010). Deductive reasoning in children with specific language impairment. *British Journal of Developmental Psychology*, *28*(1), 71–87. https://doi.org/10.1348/026151009X480185

Nicola, K., & Watter, P. (2015). Health-related quality of life from the perspective of children with severe specific language impairment. *Health and Quality of Life Outcomes 2015 13:1*, *13*(1), 1–9. https://doi.org/10.1186/S12955-015-0326-1

Nitido, H., & Plante, E. (2020). Diagnosis of developmental language disorder in research studies. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2777–2788. https://doi.org/10.1044/2020_JSLHR-20-00091

Norbury, C. F. (2005a). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, *90*(2), 142–171. https://doi.org/10.1016/j.jecp.2004.11.003

Norbury, C. F. (2005b). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383–399. https://doi.org/10.1348/026151005X26732

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language*, *41*(3), 485–510. https://doi.org/10.1017/S030500091300007X

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles,

A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist - 2. *International Journal of Language and Communication Disorders*, *39*(3), 345–364. https://doi.org/10.1080/13682820410001654883

Novogrodsky, R., & Friedmann, N. (2006). The production of relative clauses in syntactic SLI: A window to the nature of the impairment. *International Journal of Speech-Language Pathology*, *8*(4), 364–375. https://doi.org/10.1080/14417040600919496

Oetting, J. B., & Rice, M. L. (1993). Plural acquisition in children with specific language impairment. *Journal of Speech and Hearing Research*, *36*(6), 1236–1248. https://doi.org/10.1044/jshr.3606.1236

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, *38*(2), 434–445. https://doi.org/10.1044/jshr.3802.434

Omar, M. (1973). *The acquisition of Egyptian Arabic as a native language* (Janua ling). The Hague, Paris: Mouton & Co N.V. https://doi.org/10.1515/9783110819335

Ouali, H. (2018). The Syntax of Tense in Arabic. In Elabbas Benmamoun & R. Bassiouney (Eds.), *The Routledge Handbook of Arabic Linguistics* (pp. 89–103). New Yor, NY: Routledge. https://doi.org/10.4324/9781315147062-6

Owen, A. J., Dromi, E., & Leonard, L. B. (2001). The phonology-morphology interface in the speech of Hebrew-speaking children with specific language impairment. *Journal of Communication Disorders*, *34*(4), 323–337. https://doi.org/10.1016/S0021-9924(01)00053-3

Owen, A. J., & Leonard, L. B. (2006a). The production of finite and nonfinite complement clauses by children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research*, *49*(3), 548–571. https://doi.org/10.1044/10902-4388(2006/040)

Owen, A. J., & Leonard, L. B. (2006b). The production of finite and nonfinite complement clauses by children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research*, *49*(3), 548–571. https://doi.org/10.1044/10902-4388(2006/040)

Paradis, J., & Crago, M. (2001). The Morphosyntax of Specific Language Impairment in French: An Extended Optional Default Account. *Language Acquisition*, *9*(4), 269–300. https://doi.org/10.1207/S15327817LA0904_01

Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech and Hearing Research*, *35*(1), 99–107. https://doi.org/10.1044/jshr.3501.99

Pauls, L. J., & Archibald, L. M. D. (2016). Executive Functions in Children With Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language and Hearing Research*, *59*(5), 1074–1086. https://doi.org/10.1044/2016_JSLHR-L-15-0174

Pawłowska, M. (2014). Evaluation of three proposed markers for language impairment in English: A Meta- Analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research*, *57*, 2261–2273. https://doi.org/10.1044/2014_JSLHR-L-13-0189.

Pham, G., & Ebert, K. D. (2020). Diagnostic accuracy of sentence repetition and nonword repetition for developmental language disorder in vietnamese. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1521–1536. https://doi.org/10.1044/2020_JSLHR-19-00366

Pharr, A. E., Ratner, N. B., & Rescorla, L. (2000). Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics*, *21*, 429–449. https://doi.org/10.1017/S014271640000401X

Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., & Morgan, A. T. (2019). What predicts nonword repetition performance? *Child Neuropsychology*, *26*(4), 518–533. https://doi.org/10.1080/09297049.2019.1674799

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders*, *50*(1), 106–118. https://doi.org/10.1111/1460-6984.12126

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research*, *53*(2), 414–

429. https://doi.org/10.1044/1092-4388(2009/08-0016)

Poll, G. H., Miller, C. A., & Hell, J. G. van. (2016). Sentence Repetition Accuracy in Adults With Developmental Language Impairment: Interactions of Participant Capacities and Sentence Structures. *Journal of Speech, Language, and Hearing Research*, *59*(2), 302–316. https://doi.org/10.1044/2015_JSLHR-L-15-0020

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language & Communication Disorders*, *48*(3), 329–342. https://doi.org/10.1111/1460-6984.12014

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*(3), 265–282. https://doi.org/10.1006/jmla.1997.2546

Prévost, P., Strik, N., & Tuller, L. (2014). Wh- questions in child L2 French: Derivational complexity and its interactions with L1 properties, length of exposure, age of exposure, and the input. *Second Language Research*, *30*(2), 225–250. https://doi.org/10.1177/0267658313519814

Qasem, F., & Sircar, S. (2017). Imperative as root infinitive analogue in Yemeni Ibbi Arabic: Two case studies. *Arab Journal of Applied Linguistics*, *2*(1), 2490–4198.

Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, *10*(1), 37–52. https://doi.org/10.1080/21622965.2019.1596113

Rakhlin, N. V., Li, N., Aljughaiman, A., & Grigorenko, E. L. (2020). Narrative language markers of Arabic language development and impairment. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3472–3487. https://doi.org/10.1044/2020_JSLHR-20-00082

Ramachandra, V., Hewitt, L. E., & Brackenbury, T. (2011). The Relationship Between Phonological Memory, Phonological Sensitivity, and Incidental Word Learning. *Journal of Psycholinguistic Research*, *40*, 93–109. https://doi.org/10.1007/s10936-010-9157-8

Ramus, F., Marshall, C. R., Rosen, S., & Van Der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, *136*(2), 630–645. https://doi.org/10.1093/brain/aws356

Rapin, I., & Allen, D. A. (1983). Developmental Language Disorders: Nosologic Considerations. In U. Kirk (Ed.), *Neuropsychology of Language, Reading and Spelling* (2nd ed., pp. 155–184). Cambridge,MA: Academic Press. https://doi.org/10.1016/B978-0-12-409680-6.50014-7

Rapin, I., & Allen, D. A. (1988). Syndromes in developmental dysphasia and adult aphasia. *Research Publications - Association for Research in Nervous and Mental Disease*, *66*, 57–75. https://europepmc.org/article/med/2451853

Raven, J. (2007). Coloured progressive matrices. London: Pearson.

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, *19*(56), 187–206. https://doi.org/10.1177/014272379901905603

Redmond, S. M. (2011). Peer victimization among students with specific language impairment, attention-deficit/hyperactivity disorder, and typical development. *Language, Speech, and Hearing Services in Schools*, *42*(4), 520–535. https://doi.org/10.1044/0161-1461(2011/10-0078)

Redmond, S. M., Ash, A. C., Christopulos, T. T., & Pfaff, T. (2019). Diagnostic accuracy of sentence recall and past tense measures for identifying children's language impairments. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2438–2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388

Redmond, S. M., & Rice, M. L. (2001). Detection of Irregular Verb Violations by Children With and Without SLI. *Journal of Speech, Language, and Hearing Research*, *44*(3), 655–669. https://doi.org/10.1044/1092-4388(2001/053)

Redmond, S. M., Thompson, H. L., & Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivitydisorder. *Journal of Speech Language & Hearing Research*, *54*, 99–117.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language and Communication Disorders*, *49*(4), 416–451. https://doi.org/10.1111/1460-6984.12102

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, *39*(1), 153–165. https://doi.org/10.1044/jshr.3901.153

Rice, M. L., & Blossom, M. (2013). What do children with Specific Language Impairment do with multiple forms of DO? *Journal of Speech, Language, and Hearing Research : JSLHR*, *56*(1), 222. https://doi.org/10.1044/1092-4388(2012/11-0107)

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders*, *55*(1), 33–42. https://doi.org/10.1044/jshd.5501.33

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific Language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, *58*(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of Omitted BE and DO in Questions as Extended Finiteness Clinical Markers of Specific Language Impairment (SLI) to 15 Years: A Study of Growth and Asymptote. *Journal of Speech, Language and Hearing Research*, *52*(6), 1417–1433. https://doi.org/10.1044/1092-4388(2009/08-0171)

Rice, M. L., Noll, K. R., & Grimm, H. (1997). An Extended Optional Infinitive Stage in German-Speaking Children With Specific Language Impairment. *Language Acquisition*, *6*(4), 255–295. https://doi.org/10.1207/S15327817LA0604_1

Rice, M. L., Sell, M. A., & Hadley, P. A. (1991). Social Interactions of Speech, and LanguageImpaired Children. *Journal of Speech, Language and Hearing Research*, *34*, 1299–1307. https://doi.org/10.1044/jshr.3406.1299

Rice, M. L., Smolik, F., Rytting, N., & Blossom, M. (2010). Mean Length of Utterance Levels in 6-month Intervals for Children 3 to 9 Years with and without Language Impairments. *Hearing Research*, *53*(April), 333–349.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific Language Impairment as a Period of Extended Optional Infinitive. *Journal of Speech, Language, and Hearing Research*, *38*(4), 850–863. https://doi.org/doi.org/10.1044/jshr.3804.850

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1412–1431.

https://doi.org/10.1044/jslhr.4106.1412

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality Judgments of an Extended Optional Infinitive Grammar. *Journal of Speech, Language and Hearing Research*, *42*(4), 943–961. https://doi.org/10.1044/JSLHR.4204.943

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *International Journal of Language & Communication Disorders*, *47*(5), 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Riches, N. G. (2017). Complex sentence profiles in children with Specific Language Impairment: Are they really atypical? *Journal of Child Language*, *44*(2), 269–296. https://doi.org/10.1017/S0305000915000847

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language and Communication Disorders*, *45*(1), 47–60. https://doi.org/10.3109/13682820802647676

Rinaldi, W. (2000). Pragmatic comprehension in secondary school-aged students with specific developmental language disorder. *International Journal of Language and Communication Disorders*, *35*(1), 1–29. https://doi.org/10.1080/136828200247223

Rispens, J., & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *British Journal of Developmental Psychology*, *28*(1), 177–188. https://doi.org/10.1348/026151009x482633

Roberts, J., Rescorla, L., Giroux, J., & Stevens, L. (1998). Phonological skills of children with specific expressive language impairment (SLI-E): Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, *41*(2), 374–384. https://doi.org/doi.org/10.1044/jslhr.4102.374

Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics*, *31*(1), 141–165. https://doi.org/10.1017/S0142716409990208

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). *pROC: an open-source package for R and S+ to analyze and compare ROC curves*. BMC *Bioinf*, 1471–2105. https://doi.org/10.1186/1471-2105-12-77

Rothweiler, M., Chilla, S., & Clahsen, H. (2012). Subject-verb agreement in Specific Language

Impairment: A study of monolingual and bilingual German-speaking children. *Bilingualism*, *15*(1), 39–57. https://doi.org/10.1017/S136672891100037X

Roy, P., & Chiat, S. (2004). A Prosodically Controlled Word and Nonword Repetition Task for 2- to 4-Year-Olds: Evidence from Typically Developing Children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223–234.

Rujas, I., Mariscal, S., Murillo, E., & Lázaro, M. (2021). Sentence Repetition Tasks to Detect and Prevent Language Difficulties: A Scoping Review. *Children*, *8*(7), 598. https://doi.org/10.3390/CHILDREN8070578

Ryding, K. C. (2005). Verb inflection: a summary. In *A reference grammar of modern standard Arabic* (1st Editio, pp. 438–446). Cambridge: Cambridge university press. https://doi.org/10.1017/CBO9780511486975. 022

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little, Brown and Co.

Saiegh-Haddad, E., & Ghawi-Dakwar, O. (2017). Impact of diglossia on word and non-word repetition among language impaired and typically developing Arabic native speaking children. *Frontiers in Psychology*, *8*(NOV), 1–17. https://doi.org/10.3389/fpsyg.2017.02010

Sandgren, O., Salameh, E. K., Nettelbladt, U., Dahlgren-Sandberg, A., & Andersson, K. (2020). Using a word association task to investigate semantic depth in swedish-speaking children with developmental language disorder. *Logopedics Phoniatrics Vocology*, 1–7. https://doi.org/10.1080/14015439.2020.1785001

Schütze, C. T., & Wexler, K. (1996). Subject Case Licensing and English Root Infinitives. In A. Stringfellow, D. Cahma-Amitay, E. Hughes, & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development* (pp. 670–681). Somerville, MA: Cascadilla Press. https://linguistics.ucla.edu/people/cschutze/BU20.pdf

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language & Hearing Research*, *64*(7), 2750–2765. https://doi.org/10.1044/2021_JSLHR-20-00552

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties.

International Journal of Language & Communication Disorders, *45*(6), 691–702. https://doi.org/10.3109/13682820903509432

Seeff-Gabriel, B. K., Chiat, S., & Roy, P. (2008). *Early repetition battery*. London: Pearson Assessment.

Shaalan, S. (2010). *Investigating grammatical complexity in Gulf Arabic speaking children with specific language impairment (SLI).(Unpublished Thesis)*. University College London.

Shaalan, S. (2020). Nonword Repetition Skills in Gulf Arabic–Speaking Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3700–3713. https://doi.org/10.1044/2020_JSLHR-20-00040

Shaalan, S., & Khater, M. (2006). A comparison of two measures of assessing spontaneous language samples in Arabic speaking children. *[Poster Presentation]. Child Language Seminar 2006, University of Newcastle upon-Tyne, United Kingdom*.

Shahin, K. (2010). Fostering language development in Arabic. In *Encyclopedia of language and literacy development* (pp. 1–9). Canadian Language and Literacy Research Network.

Shahin, Kimary. (2007). Palestinian Arabic. In M. Eid, A. Elgibali, K. Versteegh, M. Woldich, & A. Zaborski (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. 3, pp. 526–538).

Shahmahmood, T. M., Soleymani, Z., Kazemi, Y., Haresabadi, F., Eghbal, N., Kazemi, H., & Amin, S. (2020). The Effects of Phonological Short Term Memory on Lexical and Grammatical Production Skills in Persian Children with Developmental Language Disorder. *Journal of Psycholinguistic Research*, *49*(6), 1067–1082. https://doi.org/10.1007/S10936-020-09734-3

Slobin, D., & Welsh, C. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, & Winston. https://eric.ed.gov/?id=ED012892

Smolík, F., & Vávrů, P. (2014). Sentence imitation as a marker of SLI in Czech: Disproportionate impairment of verbs and clitics. *Journal of Speech, Language, and Hearing Research*, *57*(3), 837–849. https://doi.org/10.1044/2014_JSLHR-L-12-0384

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(3), 369–373. https://doi.org/10.1017/S0142716400009279

Snowling, M. J., Adams, J. W., Bishop, D. V. M., & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International*

Journal of Language & Communication Disorders, *36*(2), 173–183. https://doi.org/10.1080/13682820120976

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., & Hulme, C. (2020). Dyslexia and Developmental Language Disorder: comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*, *61*(6), 672–680. https://doi.org/10.1111/JCPP.13140

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., & Hulme, C. (2019). Developmental Outcomes for Children at High Risk of Dyslexia and Children With Developmental Language Disorder. *Child Development*, *90*(5), e548–e564. https://doi.org/10.1111/CDEV.13216

Soltan, U. (2007). On agree and postcyclic merge in syntactic derivations: First conjunct agreement in standard Arabic. *Perspectives on Arabic Linguistics, 19*, 191–213. https://doi.org/10.1075/ cilt.289.15sol

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility Criteria for Language Impairment. *Language, Speech, and Hearing Services in Schools*, *37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders*, *44*(2), 186–199. https://doi.org/10.1016/j.jcomdis.2010.09.004

Stanford, E., Durrleman, S., & Delage, H. (2019). The Effect of Working Memory Training on a Clinical Marker of French-Speaking Children With Developmental Language Disorder. *American Journal of Speech-Language Pathology*, *28*(4), 1388–1410. https://doi.org/10.1044/2019_AJSLP-18-0238

Stavrakaki, S. (2006). Developmental perspectives on Specific Language Impairment: Evidence from the production of wh-questions by Greek SLI children over time. *International Journal of Speech-Language Pathology*, *8*(4), 384–396. https://doi.org/10.1080/14417040600880714

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of cantonese. *Journal of Speech, Language, and Hearing Research*, *49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Storkel, H. L., Komesidou, R., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying adequate progress and successful learning patterns. *Language, Speech, and Hearing Services in Schools*, *48*(2), 108–124. https://doi.org/10.1044/2017_LSHSS-16-0058

Stromswold, K. (1995). The Acquisition of Subject and Object Wh-Questions. *Language Acquisition*, *4*(1–2), 5–48. https://doi.org/doi.org/10.1080/10489223.1995.9671658

Surian, L., Baron-Cohen, S., & Van Der Lely, H. (1996). Are children with autism deaf to gricean maxims? *Cognitive Neuropsychiatry*, *1*(1), 55–72. https://doi.org/10.1080/135468096396703

Swanson, L. A., Fey, M. E., Mills, C. E., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology*, *14*(2), 131–143. https://doi.org/10.1044/1058-0360(2005/014)

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36.

Tabachnick, B., & Fidell, L. (2007). *Using Multivariate Statistics* (5th ed.). Allyn & Bacon/Pearson Education

Taha, J., & Chondrogianni, V. (2017). *Nonword repetition as a risk marker in Palestinian children with and without delayed language development.(Unpublished dissertation)*. The University of Edinburgh

Taha, J. (2019). Arabic noun plurals test (Unpublished Material).

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021a). Sentence repetition as a clinical marker of Developmental Language Disorder: Evidence from Arabic. *Journal of Speech, Language, and Hearing Research*, 64(12),4876-4899. https://doi.org/10.1044/2021_JSLHR-21-00244

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021b). Expressive verb morphology deficits in Arabic-speaking children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 64(2), 561-578. https://doi.org/10.1044/2020_JSLHR-19-00292

Tallal, P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia . *Nature*, *241*(5390), 468–469. https://doi.org/10.1038/241468A0

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, *12*(1), 83–93. https://doi.org/10.1016/0028-3932(74)90030-X

Team R Core. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Theodorou, E., Kambanaros, M., & Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilectal children with SLI. *Frontiers in Psychology*, *8*(DEC), 2104. https://doi.org/10.3389/fpsyg.2017.02104

Thomas, M., Ansari, D., Jarrold, C., & Karmiloff-Smith, A. (2009). *Using Developmental Trajectories to Understand Developmental Disorders*. *52*(April), 336–359.

Thomas, M. S., Purser, H. R., & Van Herwegen, J. (2012). Cognition: The developmental trajectory approach. In A. Farran, E. K., & Karmiloff-Smith (Ed.), *Neurodevelopmental disorders across the lifespan: A neuroconstructivist approach* (pp. 13–36). Oxford University Press.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16. https://doi.org/10.1016/J.JCOMDIS.2012.08.002

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Thordardottir, E. T., & Namazi, M. (2007). Specific Language Impairment in French-Speaking Children: Beyond Grammatical Morphology. *Journal of Speech, Language, and Hearing Research*, *50*(3), 698–715. https://doi.org/10.1044/1092-4388(2007/049)

Timler, G. R. (2008). Social knowledge in children with language impairments: Examination of strategies, predicted consequences, and goals in peer conflict situations. *Clinical Linguistics and Phonetics*, *22*(9), 741–763. https://doi.org/10.1080/02699200802212470

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245–1260.

https://doi.org/10.1044/jslhr.4006.1245

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 and Phonetics*, *28*(7–8), 602–616. https://doi.org/10.3109/02699206.2014.927003

Toseeb, U., & St Clair, M. C. (2020). Trajectories of prosociality from early to middle childhood in children at risk of Developmental Language Disorder. *Journal of Communication Disorders*, *85*(May-June), 105984. https://doi.org/10.1016/j.jcomdis.2020.105984

Tsimpli, I. M. (2001). LF-Interpretability and Language Development: A Study of Verbal and Nominal Features in Greek Normally Developing and SLI Children. *Brain and Language*, *77*(3), 432–448. https://doi.org/10.1006/BRLN.2000.2413

Tucker, M. A. (2011). The morphosyntax of the Arabic verb: Toward a unified syntax-prosody. In N. LaCara, A. Thompson, & M. A. Tucker (Eds.), *Morphology at Santa Cruz: Papers in Honor of Jorge Hankamer* (pp. 177–211). Linguistics Research Center.

Tuller, L., Delage, H., Monjauze, C., Piller, A. G., & Barthez, M. A. (2011). Clitic pronoun production as a measure of atypical language development in French. *Lingua*, *121*(3), 423–441. https://doi.org/10.1016/j.lingua.2010.10.008

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language and Communication Disorders*, *53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

Tyack, D. L., & Gottsleben, R. H. (1986). Acquisition of complex sentences. *Language, Speech, and Hearing Services in Schools*, *17*(3), 160–174. https://doi.org/10.1044/0161-1461.1703.160

Ullman, M., & Pierpont, E. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Umemneku Chikere, C. M., Wilson, K. J., Allen, A. J., & Vale, L. (2021). Comparative diagnostic accuracy studies with an imperfect reference standard – a comparison of correction methods. *BMC Medical Research Methodology 2021 21:1*, *21*(1), 1–12. https://doi.org/10.1186/S12874-021-01255-4

van der Lely, H. K. J. (1998). SLI in Children: Movement, Economy, and Deficits in the

Computational-Syntactic System. *Language Acquisition*, *7*(2–4), 161–192. https://doi.org/10.1207/S15327817LA0702-4_4

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends in Cognitive Sciences*, *9*(2), 53–59. https://doi.org/10.1016/J.TICS.2004.12.002

Van Der Lely, H. K. J., & Marshall, C. R. (2010). Assessing Component Language Deficits in the Early Detection of Reading Difficulty Risk. *Journal of Learning Disabilities*, *43*(4), 357–368. https://doi.org/10.1177/0022219410369078

van der Lely, H. K. J., & Ullman, M. T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes*, *16*(2–3), 177–217. https://doi.org/10.1080/01690960042000076

van der Lely, H., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition*, *62*(3), 245–290. https://doi.org/10.1016/S0010-0277(96)00783-4

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2010). Who is at risk for dyslexia? Phonological processing in five-to seven-year-old dutch-speaking children with SLI. *Scientific Studies of Reading*, *14*(1), 58–84. https://doi.org/10.1080/10888430903242035

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Development of phonological processing skills in children with specific language impairment with and without literacy delay: A 3-year longitudinal study. *Journal of Speech, Language, and Hearing Research*, *55*(4), 1053–1067. https://doi.org/10.1016/j.ridd.2011.11.005

Vang Christensen, R. (2019). Sentence Repetition: A Clinical Marker for Developmental Language Disorder in Danish. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4450–4463. https://doi.org/10.1044/2019_JSLHR-L-18-0327

Verhagen, J., & Blom, E. (2014). Asymmetries in the acquisition of subject–verb agreement in Dutch: Evidence from comprehension and production. *First Language*, *34*(4), 315–335. https://doi.org/10.1177/0142723714544412

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics*, *12*(1), 54–73. https://doi.org/10.1111/1473-4192.00024

Voci, S. C., Beitchman, J. H., Brownlie, E. B., & Wilson, B. (2006). Social anxiety in late adolescence: The importance of early childhood language impairment. *Journal of Anxiety Disorders*, *20*(7), 915–930. https://doi.org/10.1016/J.JANXDIS.2006.01.007

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: a meta-analysis. *Research in Developmental Disabilities*, *34*(9), 2586–2597. https://doi.org/10.1016/J.RIDD.2013.05.014

Wadman, R., Durkin, K., & Conti-Ramsden, G. (2008). *Self-Esteem, Shyness, and Sociability in Adolescents With Specific Language Impairment (SLI)*. *51*(4), 938–952. https://doi.org/10.1044/1092-4388(2008/069)

Wallan, A. A. (2018). *Evaluation of Arabic tests of sentence repetition and verbal short term memory for Saudi preschoolers. (Unpublished Thesis)*. City, University of London.

Watson, J. C. E. (2011). Word Stress in Arabic. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1–29). Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0124

Weismer, S. E. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Research*, *28*(2), 175–184. https://doi.org/10.1044/jshr.2802.175

Weismer, S. E. (1996). Capacity limitations in working memory: The impact on lexical and morphological learning by children with language impairment. *Topics in Language Disorders*, *17*(1), 33–44.

Weismer, S. E., Evans, J., & Hesketh, L. J. (1999). *An Examination of Verbal Working Memory Capacity in Children With Specific Language Impairment*. *42*(5), 1249–1260. https://doi.org/10.1044/JSLHR.4205.1249

Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword Repetition Performance in School-Age Children with and Without Language Impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865–878. https://doi.org/10.1044/jslhr.4304.865

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, *106*(1–4), 23–79. https://doi.org/10.1016/s0024-3841(98)00029-1

Wexler, K. (2003). Lenneberg's Dream: Learning, Normal Language Development, and Specific Language Impairment. In Y. Levy & J. C. Schaeffer (Eds.), *Language competence across populations : toward a definition of specific language impairment* (pp. 11–62). Mahwah, NJ: Erlbaum Associates.

Wexler, K., Schaeffer, J., & Bol, G. (2004). Verbal Syntax and Morphology in Typically Developing Dutch Children and Children With SLI: How Developmental Data Can Play an Important

Role In Morphological Theory. *Syntax*, *7*(2), 148–198. https://doi.org/10.1111/J.1467-9612.2004.00006.X

Wexler, K., Schütze, C. T., & Rice, M. (1998). Subject Case in Children with SLI and Unaffected Controls: Evidence for the Agr/Tns Omission Model. *Language Acquisition*, *7*(2–4), 317–344. https://doi.org/10.1207/s15327817la0702-4_8

Whitehouse, A. J. O., Watt, H. J., Line, E. A., & Bishop, D. V. M. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language and Communication Disorders*, *44*(4), 511–528. https://doi.org/10.1080/13682820802708098

Wiig, E., Secord, W., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Bloomington, MN: NCS Pearson.

Willis, B. H. (2008). Spectrum bias--why clinicians need to be cautious when applying diagnostic test studies. *Family Practice*, *25*(5), 390–396. https://doi.org/10.1093/FAMPRA/CMN051

Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, *9*(4–6), 349–363. https://doi.org/10.1080/09658210143000155

Windfuhr, K. L., Faragher, B., & Conti-Ramsden, G. (2002). Lexical learning skills in young children with specific language impairment (SLI). *International Journal of Language and Communication Disorders*, *37*(4), 415–432. https://doi.org/10.1080/1368282021000007758

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-Language Nonword Repetition by Bilingual and Monolingual Children. *American Journal of Speech-Language Pathology*, *19*(4), 298–310. https://doi.org/10.1044/1058-0360(2010/09-0064)

Winstanley, M., Webb, R. T., & Conti-Ramsden, G. (2018). More or less likely to offend? Young adults with a history of identified developmental language disorders. *International Journal of Language & Communication Disorders*, *53*(2), 256–270. https://doi.org/10.1111/1460-6984.12339

Wittke, K., & Spaulding, T. J. (2018). Which Preschool Children With Specific Language Impairment Receive Language Intervention? *Language, Speech, and Hearing Services in Schools*, *49*(1), 59–71. https://doi.org/10.1044/2017_LSHSS-17-0024

Wong, A. M. Y., Kidd, J. C., Ho, C. S. H., & Au, T. K. F. (2010). Characterizing the overlap between SLI and dyslexia in Chinese: The role of phonology and beyond. *Scientific Studies of*

Reading, *14*(1), 30–57.

Yew, S., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(5), 516–524. https://doi.org/10.1111/JCPP.12009

Zebib, R., Henry, G., Messarra, C., Hreich, E. K., & Khomsi, A. (2019). ELO-L: A Norm-Referenced Language Screening Test for 3 to 8-Year-Old Lebanese Children. *Arab Journal of Applied Linguistics*, *4*(2), 24–53.

Zebib, R., Tuller, L., Hamann, C., Ibrahim, L. A., & Prévost, P. (2020). Syntactic complexity and verbal working memory in bilingual children with and without Developmental Language Disorder. *First Language*, *40*(4), 461–484. https://doi.org/10.1177/0142723719888372

Appendices

Appendix A Individual raw scores of the background measures and the verb elicitation task for the TD and DLD groups in Study 1

ID	gender	age	*MPU*	A-QU-LITMUS-NWRT (/ 30)	*CPM* (/36)	% of correct verbs
DLD1	M	67	2.88	17	20	55
DLD2	M	69	3.14	19	21	70
DLD3	F	84	4.06	16	23	60
DLD4	F	85	4.1	12	20	51
DLD5	M	52	3.21	13	12	90
DLD6	M	58	3.12	10	18	65
DLD7	M	50	2.62	11	11	68
DLD8	M	94	6.27	23	19	91
DLD9	M	54	3.22	16	12	88
DLD10	M	48	2.19	9	10	78
DLD11	F	56	3.21	16	12	86
DLD12	M	66	4.98	18	16	96
DLD13	M	61	3.36	21	9	100
DLD14	F	89	3.77	17	14	78
TD1	M	57	6.47	30	19	100
TD2	M	59	5.21	30	14	98
TD3	M	71	4.19	30	18	100
TD4	F	75	5.46	30	16	100
TD5	F	42	2.97	19	8	91
TD6	M	60	5.1	30	17	100
TD7	F	66	5.26	29	21	100
TD8	F	56	3.46	28	18	96
TD9	F	84	6.31	30	21	100
TD10	F	54	3.93	30	14	93
TD11	F	56	5.11	28	15	96
TD12	F	36	2.41	16	NA	65
TD13	M	83	5.89	30	22	100
TD14	F	54	4.9	27	17	98
TD15	M	48	3.93	24	15	96
TD16	M	85	6.01	30	21	100
TD17	M	80	5.68	30	15	100
TD18	M	79	5.13	30	19	98
TD19	M	68	4.88	29	19	98
TD20	F	51	3.79	27	14	98

TD21	M	65	3.92	25	21	98
TD22	M	96	7.61	30	23	100
TD23	M	87	6.58	30	20	100
TD24	M	41	2.83	19	9	80
TD25	M	90	7.24	30	20	100
TD26	M	73	5.96	30	18	100
TD27	F	39	3.87	19	NA	73
TD28	F	43	4.21	21	8	80
TD29	F	47	4.53	25	10	91
TD30	M	49	4.69	23	15	95
TD31	M	43	3.91	20	10	78
TD32	M	55	5.45	30	23	100

Note. MPU = Mean Morpheme per Utterance. A-QU-LITMUS-NWRT = Arabic version of a Quasi-Universal Nonword Repetition Test. CPM = Colored Progressive Matrices.

.

Appendix B List of verbs used in the verb elicitation task in Study 1

Pair	Number Agreement	Gender Agreement	Tense: Present (A)	Tense: Past (B)
	Practice items		A. bit.qatˤ.tˤiʕ* cut-PRES-3FS	A. qatˤ. tˤa. ʕat cut-PAST-3FS
			B. byir.bu.tˤu tie-PAST-3P	B. ra.ba.tˤu tie-PAST-3P
1.			1. biyo:.kil **eat**-PRES-3MS	1. ʔa.kal **eat**-PAST-3MS
2.			2. byiʃ.rab **drink**-PRES-3MS	2. ʃi.rib **drink**-PAST-3MS
3.			3. byi.ɣas.sil **wash**-PRES-3MS	3. ɣas.sal **wash**-PAST-3MS
4.		Masculine	4. biy.maʃ.ʃit **brush**-PRES-3MS	4. maʃ.ʃat **brush**-PAST-3MS
5.			5. byir.sum **draw**-PRES-3MS	5. ra.sam **draw**-PAST-3MS
6.			6. byi.law.win **paint**-PRES-3MS	6. law.wan **paint**-PAST-3MS
7.			7. byi.yib.ri **sharpen**- PRES-3MS	7. ba.ra **sharpen**- PAST-3MS
8.			8. byif.taħ **open**-PRES-3MS	8. fa.taħ **pray**-PAST-3MS
9.	Singular		9. bit.far.ʃi **brush**-PRES-3FS	9. far.ʃat **brush**-PAST-3MS
10.			10. btik.tub **write**-PRES-3FS	10. kat.bat **write**-PAST-3FS
11.			11. bit.qusˤ **cut**- PRES-3FS	11. qasˤ.sˤ.at **cut**-PAST-3FS
12.		Feminine	12. bit.naʃ.ʃif **dry**-PRES-3FS	12. naʃ.ʃa.fat **dry**-PAST-3FS
13.			13. bitʃ.rab **drink**- PRES-3FS	13. ʃir.bit **drink**-PAST-3FS
14.			14. btaʕ.tˤi: **give**- PRES-3FS	14. aʕ.tˤat **give**-PAST-3FS
15.			15. bit.taʕ.mi **feed**- PRES-3FS	15. taʕ.mat **feed**-PAST-3FS

.16		16. byil.bi.su wear-PRES-3PL	16. lib.su wear-PAST-3PL
17.		17. byi.nadˤ.fu clean-PRES-3PL	17. nad.dˤa.fu clean-PAST-3PL
18.		18. biʃ.ra.bu drink-PRES-3PL	18. ʃir.bu drink-PAST-3PL
19.		19. byi.law.nu paint-PRES-3PL	19. law.wa.nu paint-PAST-3PL
20.		20. byik.tu.bu write-PRES-3PL	20. ka.ta.bu write-PAST-3PL
21.		21. byin.fu.χu blow-PRES-3PL	21. na.fa.χu blow-PAST-3PL
22.		22. byil.ʕa.bu play-PRES-3PL	22. liʕ.bu play-PAST-3PL
23.		23. bin.ʃu.ru hang-PRES-3PL	23. na.ʃa.ru hang-PAST-3PL
24.	Plural	24. byif.ta.ħu open-PRES-3PL	24. fa.ta.ħu open-PAST-3PL
25.		25. byi.maʃ.tu brush- PRES-3PL	25. maʃ. ʃa.tu brush- PAST-3PL
26.		26. byi.far.ʃu brush- PRES-3PL	26. far.ʃu brush- PAST-3PL
27.		27. byir.bu.tu tie- PRES-3PL	27. ra.ba.tu tie- PAST-3PL
28.		28. byi.naʃ.fu dry- PRES-3PL	28. naʃ.ʃa.fu dry- PAST-3PL
29.		29. byi.qusˤ.sˤu cut- PRES-3PL	29. qasˤ.sˤu cut- PAST-3PL
30.		30. byiʃ.la.ħu takeoff- PRES-3PL	30. ʃil.ħu takeoff- PAST-3PL

Note. PRES-3MS = present 3rd person masculine singular. PAST-3MS= past 3rd person masculine singular. PRES-3FS= present 3rd person feminine singular. PAST-3FS= past 3rd person feminine singular. PRES-3P= present 3rd person plural.PAST-3P= past 3rd person plural.

*underlined syllable are stressed.

Appendix C Individual raw scores for the TD and DLD groups of Study 2 and Study 3.

ID	age	LITMUS-SR-PA-72	AVET	ANPT	MPU	CPM	A-QU-LITMUS-NWRT
DLD-1	48	1.39	51.04	10	2.58	13	36.67
DLD-2	49	47.22	21.88	6.67	3.41	14	50
DLD-3	49	12.5	64.58	6.67	2.68	13	26.67
DLD-4	49	22.22	71.88	23.33	3.3	17	53.33
DLD-5	51	40.28	75	20	3.43	10	56.67
DLD-6	51	8.33	32.29	13.33	1.89	9	3.33
DLD-7	53	11.11	14.58	10	1.95	19	44.83
DLD-8	53	16.67	47.92	20	3.33	10	43.33
DLD-9	53	13.89	56.25	13.33	3.02	12	33.33
DLD-10	55	15.28	42.71	16.67	1.98	NA	40
DLD-11	60	27.78	82.29	23.33	3.22	12	80
DLD-12	57	1.39	37.5	3.33	2.95	10	43.33
DLD-13	57	6.94	47.92	3.33	3.68	12	50
DLD-14	60	2.78	47.92	16.67	3.4	10	46.67
DLD-15	61	43.06	38.54	16.67	2.92	13	66.67
DLD-16	60	20.83	84.38	23.33	3.22	14	56.67
DLD-17	62	23.61	78.13	10	3.97	13	43.33
DLD-18	67	43.06	89.58	26.67	2.95	18	73.33
DLD-19	70	43.33	71.88	36.67	4.29	22	56.67
DLD-20	81	6.94	76.04	23.33	2.8	13	53.33
DLD-21	81	27.78	71.88	33.33	4.23	19	60
DLD-22	59	51.39	48.96	43.33	2.6	15	76.67
DLD-23	82	47.22	72.92	36.67	4.51	23	86.67
DLD-24	73	56.94	70.83	73.33	3.11	13	83.33
DLD-25	82	27.78	72.92	23.33	3.93	19	26.67
DLD-26	82	2.78	78.13	36.67	4.51	23	16.67
DLD-27	60	2.78	62.5	30	2.16	14	70
DLD-28	69	23.61	72.92	0	4.61	19	70
DLD-29	64	44.44	78.13	36.67	3.47	16	43.33
DLD-30	48	50	63.54	23.33	3.39	13	73.33

TD-1	79	69.44	98.96	90	6.49	20	83.33
TD-2	65	91.67	100	96.67	6.28	14	100
TD-3	71	91.67	100	100	5.16	22	96.67
TD-4	48	73.61	94.79	56.67	3.15	9	96.67
TD-5	49	86.11	92.71	73.33	3.9	15	96.67
TD-6	50	37.5	90.63	53.33	3.96	12	66.67
TD-7	50	70.83	97.92	80	4.39	10	100
TD-8	51	70.83	93.75	23.33	4.29	11	100
TD-9	51	80.56	84.38	26.67	4.88	14	93.33
TD-10	51	59.72	95.83	26.67	4.66	16	86.67
TD-11	53	87.5	97.92	80	4.59	15	100
TD-12	54	90.28	100	76.67	5.53	12	100
TD-13	56	93.06	100	83.33	5.2	13	93.33
TD-14	56	83.33	100	56.67	5.03	13	96.67
TD-15	56	30.56	100	43.33	4.83	13	83.33
TD-16	56	65.28	96.88	36.67	4.14	15	93.33
TD-17	57	80.56	97.92	70	4.43	12	86.67
TD-18	57	69.44	87.5	40	5.57	11	93.33
TD-19	57	81.94	100	93.33	4.63	17	90
TD-20	60	79.17	100	36.67	5.61	13	90
TD-21	60	84.72	100	90	5.02	12	100
TD-22	62	95.83	100	70	4.9	12	96.67
TD-23	62	90.28	100	90	5.4	22	100
TD-24	64	95.83	100	83.33	5.11	12	100
TD-25	56	66.67	90.63	30	7.03	14	83.33
TD-26	61	76.39	100	63.33	5.5	14	90
TD-27	66	90.28	100	90	5.27	16	96.67
TD-28	71	95.83	100	100	6.42	16	96.67
TD-29	76	80.56	100	53.33	6.03	20	83.33
TD-30	72	87.5	100	100	6.64	21	90
TD-31	73	94.44	100	90	5.69	12	100
TD-32	60	84.72	97.92	83.33	6.31	12	96.67
TD-33	48	95.83	85.42	53.33	3.35	10	96.67
TD-34	48	84.72	91.67	43.33	4.49	14	96.67

TD-35	73	83.33	98.96	93.33	6.78	18	83.33
TD-36	57	68.06	83.33	40	4.03	18	63.33
TD-37	67	86.11	88.54	73.33	4.83	18	100
TD-38	57	81.94	73.96	50	5.76	16	100
TD-39	56	59.72	82.29	63.33	3.79	14	40
TD-40	70	90.28	100	90	5.11	18	100
TD-41	48	58.33	85.42	20	4.12	15	90.63
TD-42	70	90.28	100	90	5.11	18	100
TD-43	63	95.83	96.88	60	4.93	23	100
TD-44	80	97.22	100	100	5.69	23	100
TD-45	77	95.83	100	100	6.23	22	100
TD-46	66	87.5	98.96	66.67	5.15	14	90
TD-47	64	95.83	100	86.67	4.78	17	100
TD-48	74	90.28	100	100	5.13	16	100
TD-49	78	69.44	94.79	86.67	6.87	17	90
TD-50	76	88.89	100	100	5.97	20	96.67
TD-51	76	86.11	100	100	6.78	17	100
TD-52	79	83.33	100	96.67	5.89	22	100
TD-53	71	93.06	100	93.33	7.16	14	100
TD-54	67	98.61	100	100	5.67	17	100
TD-55	73	87.5	100	100	6.08	16	100
TD-56	78	94.44	100	96.67	6.43	17	100
TD-57	80	84.72	100	100	7.48	18	100
TD-58	69	87.5	100	96.67	5.19	22	100
TD-59	78	100	100	100	6.19	23	100
TD-60	78	95.83	100	93.33	5.98	15	100

Note. **DLD =** Developmental Language Disorder. **TD =**Typically Developing. **A-SR =** Arabic Sentence Repetition Test. **AVET =** Arabic Verb Elicitation Task. **ANPT =** Arabic Noun Plurals Test. **MPU =** Mean Morpheme per Utterance. **CPM =** Colored Progressive Matrices (Ravens, 2007). **A-QU-LITMUS-NWRT** = Arabic version of a Quasi-Universal Nonword Repetition Test.

Appendix D Items on the Arabic Version of Quasi-Universal LITMUS Nonword Repetition Test in Study 2 (dos Santos et al., n.d.)

Number of syllables	Number of CC	No.	Non-word (IPA)	Wordlikeness rating M(SD)	Wordlikeness
1	0	1	ˈfuk*	3.84(1.25)	HW
		7	ˈkib	4.28(.84)	HW
		12	ˈbaf	2.16(.99)	LW
	1	22	ˈklu	2.66(1)	HW
		6	ˈfla	1.36(.57)	LW
		24	ˈbli	2.36(.95)	LW
2	0	28	ˈkaˌbi	2.88 (.83)	HW
		2	ˈlaˌfi	1.92(.86)	LW
		3	ˈkaˌfib	1.36(.70)	LW
		13	ˈfaˌku	1.96(1.06)	LW
		20	ˈbuˌkif	2(1)	LW
		26	ˈbiˌlu	1.80(.82)	LW
	1	8	ˈbuˌkli	4.36(.86)	HW
		25	ˈkliˌfak	2.80(1.26)	HW
		15	ˈfluˌkif	1.56(.82)	LW
		16	ˈbluˌfa	1.36(.57)	LW
		19	ˈfaˌblu	1.16 (.37)	LW
		21	ˈfliˌku	1.92(.86)	LW
	2	10	ˈblaˌklu	1.56(.65)	LW
		4	ˈflaˌblu	1.32(.56)	LW
3	0	9	kifaˌbu	3.32(.85)	HW
		5	ˈbuˈfaˌki	1.36(.70)	LW
	1	14	ˈkaˈbuˌfik	1.56(.77)	LW
		29	ˈbiˈfaˌkub	1.56(.71)	LW
		11	ˈkuˈflaˌbi	1.32(.63)	LW
		17	ˈkuˈbaˌfli	1.32(.56)	LW
		23	ˈfiˈkuˌbla	1.84(.99)	LW
		18	ˈbiˈklaˌfu	2.08(1.22)	LW
		27	ˈfliˈbuˌka	1.80(.91)	LW
		30	ˈkliˈbaˌfu	1.32(.63)	LW

Note. HW: high wordlikeness. LW: low wordlikeness.

Appendix E Summary of model fitting in Study 2

Model	Term added	AIC	LR test	*p*
*M**	Intercept only (null model)	2731		
M0	*M* + Participant/Item intercepts	1708	$X^2(2)$ = 1027	< .001
M1	*M0* + age	1704	$X^2(1) = 6.94$	< .01
M2	*M1* + group	1626	$X^2(1) = 79.4$	< .001
M3	*M3* + nonword length	1613	$X^2(2) = 17.6$	< .001
M4	*M3* + CC number	1606	$X^2(2) = 10.6$	< .01
M5	*M4* + wordlikeness	1602	$X^2(1) = 5.49$	< .05
M6	*M5 +group:age*	1603	$X^2(1) = 1.62$	.20
M7	*M5* + group:nonword length	1605	$X^2(2) = .82$	.66
M8	*M6* + group:CC number	1596	$X^2(2) = 9.64$	< .001
M9	*M7* + group:Wordlikeness	1598	$X^2(1) = 0.52$	.47
M10	*M8* + nonword length:wordlikeness	1598	$X^2(2) = 2.48$	.29
M11	*M9* + nonword length:CC number	1596	$X^2(2) = 4.41$	.11
M12	*M10* + CC number :wordlikeness	1596	$X^2(1) = 2.11$	.15

Note. **M =** model.**AIC:** Akaike information criterion. **CC =** Consonant cluster. **LR =**likelihood ratio. For LR tests and p values refer to the comparison between the current model and the last significant model.

*LR tests and p values are not listed for *M* (null model) as it was not compared to a simpler model.

Appendix F List of items in the LITMUS-SR-PA-72 task of Study 3

Level	Target structure Mean length (SD)	substructure	Item	Arabic sentences "Approximate English translation"
1	*Past tense* 7.83 syllables (.75)	PAST-3MS	1	بابا اشترى* سيارة "Daddy bought a car"
			6	رسم الولد شجرة "The boy drew a tree"
		PAST-3FS	7	ماما غسلت الصحن "Mommy washed the dish"
			24	أكلت البسة جبنة "The cat ate cheese"
		PAST-3PL	12	لحقوا البقرات الولد "The cows chased the boy"
			21	الأولاد شربوا عصير "The boys drank juice"
	Present tense 7.66 syllables (.82)	PRES-3MS	14	سيدو بسوق السيارة "Grandpa is driving the car"
			17	الولد بلقط وردة "The boy is picking a flower"
		PRES-3FS	2	ماما بتقرأ قصة " Mommy is reading a story"
			11	بتلون البنت الحيط " The girl is painting the wall"
		PRES-3PL	4	بلعبو الاولاد فوتبول " The boys are playing football"
			13	البنات بنضفوا البيت "The girls are cleaning the house"
	Noun plurals 7.83 syllables (1.17)	Feminine sound plurals	3	تيتا كسرت الكاسات " Grandma broke the glasses"
			22	رمى الولد الطابات " The boy threw the balls"
		Masculine sound plurals	9	تيتيا نادت الطباخين "Grandma called the cooks"
			16	نادى الولد االبياعين "The boy called the salesmen"
		Broken plurals	5	أكل القرد الموز " The monkey ate the bananas"

			10	البنت ضيعت المفاتيح
				“The girl lost the keys”
	Possessive pronouns	CL-3MS	15	الولد ضيع طابته
	8.33 syllables (.82)			“The boy lost his ball”
			20	الولد كسر لعبته
				“The boy broke his game”
		CL-3FS	18	حممت البنت لعبتها
				“The girl washed her doll”
			23	البنت مشطت شعراتها
				“The girl brushed her hair”
		CL-3PL	8	البنات نضفوا بيتهم
				“The girls cleaned their house”
			19	الاولاد نشفوا ايديهم
				“The boys dried their hands”
2	*Passives*		25	الكاسة انكسرت من الولد
	9.5 syllables (.58)			“The glass got broken by the boy”
			34	الشباك انفتح من الهوا
				“The window got opened by the wind”
			38	السيارة توسخت من المطر
				“The car got dirty by the rain”
			41	الولد انضرب على بطنه
				“The boy got on his stomach”
	Clitic left dislocation		26	البلون الولد فقعه
	8.5 syllables (1.29)			“It is the balloon the boy popped”
			36	الكعكة ماما عملته
				“It is the cake mommy made”
			39	الهدية فتحتها البنت
				“It is the gift the girls opened”
			45	البنطلون لِبسُه الولد
				“It is the shirt the boy put on”
	Complement clauses	finite	29	فكرت ماما إنو البسة أكلت فار
	12 syllables (2.31)			“Mommy thought that the cat ate the mouse”
			44	فكر بابا انو الولد ضيع اللعبة
				“Dad thought that the boy lost the toy”
		Non-finite	32	بدها البنت تاكل جزرة
				“The girl wants to eat a carrot”
			47	سيدو بحب ياكل شوكلاطه
				“Grandpa likes to eat chocolate”

	Wh Object	Who	27	مين البنت اللي تيتا باستها
	10.25 syllables (1.04)			|Who is the girl that grandma kissed?”
			31	مين الولد اللي ساعدو بابا
				“Who is the boy that dad helped?”
			42	مين الولد اللي الكلب لحقه
				Who is the boy that the dog chased?”
			46	مين البيبي اللي طعمته ماما
				"who is the baby that mommy fed?”
		Which	28	أني بسة حملتها البنت
				“Which cat did the girl carry?”
			30	أنو تلفون الولد خربوا
				“Which phone did the boy break?”
			37	أني هدية البنت اعطتها
				“Which gift did the girl give?”
			43	أنو صندوق فتحه بابا؟
				“Which box did dad open?”
	Coordinates		33	ماما قرأت قصة و الولد نام
	13.9 syllables (1.63)			“Mommy read a story and the boy slept”
			35	القرد طلع عالشجرة و العصفور طار
				“The monkey climbed the tree and the bird flew”
			40	تيتا عملت شاي و بابا أكل بسكوت
				“Grandma made tea and dad ate biscuits”
			48	سيدو غسل السيارة و تيتا كتبت رسالة
				“Grandpa washed the car and grandma wrote a letter”
3	*Object relatives*	Reversible	52	هاي البنت اللي نيمتها تيتيا
	11.5 syllables (.93)			“This is the girl that mom put to sleep”
			59	هاد الحمار اللي الكلب دفعه
				“This is the donkey that the dog chased”
			64	هاد سيدو اللي ساعدو الولد
				“This is the grandpa that the boy helped”
			69	هاي البسة اللي السلحفاة عضتها
				“This is cat that the turtle bit”
		Irreversible	53	هاي القصة اللي تيتيا قرأتها
				“This is the story that grandma read”
			56	هاي البسة اللي لاقتها البنت
				“This is the cat that the girl found”
			65	هاد العصير اللي شربوا الولد
				“This is the juice that the boy drank”

		70	هاد الشباك اللي الولد سكره
			“This is the window that the boy closed”
Subject relatives	Reversible	49	هاد الولد اللي باس البنت
10.7 5 syllables (.46)			“This is the boy that kissed the girl”
		54	هاد الولد اللي حضن ماما
			“This is the boy that hugged mom”
		60	هاي تيتا اللي صورت البنت
			“This is the grandpa who took a picture of the girl”
		67	هاي البنت اللي شافت الارنب
			“This is the girl that saw the rabbit’
	Irreversible	50	هاي البنت اللي اشترت كتاب
			“This is the girl that bought a book”
		63	هاد الولد اللي غسل السيارة
			“This is the boy that washed the car”
		72	هاد الولد اللي نفخ البلون
			“This is the boy that blew the balloon”
		57	هاي البنت اللي رسمت بيت
			“This is the girl who drew a house”
Conditionals		51	اذا الولد بيعمل الواجب، راح ياخد نجمة
13.75 syllables (1.50)			“If the boy does the homework, he will take a sticker”
		62	اذا الولد برتب الغرفة، راح يروح عالحديقة
			“If the boy tidies the room, he will go to the park”
		66	اذا ماما بتشتري بيض، راح تعمل كعكة
			“If mommy buys eggs, she will make a cake”
		71	اذا البنت بتساعد ماما، راح تشتري فستان
			"If the girl helps mommy, she will buy a dress”
Subordinates		55	البنت وقعت عشان الأرض مبلولة
12.25 syllables (1.71)			“The girl fell because the floor is wet”
		58	الولد عيط عشان عشان ضيع اللعبة
			“The boy cried because he lost the toy”
		61	الولد بيدرس عشان عنده امتحان
			“The boy is studying because he has a test”
		68	البنت بتشرب مي عشنها عطشانة
			“The girl is drinking water because she is thirsty”

Note. * Underlined word is the target language-specific structure. **PAST-3MS =** 3rd person masculine singular past. **PAST-3FS** = 3rd person feminine singular past. **PAST-3PL** = 3rd person plural past. **PRES-3MS =** 3rd person masculine singular present. **PRES-3FS =** 3rd person feminine singular present. **PRES-PL =** 3rd person plural present. **CL-3MS =** 3rd person masculine singular clitic. **CL-3FS =** 3rd person feminine singular clitic . **CL-3PL =** 3rd person plural clitic.

Appendix G Average age of acquisition of 50 words included in the LITMUS-SR-PA-71. Data from Lebanese Arabic-speaking children (from Łuniewska et al., 2019)

Word	Gloss	Age of acquisition (years)
	Nouns	
1. طابة	Ball	1.69
2. موزة	banana	2.29
3. جزرة	carrot	3.19
4. مي	Water	4.43
5. بسة	cat	1.86
6. جبنة	Cheese	2.98
7. فرشاي	Toothbrush	3.38
8. بقرة	Cow	3.07
9. كلب	Dog	2.19
10. لعبة	Doll	2.12
11. وردة	Flower	3.12
12. بيت	house	2.14
13. مفتاح	Key	3.05
14. قرد	monkey	4.00
15. فار	Mouse	3.43
16. أرنب	Rabbit	3.1
17. بلوزة	Shirt	3.52
18. نجمة	Star	3.71
19. تلفون	Telephone	2.74
20. سلحفاة	turtle	3.71
21. صندوق	Box	2.67
	Verbs	
22. بيفرشي	brush	3.12
23. بيحمل	carry	3.00
24. بيمشط	brush hair	3.38
25. بيعيد	cry	2.17
26. بيرسم	draw	3.14

27. بيشرب	drink	2.02
28. بيسوق	drive	3.71
29. بينشف	dry	3.36
30. بياكل	eat	1.95
31. بيوقع	fall	2.29
32. بيطعمي	feed	3.24
33. بيطير	fly	3.24
34. بيحضن	hug	4.10
35. بيضرب	hit	4.33
36. بيبوس	kiss	2.45
37. بيفتح	open	2.88
38. بيلون	paint	5.95
39. بيلقط	pick	4.11
40. بيدفع	push	3.81
41. مطر	rain	3.02
42. بيقرأ	read	3.71
43. بينام	sleep	1.90
44. بيرمي	throw	2.95
45. بيغسل	wash	3.45
46. يطير	fly (v)	3.24
47. بيكتب	write	3.67
48. بيعض	Bite(v)	3.29
49. بينفخ	Blow(v)	3.00