

Western  Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

12-12-2017 2:00 PM

Linguistic and Cognitive Measures in Arabic-Speaking English Language Learners (ELLs) and monolingual children with and without Developmental Language Disorder (DLD)

Areej Mazin Balilah
The University of Western Ontario

Supervisor
Archibald, Lisa
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Areej Mazin Balilah 2017

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [First and Second Language Acquisition Commons](#)

Recommended Citation

Balilah, Areej Mazin, "Linguistic and Cognitive Measures in Arabic-Speaking English Language Learners (ELLs) and monolingual children with and without Developmental Language Disorder (DLD)" (2017). *Electronic Thesis and Dissertation Repository*. 5088.
<https://ir.lib.uwo.ca/etd/5088>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Understanding the current level of language knowledge in English Language Learners (ELLs) can present a challenge. The standardized language tests that are commonly used to assess language tap prior knowledge and experience. ELLs may score poorly on such ‘knowledge-based’ measures because of the low levels of exposure to each of their languages. Considerable overlap has been found on several knowledge-based measures (Paradis, 2010) between ELLs and monolingual children with an unexpected delay in language development known as Developmental Language Disorder (DLD). Measures of cognitive processing, on the other hand, are less dependent on ELLs’ linguistic knowledge because they employ nonlinguistic or novel stimuli to tap skills considered to underlie language learning. It has been suggested that processing-dependent tasks such as measures of verbal short-term memory may differentiate ELLs from children with DLD (Kohnert, Windsor, & Yim, 2006; Paradis, Schneider, & Duncan, 2013). This thesis presents three studies that investigated the performance of Arabic-speaking ELLs and monolingual children with and without DLD on linguistic and cognitive measures. Study 1 provided a description of the performance of monolingual Arabic-speaking children on a battery of Arabic language tests. The results of study 1 revealed that the majority of language measures were sensitive to developmental change in younger children between the ages of 6 and 7. Study 2 demonstrated lower standardized scores by ELLs on the Arabic and English knowledge-based language tasks. However, ELLs scored above or at age-level expectations on the cognitive measures, with the exception of an Arabic-nonword repetition task. Study 3 found a significant overlap between ELLs and monolingual Arabic-speaking children with DLD on first language (L1) knowledge-based measures. With the exception of the Arabic nonword repetition task, verbal short-term and working memory tasks distinguished ELLs from children with

underlying language impairment. The results indicated that there is a need to develop language assessment measures that evaluate a broad range of language abilities for Arabic-speaking children. The findings also suggested that unlike knowledge-based measures, cognitive measures may be valid assessment tools that minimize the role of linguistic knowledge and experiences and help distinguish between ELLs and children with DLD.

Keywords: English Language Learners (ELLs), Developmental Language Disorder (DLD), Arabic-speaking children, knowledge-based measures, and processing-dependent measures.

Acknowledgements

First, I would like to gratefully acknowledge my supervisor, Dr. Lisa Archibald, for her continuous support throughout my PhD. I am thankful for all of Dr. Archibald's advice, guidance, and motivation during the time I spent researching and writing this thesis. I also want to thank my husband, Tariq, and my three lovely daughters, Wid, Ghad, and Nad, for their continued support. To my mother, Kiriah, and my father, Dr. Mazin, thank you both for your unwavering love and support. Thank you to all of my officemates at The Language and Working Memory Lab, and Autism Spectrum and Language Disorders Lab; we have shared so many great times. I would also like to thank all of the principals, teachers, and staff of the 12 schools that participated in my PhD study (Jeddah, Saudi Arabia: Qurtobah School, Al-Noor School, Al-Thania School, Al-Taawn School, Al-Amjad School, The 58 School, The 37 School, The 57 School, The 145 School, and The 92 School; London Ontario, Canada: Al-Taqwa School and The Saudi London Learning Center). I would like to thank the children and their families whose participation enabled me to complete my PhD thesis. Finally, I would like to thank King Abdulaziz University-Saudi Arabia for my scholarship grant. Thank you all!

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of figures	x
List of Appendices	xi
Chapter 1	1
General Introduction	1
English Language Learners (ELLs)	2
ELLs' phonological acquisition	3
ELLs' lexical acquisition	4
ELLs' morphosyntactic acquisition	5
ELLs' acquisition of reading and literacy	6
Factors that Influence ELLs' Language Acquisition	7
Age of English exposure	8
Current input/output	9
Home language experiences	9
Socioeconomic status and maternal education	10
Language Shift and First Language Loss	11
Developmental Language Disorder (DLD)	12
Patterns of Language Impairment	12
Phonology.	13
Lexicon	14
Morphology and syntax	14
Pragmatics	15
Theoretical Accounts of DLD	15
Domain-specific theories	15
Domain-general theories	17
Identifying Children with DLD	18
Arabic Language	21
Identification of DLD in ELLs	22
Challenge and Complexity in the Assessment of ELLs	22
Overlap between ELLs and Children with DLD	24

Language-General Verses Language-Specific Measures	25
The Research in this Thesis	27
Chapter 2	29
The Measurement of Language Ability and Impairment in Arabic-speaking Children	29
Method	37
Participants	37
Materials and Procedure	38
Results	42
Discussion	51
Study Limitation	55
Future Direction	55
Conclusion	56
Chapter 3	58
English Language Learners (ELLs): Linguistic Gaps and Cognitive Strengths	58
Methods	67
Participants	67
Procedure	68
Arabic Language Test Battery.	69
Vocabulary	69
Oral language	69
Reading efficiency	70
English Language Test Battery.	70
Vocabulary	70
Oral language	70
Reading efficiency	71
Cognitive measures.	71
Short-term and working memory	71
Nonverbal intelligence	73
Parent questionnaire	73
Results	74
ELLs' Performance on Language Measures including Reading	74
ELLs' Performance on Cognitive Measures	75
Experience-based Predictors of Performance	76
Discussion	81

Study Limitation	87
Conclusion and Future Direction	87
Chapter 4	89
Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Arabic-Speaking English Language Learners (ELLs) Compared to Children with Developmental Language Disorder (DLD)	89
Methods	100
Participants	100
Materials and Procedure	102
Vocabulary tests	102
Language tests	103
Reading efficiency	104
Short-term and working memory	105
Nonverbal intelligence	106
Parent questionnaire	106
Results	107
Arabic Vocabulary, Language, and Reading Measures	107
Verbal Short-term and Working Memory	108
Visuospatial Short-term and Working Memory	110
Discussion	111
Conclusion and Future Directions	116
Chapter 5	118
General Discussion	118
The Measurement of Language Ability and Impairment in Arabic-Speaking Children	118
Summary of findings	119
Implications of findings	120
Linguistic and Cognitive Measures in Arabic-Speaking ELLs	120
Summary of findings	121
Implications of findings	122
Linguistic and Cognitive Measures in Arabic-speaking ELLs and Monolingual Arabic-Speaking Children With and Without DLD	122
Summary of findings	123
Implications of findings	125
Practical Implications	125
Conclusions	126

References	128
Appendix A	156
Appendix B	159
Appendix C	160
CURRUCULUM VITA	161

List of Tables

TABLE 1: Descriptive statistics for the Arabic Receptive-Expressive Vocabulary Test z-scores, and raw scores for all remaining language tasks	42
TABLE 2: Correlation between the language subtest raw scores; partial correlation (controlling for age in months) in the upper triangle and zero-order correlation in bottom triangle.	49
TABLE 3: The number of participants who scored below the cutoff on one or two language measures.	51
TABLE 4: Descriptive statistics for the language measures.	73
TABLE 5: Descriptive statistics for the cognitive measures.	74
TABLE 6: Stepwise regression models results for Arabic language measures.	77
TABLE 7: Stepwise regression models results for English language measures.	77
TABLE 8: Stepwise regression models results for cognitive measures.	78
TABLE 9: Descriptive statistics for criterion measures for all groups.	98
TABLE 10: Standard and raw scores on all language measures for the ELLs, A-TD, and A-DLD groups	104
.	
TABLE 11: Descriptive statistics of the verbal short-term and working memory subtests of AWMA and A-NWR task raw and standard scores for the ELL, A-TD, and A-DLD groups.	105
TABLE 12: Descriptive statistics of the visuospatial short-term and working memory subtests of AWMA raw and standard scores for the ELL, A-TD, and A-DLD groups.	107

List of figures

- FIGURE 1:** Mean z -scores as a function of age for the receptive and expressive subtests of the Arabic Receptive-Expressive Vocabulary Test showing significant performance growth with each increase in age band. 44
- FIGURE 2:** Mean z -scores as a function of age for the verbal and nonverbal subtests of the Arabic Language Screening Test showing significant performance growth between the 6 and 7 year olds only. 45
- FIGURE 3:** Mean z -scores as a function of age for the four subtests of the Arabic Language Test showing significant performance growth between 6 and 7-year-old only for the expressive language but not sentence repetition subtests. 46
- FIGURE 4:** Mean raw scores of sentence repetition-B subtests of the Arabic Language Test for male and female participants as a function of age, showing significantly higher scores for the males than females at 6 and 7 but not at 8 and 9 years of age. 47

List of Appendices

Appendix A: Descriptive statistic for all Arabic language tasks scores as a function of age band for the monolingual Arabic-speaking children.	152
Appendix B: Ethics Approval (1)	155
Appendix C: Ethics Approval (2)	156

Chapter 1

General Introduction

Children with an unexpected failure to develop language at the typical rate, despite normal neurological and socioemotional development, as well as average educational and experiential opportunities, are referred to as children with Developmental Language Disorder (DLD; also known as Specific Language Impairment; Bishop, Snowling, Thompson, & Greenhalgh, 2017). Another group of children who may appear to have limited language skills at school is English Language Learners (ELLs), or children who are receiving instruction in their second language (L2 English) or in a language other than their minority first language (L1). Research has found that during the early stage of L2 learning, typically developing ELLs' (TD ELLs) perform at a comparable level to monolingual children with DLD on several '*knowledge-based*' assessment tools that tap existing language knowledge and experience, such as morphosyntactic measures (Paradis, 2005; Paradis, Rice, Crago, & Marquis, 2008), and vocabulary measures (Umbel, Pearson, Fernandez, & Oller, 1992; Windsor & Kohnert, 2004). On the other hand, '*processing-dependent*' measures probing the abilities supporting language learning may be less dependent on ELLs' linguistic knowledge, and may distinguish ELLs from children with underlying language impairment. It should be noted that in any consideration of bilingual language development, the specific languages being learned must be considered. This thesis considers the case of Arabic-speaking children, including both monolingual Arabic-speaking children and Arabic-speaking ELLs. Only a few studies have focused on the measurement of language ability and impairment in Arabic-speaking children (Wiig & El-Halees, 2000). This thesis adopted an epidemiological approach to assess sensitivity to developmental change in several language measures of school age Arabic-speaking children.

Moreover, this thesis is concerned with examining ELLs' performance on a wide range of cognitive and linguistic measures. Serving as an introduction for the three studies presented in this thesis, this chapter will provide a general overview of language acquisition for ELLs and children with DLD, factors that influence ELLs' language acquisition, challenges in the assessment of ELLs, and examine language processing in ELLs.

English Language Learners (ELLs)

Children from minority ethnolinguistic communities who live in a majority English social context and attend school where English is the language of instruction are commonly referred to as *English Language Learners* (ELLs). Over the past decade, the number of ELLs has significantly increased in Canada (Paradis, Emmerzael, & Duncan, 2010) and the United States (Goldstein, 2012). *Interlanguage* is a term used to describe L2 learner language development, mapping a continuum starting from using the language productively and ending with reaching native-speaker proficiency. Interlanguage is a dynamic system that is somewhat similar to the target language system (Paradis, Genesee, & Crago, 2011). That is, interlanguage balances the processes of transfer from the L1 and the processes of language development with the target language (Paradis, 2007). It should be noted that the majority of interlanguage patterns have been found to be developmental rather than transfer based (Paradis et al., 2011). The following section will describe some developmental and transfer-based patterns in ELLs' language acquisition in several areas of language such as phonology, lexical, morphosyntactic, and reading. The review of the developmental patterns of English in ELLs is important to establish proper expectations regarding the typical patterns of English L2 in children. In addition, understanding how long it takes ELLs to attain native-speaker proficiency in different domains of language can be very important for teachers and clinicians involved in assisting ELLs.

ELLs' phonological acquisition. Although most researchers agree that younger ELLs can achieve native-like pronunciation in English, it can take more than two years to achieve this level (Gilhool, Burrows, Goldstein, & Paradis, 2009; Paradis et al., 2011). One study that examined ELLs' phonological acquisition of English found that after one year of exposure to English, ELLs achieved greater than 90% accuracy in the production of many types of consonants and vowels (Paradis et al., 2011). Acquiring fricative consonants, however, can take longer for ELLs to acquire, with less than 80% accuracy in pronunciation (Gilhool et al., 2009; Paradis et al., 2011). This finding supports the observation that it can take more than two years for ELLs to achieve native-like pronunciation in their L2 (Gilhool et al., 2009). In general, the phonological development patterns of ELLs are largely parallel to the developmental patterns of younger, English-speaking, monolingual children. For example, consonant clusters (groups of consonants in a row) are considered to be an area of difficulty and develop later for monolingual English-speaking children. Monolingual English-speaking children go through a stage where they omit one of the consonants in a cluster thereby easing the motor demands (e.g., *play* as /*pei*/). Similar phonological errors have also been reported in ELLs, even for ELLs with an L1 that has consonant clusters in the phonology (Gilhool et al., 2009). Moreover, the L1 phonological system is often considered to be a primary source that influences L2 speech development. For example, ELLs with a Spanish L1 pronounced phonetic segments (sounds) that are shared between English and Spanish more accurately than the phonemes that are present only in one of their languages (Goldstein, 2012). On the other hand, when the L1 and L2 differ in pronunciation rules, transfer errors in L2 pronunciation may arise. For instance, in Spanish when consonant clusters appear at the beginning of words (like /*st*/ or /*sp*/), they are often preceded by a vowel. This type of pronunciation rule can be transferred by ELLs from their Spanish L1, and

errors such as pronouncing the word *stop* /stap/ as “estop” /ɛstap/ can occur in the speech production of ELLs, at least initially (Paradis et al., 2011).

ELLs’ lexical acquisition. Available evidence suggests that it could take several years of schooling in English for ELLs to achieve similar vocabulary size and composition as their native-speaker peers, ranging from 3 years of schooling (Paradis et al., 2011) to 6 years of schooling (Oller & Eilers, 2002), depending on the circumstance of acquisition. Studies that examine the lexical developmental patterns in ELLs indicate that the vocabulary knowledge of ELLs is shared and distributed across both of their languages (Pearson, Fernández, & Oller, 1993). The distributed nature of the lexical-semantic knowledge of ELLs can influence vocabulary size in both languages (Oller, Pearson & Cobo-Lewis, 2007). Indeed, ELLs have been repeatedly found to score below their monolingual peers in L1 and L2 vocabulary measures as a consequence of the lexicon differences between both populations (Oller et al., 2007).

Receptive-expressive gap is a common feature of ELLs’ vocabulary development, in which ELLs tend to score lower on productive than receptive vocabulary in English (Kan & Kohnert, 2005; Paradis et al., 2011). The receptive-expressive gap in English has been found to be larger in children with limited English experience, whereas the gap narrows with increased experience in English (Gibson, Peña, & Bedore, 2014). The *general all-purpose* (GAP) words phenomenon has also been documented in the lexical developmental pattern of ELLs. That is, ELLs in the early stage of learning English have been found to use verbs such as *do* that have broad and flexible meaning as a GAP verb when more specific and advanced words would be appropriate (Golberg, Paradis, & Crago, 2008). Interestingly, research on ELLs has revealed that lexical acquisition can be faster for older ELLs who are more cognitively mature and who have already established their L1 as compared to younger monolingual speakers. That is, in addition to being

more cognitively mature, ELLs can also benefit from their L1 existing lexical knowledge to accumulate vocabulary at a faster rate than the younger monolingual speakers of the same language (Paradis, 2007).

ELLs' morphosyntactic acquisition. The available evidence suggests that it can take between 3 to 5 years for ELLs to attain the same oral English proficiency as native speakers (Hakuta, Butler, & Witt, 2000). Studies that have examined the ELLs' morphosyntactic acquisition of English found that the order sequence of the morpheme acquisition pattern in ELLs is similar to what has been reported in monolingual English-speaking children. For instance, ELLs tend to master certain morphemes before others (e.g., plurals *-s* and progressive *-ing* before past tense *-ed*, and third-person singular *-s*) (Jia, 2003; Jia & Fuse, 2007; Paradis, 2008). In addition, ELLs acquired finite verb morphology in English later than non-finiteness-related morphology, which is similar to what has been found in monolingual English-speaking children (Paradis, 2005). ELLs' errors with grammatical morphemes and syntactic structures has also been of interest to researchers. In general, ELLs' errors with grammatical morphemes appear to parallel the common errors that have been reported in younger typically developing (TD) monolingual English-speaking children and monolingual children with language impairment. For example, more recent research shows that the omission of tense/agreement (finiteness) markers can be described as a significant characteristic of ELLs' interlanguage errors (Paradis, 2005). ELLs have also been found to incorrectly choose articles, such as using the definite article *the* instead of an indefinite article (Paradis, 2005). *Overregularization* forms are also a common phenomenon in English-speaking monolingual children during their school years, occurring when a child applies the regular morphological rule to an irregular form. ELLs with a Spanish L1 have been found to produce and accept as correct overregularization forms with a

higher frequency than monolingual English-speaking children (Jacobson & Cairns, 2008).

Finally, on the whole, studies that examined the morphosyntax transfer from L1 to L2 in ELLs suggest that the morphosyntactic interlanguage patterns in ELLs are mostly developmental rather than transfer based (Dulay & Burt, 1973, 1974; Zdorenko & Paradis, 2012).

ELLs' acquisition of reading and literacy. Oller et al. (2007) found that ELLs with a Spanish L1 perform the same as their monolingual peers on word-decoding skills, whereas ELLs remained behind their monolingual peers in English vocabulary. According to Oller et al. (2007) ELLs approach native-speaker performance faster for some linguistic domains such as basic word-decoding skills, while some other linguistic domains appeared to develop slower in ELLs such as lexical acquisition (Oller et al., 2007). The concepts of common underlying proficiency can provide more explanation regarding ELLs' different developmental trajectories (Oller et al., 2007; Paradis et al., 2011). That is, the primary skills that support learning how to read and write have been found to be transferable from L1 to L2 learning, especially in children who are in the early stages of learning to read and write (Genesee, Lindholm-Leary, Saunders, & Christian, 2006). Evidence from a large body of research shows that ELLs who acquire knowledge of names and the sounds of the alphabet, and critical skills for word decoding such as certain types of phonological awareness at home before school entry, can transfer these skills to their L2 reading and writing (Genesee et al., 2006). As a result, ELLs who acquire the foundational skills that support learning to read and write at home in their L1, learn to read and write more quickly in their L2 than children who do not acquire these skills (Genesee et al., 2006). The transfer of these skills is found to be greater when a child's L1 and L2 shared similar typology such as the case of English and Spanish or French. A growing body of evidence, however, shows that transfer can occur even between two phonologically and orthographically distinct languages such

as Arabic and English (Abu-Rabia & Siegel, 2002). Indeed, many researchers have shown that there is a positive relationship between children's scores on tests of phonological awareness and alphabetic knowledge in their L1 before school entry and their later reading comprehension and word decoding skills in their L2 (Erdos, Genesee & Savage, 2011; Genesee et al., 2006). It should be noted that reading achievement in later stages of L2 development, however, required more advanced oral language skills by ELLs (Genesee, Lindholm-Leary, Saunders, & Christian, 2005).

The above section has reviewed the developmental patterns of English in ELLs' language acquisition in several areas of language such as phonological, lexical, and morphosyntactic. The finding of the studies above suggested that ELLs can take 3-5 years to achieve the same oral language proficiency as native-speakers. Importantly, ELLs achieved native-like mastery faster for some linguistic domains such as basic word-decoding skills, while some other linguistic domains appeared to develop slower in ELLs. The developmental pattern of ELLs who are in the early stage of their L2 learning mirrors that of the developmental patterns of younger, English-speaking, monolingual children in several areas of language such as phonology and morphosyntax.

Factors that Influence ELLs' Language Acquisition

ELLs' rate of their L2 acquisition is highly variable and dependent on a number of relevant experiential factors. Various factors have been identified as a source of individual differences in ELLs' language acquisition including age of English exposure, current input and output, home language experiences, maternal education and socioeconomic status. The following section will discuss factors related to ELLs' exposure to English that have been shown to impact their language acquisition.

Age of English exposure. Broadly speaking, ELLs who learn their L2 early in life, at least before 3 years of age, have more experience and fluency in their second language compared to ELLs who begin the acquisition of their L2 later in life or after establishing their L1 (Paradis et al., 2011). For example, Davison and Hammer (2012) found that ELLs with a Spanish L1 who began to learn English before preschool entry were more likely to master English grammatical forms by grade one than ELLs who did not learn English until entering preschool. Similarly, ELLs who learn their L2 simultaneously with their L1 have been found to have higher initial English vocabulary and oral comprehension abilities than ELLs who begin to learn their L2 after they have established their L1 (Hammer, Lawrence, & Miccio, 2008; Oller & Eilers, 2002). Moreover, Jia, Aaronson, and Wu (2002) reported that age of English onset in ELLs with a Mandarin L1 correlated with a higher performance on grammatical judgment tasks.

Nevertheless, other researchers who have looked at age of acquisition observed advantages for older ELLs who begin to learn English later such as in their middle school years, over younger ELLs. That is, older ELLs have more developed cognitive skills, and older ELLs can also benefit from their previous experience with schooling and literacy; these skills were found to transfer from L1 to L2 (Golberg et al., 2008). Indeed, older ELLs have been found to learn faster and develop a larger vocabulary in English than younger ELLs (Golberg et al., 2008). Similarly, Jia and Fuse (2007) found that older children/adolescents acquired grammatical morphemes at a faster rate than younger ELLs. For long-term learning, however, younger ELLs achieved 90% correct use of grammatical morphemes as compared to older ELLs. Therefore, older ELLs experienced a faster rate of growth in short-term learning, whereas younger ELLs seem to be better in long-term learning (Paradis et al., 2011).

Current input/output. Researchers who have examined the impact of language use on ELLs' language skills show a positive relationship between language use, such as the amount of L2 practiced at school, home, and in the community, and ELLs' language skills. For example, work conducted by Marchman and Martinez-Sussmann (2002) found that children's vocabulary size was significantly correlated with the amount of English spoken in the child's language-learning environment. ELLs' vocabulary growth has been found to be strongly associated with the total number of words produced by teachers (Bowers & Vasilyeva, 2011). Moreover, Bohman, Bedore, Peña, Mendez-Perez, and Gillam (2010) found that children's use of English is a significant predictor of their English semantic and morphosyntactic abilities. As the quantity of L2 input has been found to facilitate L2 learning, studies have highlighted the importance of the quality of L2 exposure on L2 learning for ELLs. That is, exposure to rich and diverse vocabulary and complex grammar by ELLs either at school or outside of school has been found to play a major role in a child's L2 acquisition (Paradis et al., 2011). For example, Jia (2003) and Jia and Aaronson (2003) showed that increasing the richness of the L2 environment around ELLs, such as the number of hours of English TV watched, the number of English books read, the number of English native-speaking friends, and the percentage of time English was spoken at home, was associated with the faster acquisition of L2 skills, including morphosyntactic abilities.

Home language experiences. Language spoken by the family members of ELLs is highly variable, which results in differences in the exposure and usage of both the L1 and the L2 among ELLs (Goldstein, 2012). Moreover, ELLs' home language exposure is dependent on the context in which the communication occurs and the topics of conversation. For example, family members may use the L1 in conversations that occur at home and English in conversations that occur outside the home. In addition, discussion about home life by family members may occur in

the L1, whereas conversation about school may occur in English. As a result, ELLs generally build different vocabulary content for each language depending on contexts in which they have been exposed to various words (Engel de Abreu, Baldassi, Puglisi, & Befi-Lopes, 2013; Goldstein, 2012). ELLs generally learned vocabulary associated with their home environment in their L1, whereas vocabulary associated with items or concepts learned in school was learned in English (Goldstein, 2012). Furthermore, the quality of English used at home has also been found to play an important role in ELLs' L2 acquisition. Several studies suggested that the frequency of English used by family members at home did not necessarily promote ELLs' L2 language acquisition (Duursma et al., 2007; Paradis et al., 2011). Paradis (2010) reported that the benefits of exposure to English at home in ELLs depended on the parents' proficiency in English. Parents with limited English skills may not be able to provide rich L2 input, especially after the children reached the early stage of English language acquisition.

Socioeconomic status and maternal education. Differences in families' socioeconomic status (SES) measured most commonly by mother's level of education are consistently associated with individual variation in the language acquisition of monolingual children (Brooks-Gunn, Han, & Waldfogel, 2002), and ELLs (Oller & Eilers, 2002). Research on ELLs shows that higher parental education is a strong predictor of ELLs' language abilities (Bohman et al., 2010). Golberg et al. (2008) found that ELLs' with higher-educated mothers have a larger vocabulary size than ELLs with less-educated mothers. In addition, ELLs with higher-SES have been found to have a better proficiency in English, and achieved native-speaking levels of proficiency in English faster than ELLs with lower SES (Cobo-Lewis, Eilers, Pearson, & Umbel, 2002; Hakuta et al., 2000; Oller & Eilers, 2002). Interestingly, Portes and Rumbaut (2001) reported that parents with higher compared to lower SES promoted language development in both languages

of their children. That is, parents with higher SES have more resources that allow them to provide their children with more support in their L1 and L2 than parents with lower SES.

Language Shift and First Language Loss

ELLs who speak a minority language at home and who are in the process of learning the majority language as their L2 through schooling and through contact with the majority societal language are L1 minority L2 learners. The minority-majority language contact situation of ELLs has been widely studied as a factor affecting L1 skills. That is, little community support and the lack of educational opportunities in ELLs' L1 have been found to impact their L1 linguistic knowledge and use, especially their lexical and grammatical abilities, which are skills primarily learned in schools (Anderson, 2012). The *Language shift* is a common phenomenon related to the minority-majority sociolinguistic status of ELLs in which the use pattern of the two languages gradually changes over time (Anderson, 2012). Shift from the use of the minority L1 to the use of the majority language can be the result of reduced or diminished ELLs' L1 abilities over time, a process described as *L1 loss*. Another phenomenon that co-occurs with L1 loss is *L1 incomplete acquisition* in which ELLs' skills in their L1 do not advance or develop further, rather than losing acquired L1 skills over time (Anderson, 2012; Paradis et al., 2011). Several patterns have been identified by Anderson (2012) as examples of L1 lexical/morphosyntactic language loss among ELLs with a Spanish L1, such as limited vocabulary, errors in verb morphology, and error in word order (use English Verb-Object word order instead of the appropriate Spanish word order). Importantly, the phenomena of L1 loss/incomplete acquisition in ELLs may result in the over diagnosis of language impairment in ELLs, particularly in lexical and grammatical skills (Anderson, 2012).

In summary, the studies presented in this section regarding the factors that influence ELLs' language acquisition demonstrate that there is a great deal of variation among ELLs. Moreover, the findings of studies that discussed the unique contributions of each factor in ELLs' language acquisition have not been entirely consistent, such as the results of the effects of age on ELLs' L2 learning. Such findings demonstrate the need for more research to understand the impact of several factors on ELLs' language acquisition. Importantly, ELLs' abilities in both languages may not remain stable, as shifts from the minority L1 to the majority L2 is a common phenomenon among ELLs.

Developmental Language Disorder (DLD)

Children with a significant limitation in their linguistic abilities but who also have largely typical cognitive abilities are referred to as children with Developmental Language Disorder (DLD; also known as specific language impairment; Bishop, Snowling, Thompson, & Greenhalgh, 2017). Children with DLD do not exhibit any other developmental problems, such as hearing impairment, low non-verbal intelligence test scores, neurological damage, or oral motor skills (Leonard, 2014). DLD is estimated to occur in 7-8% of children in kindergarten and the first year of primary education (Tomblin, Zhang, Buckwalter, & O'Brien, 2003). The language difficulties of children who exhibit early language delay can persist throughout their school years and adulthood (Snowling, Bishop, & Stothard, 2000). DLD can affect children's academic learning, especially in reading (Catts, Fey, Tomblin, & Zhang, 2002) and writing (Bishop & Clarkson, 2003).

Patterns of Language Impairment

The language deficits in children with DLD may affect all areas of language compared to their peers (Leonard, 2014). Regardless of the heterogeneity of language profiles of children with

DLD, there are some common areas of language that pose challenges to this population. Language skills such as pragmatic, lexical, grammatical, and morphological development are expected to be impaired in children with DLD. Generally, pragmatic skills are better than lexical, grammatical, and morphological language skills. Moreover, grammatical deficits have been described as a hallmark deficit in children with DLD (Leonard, Eyer, Bedore, & Grela, 1997). It should be noted, however, that the majority of available research regarding the patterns of language impairment in children with DLD pertains to the British and American English dialect.

Phonology. Preschool-aged children who exhibit phonological problems are expected to have problems in other areas of language such as issues with morphosyntax and lexical abilities (Paul & Shriberg, 1982; Ruscello, St. Louis, & Mason, 1991). Children with DLD show a delay in acquiring speech segments, which describes the accuracy of each consonant and vowel, relative to TD children (Farwell, 1972). Moreover, children with DLD have been found to produce voicing contrasts (e.g., *coal-goal*) with less accuracy compared to their peers (Catts & Jensen, 1983). In general, the phonological development patterns of children with DLD are largely parallel the developmental patterns of younger children (Farwell, 1972). For example, vowels that cause difficulty for TD children are considered to be an area of difficulty and develop later in children with DLD. Moreover, similar phonological errors that have been reported with high frequency in the speech of younger TD children (two-year-old) have also been reported in children with DLD, such as consonant cluster reduction, liquid gliding, final consonant deletion, and word-initial weak syllable deletion (Hodson & Paden, 1981; Leonard, 1982; Schwartz, Leonard, Folger, & Wilcox, 1980). Notably, some children with DLD may continue to have some phonological process problems into adulthood (Fee, 1995).

Lexicon. Children with DLD exhibit a slower rate of vocabulary growth than their same-age peers (Leonard, 2014). For example, children with DLD show a delayed ability in acquiring their first word, and forming their first word combinations compared to TD children (Leonard, 2014). Children with DLD have also been found to have word-finding difficulties or delayed speed of word retrieval (German & Simon, 1991), poor word learning (Peña, Iglesias, & Lidz, 2001), higher error rates in naming (Sheng & McGregor, 2010), and smaller Mean Length of Utterance (MLU) compared to their age controls (Leonard, 2014). In general, the lexical ability of preschool children with DLD has not been found to differ from younger TD children who match in MLU (Leonard, 2014).

Morphology and syntax. Significant deficits in the use of several grammatical morphemes have been reported for children with DLD compared to their MLU controls (e.g., regular past, third-person singular, the copula and auxiliary *be* form, and the auxiliary *do*, wh-questions, noun plural *-s*, genitive *'s*, infinitive *to*, progressive *-ing*, and articles) (Leonard, 2014; Leonard et al., 1997; Rice, Wexler, & Cleave, 1995; Steckol & Leonard, 1979). In general, omission errors were more frequent in the speech of children with DLD (e.g., omission of copula and auxiliary *be*) (Leonard, 2014). Studies that examined nominative case pronouns (e.g., I, he, she, and they) found that children with DLD show a higher frequency of using accusative words for nominative case pronouns (e.g., him eating apple) than younger TD controls (Loeb & Leonard, 1988). Grammatical morpheme judgment is another area of grammar that has been found to be difficult for children with DLD. For example, children with DLD accepted a higher number of grammatical errors, and showed a slower response time in making grammatical judgments compared to age-controls (Wulfeck & Bates, 1991).

Pragmatics. The accumulated evidence from DLD studies shows that the significant limitation in the language ability of children with DLD may affect their relationship with peers (Bishop, 2000). Indeed, problems with social interaction and peer relationships have been reported consistently in children with DLD (Bishop, 2000). For instance, even though children with DLD interact easily with others, interaction with more than one TD child can be difficult for children with DLD (Bishop, 2000; Craig & Washington, 1993). In a classroom setting, children with DLD are likely to initiate interactions with adults more than TD children, who prefer to interact with other peers rather than adults (Bishop, 2000; Rice, Sell, & Hadley, 1991). Children with DLD also have been found to change the topic more quickly in their conversations (Schelletter, 1990), and produce their utterances with less preparation (Fujiki, Brinton, & Sonnenberg, 1990) than their TD peers.

Theoretical Accounts of DLD

Several theories have been advanced to explain the disproportionate linguistic deficit among English-speaking children with DLD. Theoretical accounting for the patterns of difficulties of children with DLD can be broadly categorized into domain-specific and domain-general theories. Domain-specific theories suggest that children with DLD have limitations in the processing of language or language-related stimuli. Domain-general theories, on the other hand, assume that children with DLD present with deficits in domain-general cognitive processes that support language learning. It must be acknowledged, however, that neither of the two theories can explain DLD clearly, suggesting that DLD is a multifactorial disorder (Bishop, 2003).

Domain-specific theories. Domain-specific theories of DLD assume a specific deficit in some aspect of language processing important to the development of linguistic knowledge. Domain-specific theories address potential phonological or grammatical/ syntactical processing

deficits. A large body of studies have focused on the difficulty with two phonological resources in children with DLD, phonological awareness or the ability to detect and manipulate sounds in the language (e.g., Bradley & Bryant, 1983; Fox & Routh, 1980), and phonological short-term memory, or the brief encoding and retention of a phonological form in short-term memory (e.g., Archibald & Gathercole, 2006). Many studies have closely linked poor phonological awareness and problems with learning to read (Lafrance & Gottardo, 2005). Moreover, much of the evidence for a phonological short-term memory deficit in children with DLD comes from nonword repetition, the ability to repeat a novel phonological form (Graf Estes, Evans, & Else-Quest, 2007). Based on the finding of poor nonword repetition by children with DLD (Archibald & Gathercole, 2006b), it has been suggested that nonword repetition may be a phenotypic marker of children with DLD in school-age children (Bishop, North, & Donlan, 1996). Similar deficits in traditional phonological short-term memory tasks have also been reported for children with DLD, such as digit and word recall (Archibald & Gathercole, 2006b; Henry, Messer, & Nash, 2012).

Other domain-specific theories of DLD implicate a specific deficit in linguistic knowledge, in particular the difficulties with verb-related morphology that characterize children with DLD (Leonard, Miller, & Gerber, 1999). According to this view, it is assumed that children with DLD are unable to formulate specific aspects of grammatical morphology such as the marking of finite verbs for agreement and tense (Rice & Wexler, 1996a; Rice et al., 1995). Such results have led to the development of several theories related to inflection or other grammatical categories related to tense marking. For example, the Extended Optional Infinitive Account or the Agreement and Tense Omission Model of DLD is one of the domain-specific theories that have been examined extensively in English-speaking children with DLD (Rice et al., 1995), and

across other languages (e.g., Rice, Noll, & Grimm, 1997). This theory focuses on a DLD limitation to represent the functional categories related to agreement and inflection, which leads to difficulties in tense marking in children with DLD (Rice et al., 1997).

Domain-general theories. It is widely accepted that the mental and neural systems supporting language learning serve some nonlinguistic functions (Bates, 1994). Evidence that children with DLD perform below age level on many cognitive processing tasks, both verbal and nonverbal, has led to proposals of domain-general deficits in children with DLD (Johnston & Smith, 1989; Miller, Kail, Leonard, & Tomblin, 2001). Domain-general theories assume that children with DLD present with deficits in domain-general cognitive processes impacting language learning. The limitation in domain-general information processing in children with DLD has been discussed in terms of reduced space or capacity (Bishop, 1992), or speed (Kail, 1994). Working memory, or the ability to briefly hold and manipulate information in the current focus of attention, is one domain-general resource that may limit information processing speed or capacity.

Deficits in working memory have been reported consistently in children with DLD (Archibald & Gathercole, 2006b; Ellis Weismer, 1996; Montgomery, 1995, 2000). Working memory is a capacity-limited system imposing processing demands in addition to storage, and is generally assessed by complex memory span paradigms (Engel de Abreu, 2011). According to the working memory model advanced originally by Baddeley and Hitch (1974), working memory incorporates the central executive that coordinates activities within working memory and is associated with attentional control and high-level processing activities. Verbal and visuospatial short-term memory are the two other components described in the working memory model as modality-specific systems. Verbal and visuospatial short-term memory are responsible

for the storage of verbal (the phonological loop) and visuospatial material (the visuospatial sketchpad). The final component of working memory is the episodic buffer (Baddeley, 2000), which is responsible for integrating representations within the subsystem of working memory and across the general cognitive system. Examples of verbal working memory tasks are counting recall and backwards digit recall, in which a participant recalls numbers after counting or reversing the order, respectively. Examples of corresponding visuospatial tasks involve recalling locations or orientations after identifying a different shape or mentally rotating an image, respectively (Alloway, Rajendran, & Archibald, 2009). Examples of verbal short-term memory tasks involve serial recall of words, letters or digits, whereas examples of corresponding visuospatial short-term memory tasks involve the retention of either visual patterns or sequences of movement (Archibald & Gathercole, 2006b; Baddeley, 2000; Conway et al., 2005). Working memory has been associated with complex cognitive activities, such as language comprehension and word decoding (Cain, Oakhill, & Bryant, 2004; Engel de Abreu & Gathercole, 2012). On the other hand, verbal short-term memory has been shown to make an important contribution to new word learning (Majerus, Poncelet, Greffe, & van der Linden, 2006; Masoura & Gathercole, 2005), and vocabulary acquisition (Gathercole, Willis, Emslie, & Baddeley, 1992). It should be noted that some researchers have reported comparable performance between monolingual children with DLD and TD peers on visuospatial short-term and working memory measures (e.g., Archibald & Gathercole, 2006b), suggesting that DLD deficits in immediate memory primarily involve the verbal domain (Archibald & Gathercole, 2006b).

Identifying Children with DLD

Identifying children with DLD and distinguishing between children with and without language impairment have been the concern of numerous studies. Two common methods in

identifying children with DLD have been employed in research: using standardized measures of language ability and using clinical markers that characterize a specific type of disorder (Bishop et al., 1996). The following section illustrates the use of standardized measures of language and clinical markers in identifying children with and without DLD. It should be noted that the majority of available research pertains to English dialects.

The need for the assessment to be based on standardized individually applied measures is highlighted in the criteria for DLD by the American Psychiatric Association (1994) and the research diagnostic criteria of the World Health Organization (*International Classification of Diseases, ICD-10; 1993*). The majority of studies of children with DLD employ standardized tests of oral language ability to identify children with language problems. In addition, speech-language pathologists (SLPs) use standardized tests of language as their primary assessment tools (Leonard, 2014). The clinical process begins with measuring general language ability to determine the presence or absence of a language disorder and to address the language concerns. Standardized tests are also used to determine the severity of children's language problems, identify strengths and weaknesses, design the appropriate intervention, and measure the effectiveness of the intervention. A strength of standardized measures of language is that they allow for comparisons of individual language abilities in relation to normative groups of the same age (Bishop & McDonald, 2009). Assessing language impairment using standardized measures, therefore, requires well-documented normative data for different ages and ranges of language development ability (Bishop, 1997). In many research studies, DLD is diagnosed when two or more composite scores more than 1.25 *SDs* below the standardized mean are achieved based on five norm-referenced standardized tests of receptive and expressive language, in three domains of language (vocabulary, grammar, and narration) (e.g., Bishop, Snowling, Thompson,

& Greenhalgh, 2016; Tomblin, Records, & Zhang, 1996). It should be noted that several researchers have recently agreed that the criteria of a large mismatch between verbal and nonverbal ability is not required for the diagnosis of DLD (Bishop et al., 2017).

Clinical markers that characterize a specific type of disorder have been used in identifying children with DLD (Bishop et al., 1996). Three markers have been proposed: nonword repetition (Bishop et al., 1996; Conti-Ramsden, Botting, & Faragher, 2001), sentence recall (Conti-Ramsden et al., 2001), and finite verb morphology (Bedore & Leonard, 1998; Leonard et al., 1999; Rice & Wexler, 1996b). Nonword repetition, which is the ability to repeat nonsense phonological form, has been shown to have good accuracy for identifying school-age children with DLD across a number of studies (e.g., Conti-Ramsden et al., 2001; Dollaghan & Campbell, 1998; Weismer et al., 2000). In addition, sentence recall, which is the ability to repeat an immediate auditory sentence (Archibald & Joanisse, 2009), is found to act as a clinical marker in children with DLD. That is, children with DLD across numerous studies have shown poor performance in sentence recall compared to TD children (Briscoe, Bishop, & Frazier Norbury, 2001; Conti-Ramsden et al., 2001). It is not surprising then that sentence recall tasks have been included as a primary subtest in many language assessment batteries such as the Clinical Evaluation of Language Fundamentals-4 (CELF-4) (Semel, Wiig, & Secord, 2003). Finally, as mentioned before, children with DLD have been found to perform more poorly than TD children in the use of verb morpheme composites in general (Bedore & Leonard, 1998) and in the use of tense morphology in particular (Rice & Wexler, 1996b). As a result, verb morphemes are considered to hold promise as a clinical marker in identifying children with DLD.

Arabic Language

It has been well established that the language acquired by children affects the manifestation of deficits (Leonard, 2014). Language-specific characteristics must be taken into account in evaluating the challenges children with DLD encounter. It is important for assessment measures to be developed in light of the linguistic characteristics of the language being learned. As mentioned, the majority of available research concerning the assessment of language abilities within a restricted age range pertains to English-speaking children. In general, normative data on language development, which are the basis of standardized assessment measures, are unavailable for many languages (Bishop, 2014). Moreover, studies measuring language ability and impairment across many languages are also unavailable. The present thesis considers the case of Arabic-speaking children.

More than 280 million people speak Arabic as their first language across the Middle East and North Africa (Prochazka, 2006). Arabic is a Semitic language with a rich inflectional morphology that is usually described as nonconcatenative morphology. The case of Arabic is particularly interesting because the features of Arabic morphology are distinct from English, the most widely-studied language (Boudelaa, Pulvermüller, Hauk, Shtyrov, & Marslen-Wilson, 2010). For instance, several words in English have only a single morpheme, such as “car” or “table”, but all surface forms in Arabic are morphologically complex (Boudelaa et al., 2010). That is, all surface forms in Arabic have at least two abstract bound morphemes, which is described as a root and a word pattern (Beeston, 1970; Boudelaa et al., 2010). The root mostly consists of three consonants that represent the lexical meaning (CCC; triliteral root) (Beeston, 1970), and the pattern is primarily composed of vowels inserted between the root consonants. The roots carry a semantic meaning shared to various degrees by the derivative words associated

with the same root (Bakalla, 1979). The phonology and orthography of Arabic are also distinct from Indo-European languages such as English. Arabic has 28 consonants and 6 vowels. Arabic is a transparent language, meaning it has a one-to-one relation between graphemes and phonemes. It should be noted that there is a lack of well-documented normative data for different ages and for the range of linguistic domains for Arabic-speaking children. Only a few studies have focused on the measurement of language ability and impairment in Arabic-speaking children (Wiig & El-Halees, 2000).

Identification of DLD in ELLs

The previous section has addressed the identification of DLD in monolingual children. However, only a few researchers have examined language acquisition in ELLs with DLD (Paradis et al., 2011). As a result, much less is known about the identification of DLD in ELLs (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013). This section will address the challenges and the complexity associated with the assessment of ELLs, the overlap between ELLs and children with DLD in English standardized tests, and the language-processing measures in ELLs as an innovative solution to the assessment of ELLs.

Challenge and Complexity in the Assessment of ELLs

As reviewed in the first section of this chapter, it could take several years for ELLs to achieve similar oral language proficiency as their native-speaker peers, ranging from 3 to 5 years. The incomplete L2 acquisition of ELLs, therefore, can impact their performance on language tests that are norm-referenced with monolingual children such as the standardized tests of grammar and vocabulary. It is widely accepted among researchers that using monolingual norm-referenced testing tools to inform the diagnosis of ELLs may not be clinically useful and may result in biased assessment. Indeed, the overidentification and underidentification of

language disabilities are problems associated with the use of monolingual norm-referenced testing tools with ELLs (Cummins, 2000; Donovan & Cross, 2002; Paradis et al., 2011). Overidentification occurs when the incomplete acquisition of ELLs' L2 is misinterpreted as a language disability, whereas underidentification occurs when the language disabilities in ELLs go undiagnosed assuming that such limitations may be the result of learning two languages (Paradis et al., 2011). Correct identification of DLD in ELLs is important in providing appropriate educational services and intervention. Nevertheless, several studies report that SLPs commonly use English norm-referenced standardized tests in order to assess ELLs' linguistic abilities (Caesar & Kohler, 2007). According to Paradis et al. (2011), a lack of the appropriate testing material for ELLs is the primary reason why English norm-referenced standardized tests are used by SLPs. The limited normative data about the trajectory of ELLs' language acquisition from a diverse first-language background can be one of the main challenges in developing appropriate language assessment measures for ELLs (Bedore & Peña, 2008). It must be recognized that developing such extensive norms is a daunting task. For example, the morphosyntactic structures that are the markers of ELLs with DLD should be developed, and the way that the two languages might interact with and influence each other also should be considered (Bedore & Peña, 2008). As a result, there is a lack of assessment tools that are valid and reliable for the identification of DLD in ELLs.

Using translated tests from the ELLs' majority language is one of the most common suggested solutions to better assess ELLs (e.g., Stow & Dodd, 2003). However, several problems have arisen when researchers have used such a method. For the most part, the norm-referencing of translated tests would be not valid for ELLs, as the trajectory of language acquisition differs across languages (Bedore & Peña, 2008; Paradis et al., 2011). Moreover, as language structure

varies across languages, features that have been described as a hallmark deficit in DLD in the source language may not be similar in the target language, and vice versa (Leonard, 2014).

Another common suggestion regarding the assessment of ELLs is assessing ELLs in their L1 or native language (Wagner, Francis, & Morris, 2005). Such an approach can be essential for the effective assessment of ELLs, as the language difficulty of ELLs with DLD should manifest across both of their languages (Gillam et al., 2013; Paradis et al., 2011). However, assessing ELLs in their L1 can be also problematic as the testing materials and the professionals who will administer such tasks may not be available in one of the child's languages (Chu & Flores, 2011). Moreover, as mentioned before, the phenomena of L1 loss/incomplete acquisition in ELLs also may result in the misdiagnosis of language abilities in ELLs (Anderson, 2012).

Overlap between ELLs and Children with DLD

Considerable overlap has been found between TD ELLs who are in the early stage of developing their L2 and monolingual children with DLD in several linguistic features (Paradis, 2010). For example, both groups tend to have errors in vocabulary choice and grammatical morphemes (Tabors, 2008). In the absence of normative data about the trajectory of ELLs' language acquisition and the lack of the appropriate tools for the identification of DLD in ELLs, differentiating ELLs from children with underlying language impairment can be difficult. Similarities between TD ELLs and monolingual children with DLD have been found in several standardized tests of vocabulary and morphosyntactic knowledge. For example, significant overlap has been found between TD ELLs who are in the early stage of learning English as their L2 and monolingual children with DLD in the accuracy rate and error pattern of various grammatical morphemes, including: third person singular [-s], past tense [-ed], irregular past tense, BE as a copula and auxiliary verb, DO as an auxiliary verb, progressive [-ing],

prepositions in and on, plural [-s], and determiners *a* and *the* (Paradis, 2005; Paradis et al., 2008). Similarly, ELLs' distributed lexical-semantic knowledge across both of their languages can also affect their performance on vocabulary measures. Indeed, TD ELLs often perform comparable to monolingual children with DLD on single language vocabulary measures (Umbel, Pearson, Fernandez, & Oller, 1992; Windsor & Kohnert, 2004). As a result, using standardized tests of language with ELLs may increase the risk of the overidentification of learning disabilities, and such tasks may not accurately distinguish TD ELLs and monolingual children with DLD. The findings reviewed above also raise questions regarding the suitability of using English language tests with ELLs, especially in the first few years of their L2 acquisition, an issue that will be discussed in the next section.

Language-General Verses Language-Specific Measures

It has been suggested that *task effects*, or whether a test probes language-specific abilities or language-general abilities, are an important factor that may account for the considerable differences observed in ELLs' performance (Paradis et al., 2011). *Knowledge-based* measures attempt to examine knowledge that is specific to one given language such as standardized tests of vocabulary or morphosyntactic knowledge. Therefore, performance on such tasks would be affected by ELLs' prior language experience with the target language (Kohnert et al., 2006). The different language-specific experiences of ELLs than their monolingual peers, such as the shorter length of exposure (learn or use) to the language being tested, and the distributed linguistic knowledge across both of the ELLs' languages, can affect their performance on such tasks (Kan & Kohnert, 2005). Therefore, ELLs' poor performance on standardized tests may reflect their lack of experience with the tests' stimuli rather than reflect their actual language ability (Peña et al., 2001).

Language-general measures, on the other hand, attempt to assess either the cognitive interface skills that are shared across the two languages of ELLs (Paradis et al., 2011) or the underlying processing system (Kohnert et al., 2006). Language-general measures are designed to be less affected by ELLs' language-specific knowledge (Paradis et al., 2011). An example of language measures that assess cognitive interface skills is basic word reading tasks that tap perceptual-cognitive skills. Language measures that tap perceptual-cognitive skills that are shared between the two languages of ELLs have been found to be less biased against ELLs (Oller et al., 2007). Indeed, ELLs have been found to perform as well as monolingual children on basic word-reading tasks (Balilah & Archibald, under review; Oller et al., 2007).

Processing-dependent measures are also considered to be language-general measures that tap abilities underlying language learning. Processing-dependent measures include stimuli designed to be equally familiar (or unfamiliar) to all children, regardless of the language they speak; therefore, such tasks are expected to minimize the role of prior language knowledge and experience by directly tapping the abilities underlying language learning (Kohnert et al., 2006). Based on the finding that the nonlinguistic processing domain underpins some of the language learning difficulties in children with DLD, considerable research attention has focused on processing-dependent measures, with the idea that such tasks may be sensitive to the underlying differences between children with DLD and ELLs (Kohnert et al., 2006; Paradis et al., 2013). It should be noted, however, that the majority of ELLs research comparing ELLs with monolingual children with DLD on processing-dependent measures have focused on nonword repetition measures. Given that the phonological forms of nonword repetition tasks are novel to all participants, this task may be less dependent on ELLs' linguistic knowledge. Accumulated evidence from ELL studies, however, shows that the use of nonword repetition tasks does not

completely eliminate the effect of children's experience with the target language (Kohnert et al., 2006). As nonword repetition tasks rely on linguistic stimuli (Kohnert, 2010), performance on such tasks may be influenced by ELLs' previous sublexical phonological knowledge and experience. Indeed, performance of TD ELLs and monolingual English-speakers with DLD did not differ on English nonword repetition tasks (Kohnert et al., 2006; Windsor, Kohnert, Lobitz, & Pham, 2010).

Evidence from research on typically developing individuals, however, indicates that using processing-dependent measures such as working memory tasks may minimize the role of linguistic knowledge and experience in the ELL population. Indeed, equivalent performances have been reported for bilingual children when compared to monolingual peers in working memory measures (Cockcroft, 2016; Engel de Abreu, 2011; Engel de Abreu et al., 2013). To summarize, evidence from several studies indicated that processing-dependent measures, including short-term and working memory measures, are considered to be a less biased form of assessment than knowledge-based measures (Kohnert, 2010; Kohnert et al., 2006; Paradis et al., 2013). Using such tasks with ELLs may provide compelling diagnostic power for distinguishing ELLs from children with underlying language impairment.

The Research in this Thesis

As indicated throughout this chapter, understanding the current level of language knowledge in ELLs can present a challenge. Considerable overlap has been found between TD ELLs and monolingual children with DLD in several knowledge-based measures (Paradis, 2010). As an alternative to knowledge-based testing of ELLs, some researchers have suggested that conducting assessments using processing-dependent measures can provide meaningful information regarding ELLs' language abilities (Kohnert, 2010; Paradis et al., 2013). That is,

processing-dependent measures may pose similar challenges and be equally familiar (or unfamiliar) to all children, regardless of the language they speak (Engel de Abreu & Gathercole, 2012). Therefore, such tasks are expected to minimize the role of prior language knowledge and experience in ELLs by directly tapping abilities underlying language learning (Kohnert et al., 2006).

The primary aim of this thesis is to investigate cognitive and linguistic markers that may differentiate Arabic-speaking ELLs from age-matched monolingual children with and without DLD. Chapter 2 examines the extent to which available language measures are sensitive to developmental change in monolingual Arabic-speaking children. Chapter 3 investigates ELLs' performance on linguistic tests (vocabulary, language, and reading) in Arabic L1 and English L2, and cognitive measures of short-term and working memory and non-verbal intelligence relative to age-level expectations for monolingual speakers. Chapter 3 also analyzes the influence of a number of relevant experiential factors on ELLs' performance. The aim of Chapter 4 was to investigate the utility of knowledge- and processing-dependent measures in distinguishing Arabic-speaking ELLs from age-matched monolingual children with and without DLD. Chapter 5 reviews the main findings of the three studies presented in this thesis and the implication of the findings.

Chapter 2¹

The Measurement of Language Ability and Impairment in Arabic-speaking Children

In the course of language development, most children move from speaking their first words to becoming sophisticated language users in an amazingly short period, with apparent ease, and without specific instruction. Understanding typical language acquisition is important not only for understanding child development, but also for identifying children who are lagging in language growth. In order to measure language development, language measures commonly test a range of abilities beginning with early-acquired forms and progressing to more complex and later developing skills. Of course, the adequacy of such measures in capturing developmental progress in language is dependent on the accuracy with which the tests tap changes in the particular language targeted. To date, the majority of available language assessments have been developed for English speaking children. The present study considers the case of Arabic-speaking children. The available Arabic measures have largely been based on English measures, and in some cases, directly translated. The purpose of this study was to examine the extent to which available Arabic language measures are sensitive to developmental change in a large unselected sample of Arabic-speaking children from Saudi Arabia.

Generally speaking, trajectories of language development follow a similar course and within similar age ranges across individuals speaking the same language (Bishop, 1994; Bishop & Edmundson, 1987). Measuring differences across growth and developmental ranges can help to explain children's language development, including how and when children typically meet

¹ A version of this chapter has been accepted for publication (Balilah, A., & Archibald, L. M. D. (in press). *The Measurement of Language Ability and Impairment in Arabic-speaking Children*. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice*. Springer).

certain milestones and acquire particular language abilities. Well-designed and sensitive language measures can be used to assess children's language abilities and evaluate whether or not the child's language is developing as expected.

Developing language measures requires good normative data on language development covering the range the ages under study (Bishop, 1997). In addition, the design of language measures should consider the characteristics of children identified as having a language impairment, and how the manifestation of such a language impairment may change over the course of development (Bishop, 1994; Bishop & Edmundson, 1987). It follows from this that the psycholinguistic knowledge required to design linguistic measures sensitive to developmental change and impairment is highly specific to the language in which the test is being designed. Although language is a universal human phenomenon, language development and the way in which children learn language varies across languages (e.g., Tardif, 2006). Nevertheless, normative data on language development and data regarding characteristics of language impairment are unavailable for many languages (Bishop, 1997). In the absence of the necessary language-specific normative data, researchers have designed language tests based on, guided by, and, in some cases, directly translated from those for which we have the greatest evidence, English language tests. This is certainly the case for available Arabic language measures. The extent to which these Arabic tests adequately capture change across language development requires careful assessment, which was the focus of the present study.

Across the Middle East and North Africa, more than 280 million people speak Arabic as their first language (Prochazka, 2006). Arabic is a semitic language with a nonconcatenative morphology. Arabic is a root and pattern language with complex interaction between syntax, morphology and phonology. Word roots mostly consist of three consonants that represent the

lexical meaning (CCC; trilateral root; Beeston, 1970), and the pattern is primarily composed of vowels inserted between the root consonants. The roots carry a semantic meaning shared to various degrees by the derivative words associated with the same root (Bakalla, 1979). For example, the root [k-t-b] carries the semantic meaning of *writing* so that [katip] means *writer*, [kitap] means *book*, and [maktabh] means *library*. The derivatives of the simple root by the variation of the vowels and the addition of the affixes (prefix, suffixes, and infixes) create different meanings. Moreover, the verbal inflection system of Arabic is relatively rich. Verbs are morphologically inflected for tense, and mood, and the verb should agree with the subject for aspects of person (first, second, and third), number (singular, dual, and plural), and gender (feminine and masculine) (Bakalla, 1979). As an example of the Arabic inflection, the base verb [katab] means *he wrote*, and inflectional endings indicate changes in gender [katabt] (*she wrote*), in number [katabu] (*they wrote*), and in tense [yaktub] (*he writes*).

Another characteristic of Arabic is the phenomenon of diglossia, or the use of different forms of a language in different situations (Ferguson, 1959). Two varieties of Arabic are used within the Arabic community simultaneously under different conditions: colloquial Arabic and Modern Standard Arabic (Daher, 1998). Colloquial Arabic can be described as the oral form of the language. There are diverse colloquial Arabic dialects across Arabic countries, and most countries have their own dialect (Aljenaie, 2001; Beeston, 1970; Bulos, 1965). On the other hand, the Modern Standard Arabic is the written form of the language used in formal institutional teaching, public media, and formal texts and newspapers. Arabic children speak colloquial Arabic in their daily communication as their mother tongues (Thompson-Panos & Thomas-Ruzic, 1983). Arabic children are also exposed simultaneously to Modern Standard Arabic through television programs and schooling (Al-Tamimi, 2011; Ibrahim, 1989).

There is a lack of well-documented normative data for different ages and for the range of ability in language development in Arabic-speaking children. Only a few studies have focused on the measurement of language ability and impairment in Arabic-speaking children (Wiig & El-Halees, 2000). In general, the acquisition of the phonology and morphology of Arabic has received more attention than other aspects of the language, such as the acquisition of semantics and pragmatics (Omar, 1973). One of the main challenges in developing Arabic language assessment measures is related to the diglossic nature of Arabic and colloquial variation (Al-Tamimi, 2011). As Arabic-speaking children are exposed naturally to the colloquial dialect, and they use it primarily in their daily oral communication, language assessment measures should address the acquisition of the colloquial dialect (Al-Tamimi, 2011). The varying degrees of similarity and difference both across and within Arabic colloquial dialects, with some even being mutually unintelligible, can pose a challenge for the assessment of Arabic (Al-Tamimi, 2011; Newman, 2002; Shahin, 2010). For example, in the western region of Saudi Arabia (Mecca, Medina, and Jeddah), Urban Hijazi is the dialect spoken in the major cities (Sieny, 1978). The high degree of contact with other languages and dialects and the mixture of ethnic groups residing in this region affect the Hijazi dialect (Al-Essa, 2006; Alahmadi, 2015). As a result, lexical and phonological variations characterize the Urban Hijazi dialect (Al-Essa, 2006; Alahmadi, 2015). There is a great need for well-documented normative data for different ages and ranges of language development ability in Arabic-speaking children in their native colloquial Arabic dialect (Shahin, 2010; Wiig & El-Halees, 2000). It must be recognized, however, that developing such extensive norms is a daunting task, which has partly fueled the interest in being guided by more extensively researched English language tests in designing Arabic measures.

In general, the majority of language assessment tests used for Arabic are translated and adapted from English. The vocabulary measure, the *Arabic Receptive-Expressive Vocabulary Test* was based on English language tests such as the *Test of Word Knowledge* (Wiig & Secord, 1992), and the *Clinical Evaluation of Language Fundamentals* (Semel, Wiig, & Secord, 1989b). Wiig and El-Halees (2000), the authors of The Arabic Receptive-Expressive Vocabulary Test, did not indicate how the task had been developed or how the Arabic words included in the task were chosen. As part of the test's development, a sample of 117 Arabic-speaking children ages 3 to 13 years in Palestine completed the test to develop criterion raw scores. The colloquial dialects represented in this sample were Palestinian dialect. The Arabic Receptive-Expressive Vocabulary Test was designed with fixed start and stop points for different age groups. For example, all 7-year-olds start with item 20 (receiving credit for previous items) and end with item 45, whereas 8-year-olds start with item 25 and end with item 50.

The *Arabic Language Screening Test* is another test developed by (El-Halees & Wiig, 1999a). The aim of the test is to evaluate children's language development, and identify children at risk for language disorders, by comprehensively sampling expressive and receptive Arabic language skills. The *Arabic Language Screening Test* (El-Halees & Wiig, 1999a) was based on English language tests such as the *Clinical Evaluation of Language Fundamentals* (Semel et al., 1989b; Semel, Wiig, & Secord, 1995; Wiig, Secord, & Semel, 1992), the *Clinical Evaluation of Language Fundamentals Screening Test* (Semel, Wiig, & Secord, 1989a), and the *Test of Word Knowledge* (Wiig & Secord, 1992). The authors of the Arabic Language Screening Test indicated that the Arabic task used stimuli that reflected Arabic culture and societal values, and that the stimuli were never translated literally or idiomatically from English. In addition, the items included in the tasks were designed in light of aspects of Arabic phonology, morphology,

syntax, and semantics, and all the illustrations in the task represented Arabic cultural experiences and values. In addition, the authors indicated that three Arabic-speaking speech and language pathologists had reviewed all the items that were included in the task. The authors explained, however, that the design of the Arabic task drew heavily on the findings of English-speaking children with language impairment (Wiig & El-Halees, 2000). That is, the authors assumed that the underlying aspect of language behaviors that differentiate English-speaking children with language impairment from their typically developing peers could be applied to Arabic-speaking children with language impairment. As part of the test's development, a normative sample of 450 school-aged Arabic-speaking children between 6 to 12 years in Jordan and Palestine completed the test. The colloquial dialects represented in this sample included Palestinian and Jordanian dialects. In the Arabic Language Screening Test, all children complete all items regardless of age. Notably, all the tasks developed by Wiig and El-Halees (2000) used Modern Standard Arabic language, and the examiners were required to modify items to match each child's spoken dialect.

Three additional Arabic tasks have been developed by Shaalan (2010) including a phonological short-term memory task (*Nonword Repetition Task*), a comprehensive language test (*Arabic Language Test*), and a receptive vocabulary test (*Arabic Picture Vocabulary Test*). Regarding the Nonword Repetition Task, Shaalan indicated that the stimuli of the task were designed to reflect the phonotactic and morphological rules of Arabic. The stimuli were designed to imitate several triconsonantal roots of Arabic (containing a sequence of three consonants) that do not appear in the Arabic lexicon. The task controls for phonological complexity included stimuli with different types of clusters (Shaalan, 2010). The range of syllable length for all stimuli included in the task was two to three-syllable. Shaalan reported data for a group of 33

children ages 5 to 7 years who completed the Nonword Repetition Task. The Arabic Language Test included three subtests: *Sentence Comprehension*, *Expressive Language*, and *Sentence Repetition*. Shalaan indicated that the stimuli of the Sentence Comprehension subtest were designed to reflect Arabic syntactic, morphological, and morphosyntactic structures; however, several of the items and illustrations in the subtest were based on English tests such as the *Clinical Evaluation of Language Fundamentals* (CELF-3, Semel et al., 1995). Shaalan reported data for a group of 114 children ages 4 to 9 years who completed the Sentence Comprehension subtest. For the Expressive Language subtest, the subtest used variants of Arabic morphosyntactic structure that are commonly used by Arabic-speaking children. The linguistic structures of the items included in the subtest were chosen based on the experiences of the author and other clinicians who worked with Arabic children. In addition, Shalaan indicated that the linguistic structures of the items used in the subtest benefited from the available evidence of three previous Arabic studies (Abdalla, 2002; Al-Akeel, 1998; Aljenaie, 2001). Various items and illustrations in the subtest were based on English tests such as the *Clinical Evaluation of Language Fundamentals* (CELF-Semel et al., 1995), and the *Preschool Language Scale* (PLS-4, Zimmerman, Steiner, & Pond, 1992). For the Sentence Repetition subtest, the linguistic structures used in the sentences were closely similar to the linguistic structures used in the Sentence Comprehension subtest. Shaalan reported data for a group of 112 children ages 4 to 9 years who completed the Expressive Language and Sentence Repetition subtests. In regards to the Arabic Picture Vocabulary Test, Shaalan did not indicate how the Arabic words that were included in the task had been chosen. In addition, some of the illustrations that were used in the subtest were based on English tests such as the *British Picture Vocabulary Scale* (BPVT-2, Dunn, Dunn, Whetton, & Burley, 1997). Shaalan reported data for a group of 107 children ages 4

to 9 years who completed the Arabic Picture Vocabulary Test. In all of the tests that were developed by Shaalan, the child completes all items regardless of age, and all the tasks tested used the Gulf Qatar dialect because the test was administered to Gulf Arabic-speaking children from Qatar.

It is clear that the design of Arabic language tests has drawn heavily on the available evidence regarding typical and atypical language development in English-speaking children (Wiig & El-Halees, 2000). Given the lack of normative data in typical and atypical language development in children speaking Arabic colloquial dialects (Al-Tamimi, 2011), current language assessment measures may not utilize the crucial aspects of Arabic necessary to capture developmental progression across ages. The present study focused on assessing the sensitivity of the available Arabic language measures to developmental change in 6 to 9 year-olds, a stage perhaps reflecting slower language change relative to the preschool years (Pence & Justice, 2008) but still an important period of growth in language complexity (Nippold, 1998). Another important factor to consider in examining Arabic language development is potential sex differences. In Arabic countries such as Saudi Arabia, a number of cultural practices are related to an individual's identification as male or female. In Saudi Arabia, male public schools provide more diverse activities than female public schools. For example, male public schools provide physical education, whereas female public schools do not (Khalaf et al., 2013). Given these cultural practices, potential sex-based language differences should be evaluated, as was the case in the current work.

In the present study, a large unselected sample of school age Arabic-speaking children in Saudi Arabia completed the currently available Arabic language measures. One purpose of the study was to investigate one psychometric property of the included language tests, sensitivity to

developmental change. Performance differences across the full 6-9 year age band studied would indicate that the existing measures capture developmental language growth whereas a lack of such age-related differences would suggest that the tests fail to capture crucial aspects of Arabic language development. A second aim of the study was to examine the consistency with which individuals are identified with low language skills across tests. Low performance across a number of measures would increase the confidence with which individual participants could be considered to have a language impairment. The final goal of the study was to explore the possible sex differences on the measures. Sex differences on language performance would signal the need for further investigation of this factor in language acquisition and consideration in test design.

Method

Participants

The study took an epidemiological approach by inviting all monolingual Arabic speakers from 6 to 9 years of age in 10 schools (5 male schools, 2 of which were public; 5 female, all public) in Saudi Arabia (Jeddah) to participate. A total of 421 school-age children (158 males) participated who were, on average, 7.94 years of age ($SD = 1.10$), with similar age ranges for both the male ($M = 7.85$, $SD = 1.16$, range = 6.16-9.92) and female participants ($M=7.99$, $SD = 1.06$, range = 6.33-9.83). Schools were located in different demographic regions in Jeddah, representing a considerable socioeconomic range. In order to account for variability in the data due to socioeconomic status, a proxy measure of socioeconomic status was employed based on parent report of the highest level of education achieved by the child's mother. The descriptors included some high school, completed high school, some college, completed college, some university, and completed university. Responses were transposed to a 3-point scale with 1

corresponding to some/completed high school, 2 to some/completed college, and 3 to some/completed university. This question was optional, and was completed by 399 of the parents in the study. In addition, according to parental reports, all students spoke Arabic as their first and only language. The dialect spoken in Jeddah is the Urban Hijazi dialect, which is characterized by the previously mentioned lexical and phonological variations. Dialectical variations commonly observed in Hijazi were employed in testing to match that spoken by the child, and scored as accurate in child responses.

Materials and Procedure

A trained native Arabic speaker tested each child individually in a quiet room in the child's school. Data were collected in three sessions of approximately 40 minutes with each session occurring about one week apart. Children completed a battery of assessment measures including tests of language, vocabulary, phonological short-term memory, and other tasks not reported here. The tests were administered in a fixed order so that session one involved the administration of the *Arabic Receptive-Expressive Vocabulary Test* (El-Halees & Wiig, 1999b), *Nonword Repetition Task* (Shaalán, 2010), and the *Arabic Language Screening Test* (El-Halees & Wiig, 1999a), session two, the *Arabic Language Test* (Shaalán, 2010), and session three, the *Arabic Picture Vocabulary Test* (Shaalán, 2010) and other tasks not reported here. All research assistants underwent a rigorous training procedure. After review of test administration, the research assistant administered all tests to a child not involved in the study in the presence of the first author who independently scored the child's performance. After completion, scored records were reviewed and any discrepancies discussed. If there were more than 3 discrepancies, this procedure was repeated until there were fewer than 3 scoring discrepancies across the entire test battery.

The Arabic Receptive-Expressive Vocabulary Test (AREVT). Both the expressive and receptive subtests of the AREVT were administered to all children. In the receptive subtest, the children pointed to a picture corresponding to a given spoken word from a choice of four. In the expressive subtest, children named or described a picture with a single word or phrase. Raw scores corresponded to the number of correct responses with maximum possible score for each subtest dependent on the child's age (6;0 to 6;11: $n=40$; 7;0 to 7;11: $n=45$; 8;0 to 8;1: $n=50$; 9;0 to 9;11: $n=55$). Total test scores were the sum of correct raw scores for each subtest. Subtest and total raw scores were converted to z -scores within each year band in order to allow comparison across age groups.

The Arabic Picture Vocabulary Test (APVT). In the APVT, the children pointed to a picture corresponding to a given spoken word from a choice of four. Each response was scored as correct or incorrect with a maximum possible score of 132. High test retest reliability has been reported for the APVT = .97 (Shaalán, 2014).

The Arabic Language Screening Test (ALST) (school-age). The ALST involved tests of verbal and nonverbal abilities. Tests of verbal abilities included 6 items each assessing nouns and verbs, adjectives, morphology, understanding sentences, forming sentences, remembering instructions, and repeating sentences. For Nouns and Verbs, children named the object, person, or activity pictured. For Adjectives, children were first required to point to a picture that illustrated a spoken sentence, and then give the opposite word. For Morphology, the children were given a sentence and asked to generate spoken sentences in reference to a picture cue. In Understanding Sentences, the children pointed to a picture from a choice of three corresponding to a spoken sentence. In Forming Sentences, the children formulated a sentence about the visual stimuli presented using target words or phrases. In Remembering Instructions, the children

pointed to pictures in response to oral directions. In Repeating Sentences, the children repeated sentences presented by the examiner. The majority of subtests were scored with 1 point for each correct response with the exception of Formulating Sentences (2=correct sentence; 1=few errors; 0=nonsense or unrelated sentence, or no response) and Repeating Sentences (2=correct; 1=1-2 errors of omission, addition, or substitution; 0=3 or more errors, or no response). For Adjectives, 1 point was given each for correctly pointing to the picture and correctly naming the opposite word. The highest possible overall raw score in the Verbal Abilities section was 60.

The Non Verbal Abilities subtest involved five short sections consisting of 5 or 6 items and requiring verbal responses. In the Missing Part activity, the children pointed to the correct picture from a choice of four to correspond with the missing object in a target picture; then the children named the object that illustrated the missing part. In Matching Words/Sentences, the children pointed to the word/sentence that was orthographically identical to the target word/sentence. In Classification by Meaning, the children chose the three pictures that were related from a choice of five. In Classification by Group Membership, the children chose the two pictures that were related from five and described the relationship. In Arranging a Story, the children ordered the four given pictures to form a logical story. The majority of the subtests were scored with 1 point for each correct response, with the exception of the Missing Part activity (2=correct pointing and naming; 1=correct pointing or correct naming; 0= incorrect pointing and naming, or no response) and Classification by Group Membership (2=correct pointing to the related pictures and correct describing of the relationship among the group members; 1=correct pointing or describing; 0=incorrect pointing and describing, or no response). The highest possible overall raw score in the Non Verbal Abilities section was 40 with a maximum possible score of 100 for the overall Arabic Language Screening Test.

The Arabic Language Test (ALT). The three subtests of the ALT were administered to all children, and each subtest was divided into a first and second section (Section A and B). In the *Sentence Comprehension* subtest, the children pointed to a picture that corresponded to the spoken sentence from a choice of three or four. In the *Expressive Language* subtest, the children were given a sentence and asked to generate a spoken word or phrase in reference to a picture cue. The *Sentence Repetition* subtest required immediate repetition of a presented audio recording of a sentence spoken by a native, adult male Arabic speaker. The Sentence Comprehension and Expressive Language tests were organized into two sections corresponding to early development or more advanced morphosyntactic structures (see Shaalan, 2010). Each response was scored as correct or incorrect with a maximum possible score of 40 for the Sentence Comprehension subtest (Section A: $n=22$; Section B: $n=18$), and 68 for the Expressive Language subtest (Section A: $n=24$; Section B: $n=44$). The 41 items of the Sentence Repetition subtest were scored on a 4-point scale (3=correct; 2=1 error; 1=2-3 errors; 0=4 or more errors or no response) for a maximum possible score of 123 (Section A: $n=18$; Section B: $n=23$). High test retest reliability has been reported for the three subtests of the ALT (the Sentence Comprehension Test = .95; the Expressive Language Test = .95, and the Sentence Repetition Test = .97) (Shaalan, 2014).

Nonword Repetition Task (NWR). In the NWR Task, children repeated nonwords of a presented audio recording of nonwords spoken by a native, adult male Arabic speaker. The stimuli selected were taken from the task employed by Shalaan (2010). Children were given one opportunity to imitate 48 nonwords varying in length (two to three syllables) and cluster type (no cluster, medial cluster, final cluster, and medial and final clusters). Each response was scored as correct or incorrect for a maximum possible score of 48.

Results

Preliminary analysis

As a first step, data were examined for distribution and evidence of floor or ceiling effects. Table 1 presents descriptive statistics for raw scores for all participants on all language tasks except the Arabic Receptive-Expressive Vocabulary Test (for which the maximum score differed across age groups and z -scores are shown). There were two subtests for which the maximum test score fell within a 1 *SD* band around the group mean: Sentence Comprehension-A and Sentence Repetition-A. This pattern was considered to reflect a ceiling effect. As a result, no further analysis was conducted on these two subtests. No corresponding floor effects were observed.

Next, the data were examined based on the four age bands of 6;0-6;11, 7;0-7;11, 8;0-8;11, and 9;0-9;11. These data are presented in Appendix A. The data were reviewed again for ceiling and floor effects. No additional ceiling or floor effects were observed within the age groups including for the Arabic Receptive-Expressive Vocabulary Test. All remaining analyses for the Arabic Receptive-Expressive Vocabulary Test were completed on the z -scores.

Table 1

Descriptive statistics for the Arabic Receptive-Expressive Vocabulary Test z -scores, and raw scores for all remaining language tasks

Measure	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>Skew</i>	<i>Kurt^b</i>	<i>Test's max</i>
Arabic Receptive- Expressive Vocabulary Test:						
Receptive	.00	1.00	-2.22 - 1.86	-.11	-.98	n/a
Expressive	.00	1.00	-2.24 - 2.53	.07	-.65	
Total score	.00	1.00	-2.09 - 2.22	.00	.12	
Arabic Picture Vocabulary Test	101.39	11.68	58 - 123	-.69	.43	132
Arabic Language Screening Test:						
Verbal abilities	38.48	7.88	14 - 56	-.37	-.38	60
Nonverbal abilities	23.92	5.60	4 - 36	-.33	-.12	40
Total score	62.40	12.21	22 - 89	-.42	-.37	100

Arabic Language Test:

Sentence Comprehension (A)	19.97 ^a	2.31	7 - 22	-2.63	9.33	22
Sentence Comprehension (B)	13.55	2.75	4 - 18	-.74	.71	18
Expressive Language (A)	18.33	3.20	7 - 24	-.48	.04	24
Expressive Language (B)	35.24	5.41	19 - 44	-.67	.20	44
Sentence Repetition (A)	51.30 ^a	4.90	22 - 54	-2.80	9.00	54
Sentence Repetition (B)	46.03	10.66	15 - 69	-.39	-.42	69
Total score	184.43	19.96	115 - 230	-.59	.10	231
Nonword Repetition task	38.65	7.12	5 - 48	-1.36	2.93	48

a - Subtests with ceiling effects; b - Kurtosis

Sensitivity to developmental changes and sex differences

In the next set of analyses, we investigated sensitivity to developmental changes across our four age bands for language tests without ceiling effects. To do this, multivariate analysis of variances (MANOVAs) were conducted separately as a function of age group (6-9) and sex (male, female) on the scores of each language test: (1) the receptive and expressive subtests of the Arabic Receptive-Expressive Vocabulary Test, (2) the verbal and nonverbal subtests of the Arabic Language Screening Test, and (3) the four subtests of the Arabic Language Test. As well, separate corresponding ANOVAs with Bonferroni correction were completed on the raw scores of the Arabic Picture Vocabulary Test and Nonword Repetition tasks. For all three MANOVAs, the Hotelling's T were significant for age, $F > 7.9$, $p < .001$ (all cases), but not for the interaction between age and sex, $F < 1.55$, $p > .05$ (all cases). The main effect of sex was also significant in each case due to higher scores for the males than females, $F > 3.1$, $p < .05$. Results of the univariate comparisons are described below.

The MANOVA performed on the z -scores of the Arabic Receptive-Expressive Vocabulary Test revealed significant univariate effects for the receptive subtest on age, $F(3,413) = 752.12$, $p < 0.05$, $\eta^2_p = .845$, and sex, $F(1,413) = 4.60$, $p < 0.05$, $\eta^2_p = .011$, but not the interaction, $F(3,413) = .038$, $p > 0.05$, and for the expressive subtest on age, $F(3,413) = 320.48$,

$p < 0.05$, $\eta^2_p = .700$. The effect of sex on the expressive subtest and the interaction between age and sex were marginal (sex: $F(1,413) = 3.68$, $p = .056$; interaction: $F(3,413) = 2.62$, $p = .05$).

Pairwise comparisons revealed significant increases with each incremental increase in age band for both subtests, which is clearly visible in the z-score plot presented in Figure 1 ($p < .001$, all cases; see also Appendix A). As well, significantly higher scores for males than females were observed for the receptive subtest (male: $M = .053$, $SE = .031$; female: $M = -.030$, $SE = .024$).

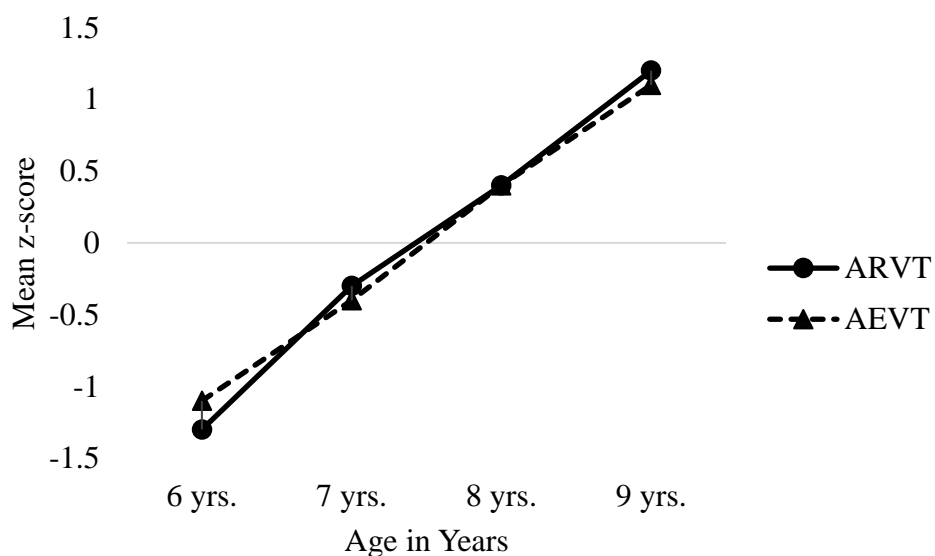


Figure 1. Mean z-scores as a function of age for the receptive and expressive subtests of the Arabic Receptive-Expressive Vocabulary Test showing significant performance growth with each increase in age band.

The MANOVA performed on the two subtests of the Arabic Language Screening Test revealed significant univariate effects for the verbal abilities subtest on age only, $F(3,413) = 19.32$, $p < 0.05$, $\eta^2_p = .123$. Remaining effects were not significant (sex: $F(1,413) = .509$, $p > 0.05$; interaction: $F(3,413) = .613$, $p > 0.05$). For the nonverbal abilities subtest, the effects of age, $F(3,413) = 17.39$, $p < 0.05$, $\eta^2_p = .112$, and sex, $F(1,413) = 6.58$, $p < 0.05$, $\eta^2_p = .016$, were significant, but the interaction was not, $F(3,413) = .566$, $p > 0.05$. As shown in Figure 2 displaying the z-scores for each age band, pairwise comparisons for both subtests revealed

significant increases between the 6 and the 7 year olds ($p < .001$, both cases) but not the 7 and 8 or 8 and 9 year olds ($p > .05$, all cases; see Appendix A). On the nonverbal abilities subtest, males scored significantly higher than females (male: $M = 24.78$, $SE = .420$; female: $M = 23.41$, $SE = .325$).

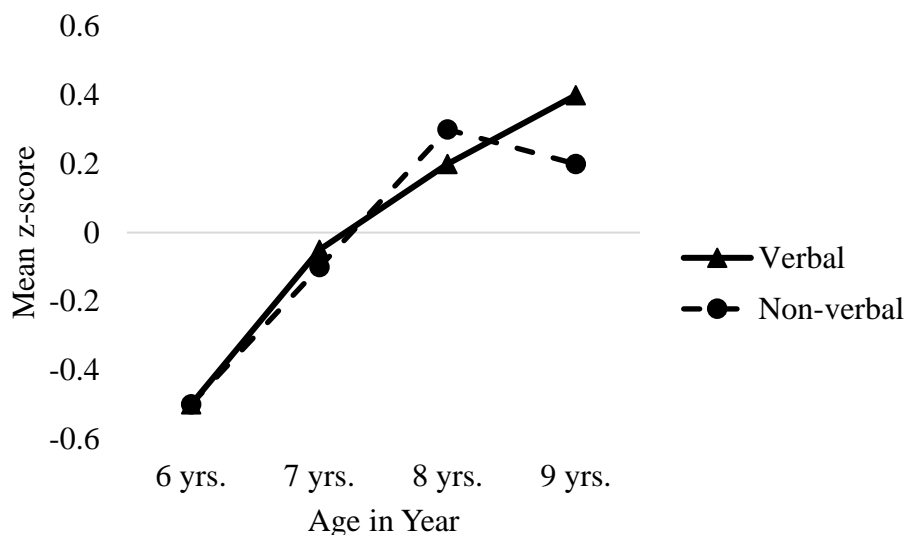


Figure 2. Mean z-scores as a function of age for the verbal and nonverbal subtests of the Arabic Language Screening Test showing significant performance growth between the 6 and 7 year olds only.

The MANOVA performed on the four subtests of the Arabic Language Test revealed significant univariate effects for the four subtests on age, $F > 9.8$, $p < 0.05$, $\eta^2_p > .065$ (all cases). For the sentence repetition-B subtest, the effect of sex, $F(1,413) = 9.39$, $p < 0.05$, $\eta^2_p = .022$, and the interaction between age and sex were significant, $F(3,413) = 3.14$, $p < 0.05$, $\eta^2_p = .022$). All remaining effects of sex and interactions were not significant, $F < 1.8$, $p > 0.05$ (all cases). For the main effects of age as shown in Figure 3, pairwise comparisons revealed significant increases between the 6 and the 7 year olds ($p < .001$, all cases) but not the 7 and 8 or 8 and 9 years old ($p > .05$, all cases; see also Appendix A) for all subtests except sentence comprehension-B (for which no significant effects were found, $p > .05$, all cases). As well, the significantly higher

scores on sentence repetition-B were observed for males compared to females (male: $M = 47.93$, $SE = .782$; female: $M = 44.89$, $SE = .606$). The significant effects of sex and age for the sentence repetition-B subtest were due to higher scores for the males than females at 6 and 7 but not 8 and 9 years of age as shown in Figure 4 ($p < .05$, all significant cases).

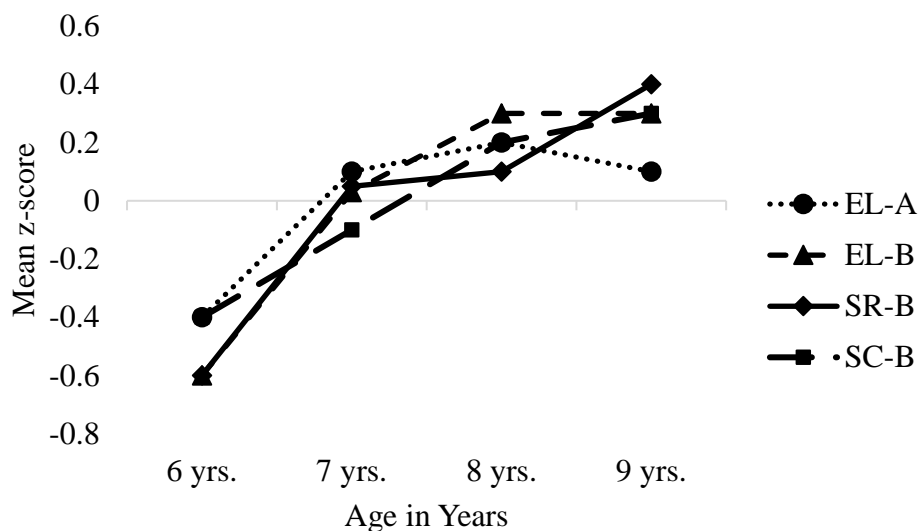


Figure 3. Mean z -scores as a function of age for the four subtests of the Arabic Language Test showing significant performance growth between 6 and 7-year-old only for the expressive language but not sentence repetition subtests.

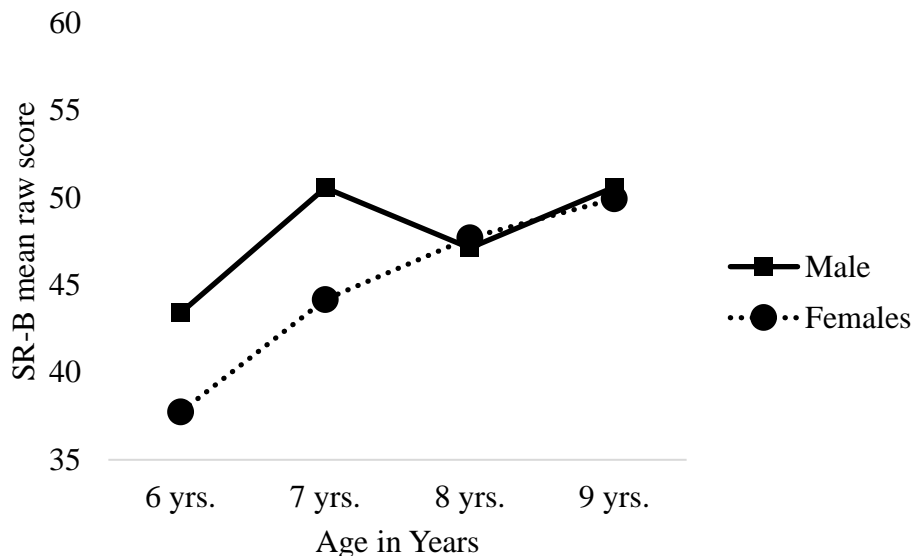


Figure 4. Mean raw scores of sentence repetition-B subtests of the Arabic Language Test for male and female participants as a function of age, showing significantly higher scores for the males than females at 6 and 7 but not at 8 and 9 years of age.

The ANOVA performed on the raw scores of the Arabic Picture Vocabulary Test revealed significant effects of age, $F(3,413) = 17.67$, $p < 0.05$, $\eta^2_p = .114$, and sex, $F(1,413) = 12.96$, $p < 0.05$, $\eta^2_p = .030$, but not the interaction, $F(3,413) = .96$, $p > 0.05$. Pairwise comparisons revealed significant increases between the 6 and the 7 year olds ($p < .05$) but not the 7 and 8 or 8 and 9 year olds ($p > .05$; see Appendix A). As well, significantly higher scores were observed for males than females (male: $M = 103.817$, $SE = .863$; female: $M = 99.88$, $SE = .669$).

The ANOVA performed on the raw scores of Nonword Repetition revealed significant effects of sex, $F(1,413) = 7.344$, $p < 0.05$, $\eta^2_p = .017$, but not age, $F(3,413) = 1.939$, $p > 0.05$, or the interaction, $F(3,413) = .520$, $p > 0.05$. Pairwise comparison revealed significantly higher scores for females than males (female: $M = 39.373$, $SE = .435$; male: $M = 37.448$, $SE = .562$).

To summarize the results for the sensitivity to developmental change in language measures, age effects were observed for all measures except the nonword repetition task. In all cases of significant age effects, performance differences were observed between the 6 and 7 year

olds with one exception: On the sentence repetition-B task of the Arabic Language Test, age increases were modified by sex such that 6 and 7 but not 8 and 9 year old males scored significantly higher than females. Only the Arabic Receptive-Expressive Vocabulary Test revealed significant differences between the remaining incremental age groups. Sex differences in favor of the male participants were observed for all subtests except the verbal abilities subtest of the Arabic Language Screening Test, and the nonword repetition task for which an advantage for females was found. It should be noted that in corresponding MANOVAs and ANCOVAs with maternal education as a covariate, the same pattern of results was observed for all subtests with one exception: Sex differences in favor of male participants in the receptive subtest of the Arabic Receptive-Expressive Vocabulary Test were no longer significant after statistically controlling for mothers' level of education.

Interrelations

To explore the degree to which the language subtests evaluate different or somewhat similar language skills, a correlation matrix was computed on the language subtests with the greatest sensitivity to developmental change in language: The receptive and expressive subtests of the Arabic Receptive-Expressive Vocabulary Test, the verbal and nonverbal subtests of the Arabic Language Screening Test, the four subtests of the Arabic Language Test, and the Arabic Picture Vocabulary test using the z -scores of the Arabic Receptive-Expressive Vocabulary Test, and raw scores for all remaining language subtests. Zero-order correlations are displayed in the lower triangle in Table 2. The intercorrelations between all language measures were moderate to small in magnitude, with r s ranging from .28 to .48 ($p < .001$, all cases). The within-test intercorrelations between subtests were large (r s ranging from 0.84 to 0.61, $p < .001$, all cases) for all measures analyzed. A partial correlation, calculated while controlling for age in months,

9.SR_B_ALT	.32	.35	.38	.42	.46	.39	.28	.60	-----
------------	-----	-----	-----	-----	-----	-----	-----	-----	-------

All values $p < .001$; values in bold $p < .05$.

Sensitivity to identified children with impairment

Of further interest was the degree to which individuals exhibited consistently low performance across tests. For this analysis, we chose the subtests with the greatest sensitivity to developmental change in language including the receptive and expressive subtests of the Arabic Receptive-Expressive Vocabulary Test (AREVT), the verbal and nonverbal subtests of the Arabic Language Screening Test, the four subtests of the Arabic Language Test, and the Arabic Picture Vocabulary test. Standard scores were calculated for each year group after replacing individual outliers (scores falling ± 3.5 *SD* band around the group mean) with the value of the next highest/lowest non-outlier. Four outliers were identified: One on the receptive AREVT subtest at age 9, one on the expressive AREVT subtest at age 8, and two on the Arabic Picture Vocabulary Test at age 7. Composite standard scores were calculated by averaging standard scores across relevant subtests for the Arabic Receptive-Expressive Vocabulary Test, the Arabic Language Screening Test, and the Arabic Language Test. Children obtaining a standard score below 86 were considered indicative of a deficit. A total of 148 participants scored below the cutoff on one or more language measures. Of these, 62% scored below the cutoff on one measure only, 28% on two measures, 7% on three measures, and 3% on all four measures. Table 3 presents the number of participants who scored below the cutoff on one or two language measures.

Table 3

The number of participants who scored below the cutoff on one or two language measures

	AREVT	APVT	ALST	ALT
AREVT	30	4	6	7
APVT		14		
ALST		5	22	
ALT		5	14	25

Discussion

The primary aim of this study was to assess the sensitivity to developmental change and sex differences of available Arabic tests when applied to school-aged Arabic-speaking children. The study also examined the degree to which individuals exhibit consistently low performance across those language measures. This is the first investigation of these Arabic language measures with Arabic-speaking children from Saudi Arabia. Age effects were observed for all measures except the Nonword Repetition Task. This study shows that the available Arabic language measures are sensitive to developmental change in younger children only between the ages of 6 and 7. Only the Arabic Receptive-Expressive Vocabulary Test revealed significant performance growth with each increase in age band. Sex differences in favour of male participants were observed for several of the language subtests.

Among the language tests, the *Arabic Receptive-Expressive Vocabulary Test* (El-Halees & Wiig, 1999b) showed the greatest sensitivity to developmental change among school-aged Arabic-speaking children. As the authors indicated, all the stimuli and illustrations in the tasks represent Arabic cultural experiences and values. The authors, however, did not indicate how the

task had been developed or how the Arabic words that were included in the task had been chosen. For example, it is not clear if the vocabulary used in the task was based on a systematic investigation of the vocabulary acquisition of Arabic-speaking children. The words included in the task should be representative of universal knowledge that is considered important for the specific population and age range of children under study (Nagy & Herman, 1987). In addition, the items selected for the vocabulary task should be guided by the nature of the vocabulary, such as the frequency with which the word appears in print (Lorge & Chall, 1963). Another point that should be highlighted is the scoring method that is used by the Arabic Receptive-Expressive Vocabulary Test. The task was designed with fixed start and stop points for different age groups, and children in certain age groups received credit for previous items. This scoring method might have influenced the test results, as the Arabic Receptive-Expressive Vocabulary Test was the only test that showed a significant performance growth with each increase in age band.

The non-significant increases between the 7 and 8 or 8 and 9 year-olds on the two language tasks—the *Arabic Language Test* and the *Arabic Picture Vocabulary Test* (Shalan, 2010)—might demonstrate that these language measures became less challenging with age. Even though the authors of the tasks indicated that the items used in the tasks are culturally appropriate for Arabic-speaking children, it may be the case that the tasks need to include more complex language skills and a broader range of language abilities to capture the level of language development in older Arabic children. For example, Abdalla and Crago (2008) found that Arabic-speaking children with language impairment have a specific difficulty with tense and subject-verb agreement forms. However, The Expressive Language (B) subtest that aims to assess more advanced morphosyntactic structures included only 3 present verb markers, and 3 past verb markers. In addition, all the responses on the *Arabic Language Test* asked children to

generate a spoken answer of two words in length, or, more often the case, only one word. Providing more comprehensive assessment by requiring complex language responses might better capture the level of language development in older Arabic children.

The *Arabic Language Screening Test* (El-Halees & Wiig, 1999a) also did not show increases between the 7 and 8 or 8 and 9 year-olds. The school-aged Arabic Language Screening Test has been designed for children between 6 to 12 years. All children are required to complete all items in this test regardless of age, and the task included 71 items completed in 10-15 minutes by typically developing children. Nevertheless, the verbal abilities subtest of the Arabic Language Screening Test included 6 items only, each assessing nouns and verbs, adjectives, morphology, understanding sentences, forming sentences, remembering instructions, and repeating sentences. The current findings suggest that a broader range of language abilities is required to capture the level of language development of all ages under study.

The Nonword Repetition Task (Shaalán, 2010) was the only measure that did not show sensitivity to developmental change. Even though the task controls for phonological complexity by including stimuli with different types of clusters, the length of all stimuli in the task is two to three-syllable. Some Arabic dialects, such as Gulf Arabic, can have up to seven-syllable words (Bukshaisha, 1985). It may be that, two to three-syllable nonwords was not sufficiently challenging for our full age range of Arabic-speaking children. In fact, many studies have shown that stimuli length affects children's performance on nonword repetition with older children reaching ceiling for short nonwords (Archibald & Gathercole, 2006b).

We observed low performance on several language measures by females compared to their male peers, even after statistically controlling for their mothers' level of education. Such differences might reflect cultural differences. Kovas et al. (2005) illustrated that after children

develop past toddlerhood, minor sex differences can be observed regarding language development. Although poorly understood at present, biological and environmental factors may account for such differences in language development (Kovas et al, 2005). That is, cultural differences in language practices and the social context in which children learn their language can influence the rate of language development (Lieven, 1994). Children acquire language at faster rates in cultures that provide rich language experiences and more opportunities for language use than in environments that provide less rich language experiences (Hoff, 2005). For example, parents who talk more often to girls than boys may provide girls with a richer environment, helping girls acquire language at a faster rate than boys (Kovas et al., 2005). Such knowledge raises questions about the impact of sociocultural differences between males and females on language practice and development in Saudi Arabia. For example, in Saudi Arabia, public schools provide many physical and recreational activities for males but not for females (Khalaf et al., 2013). In addition, males are allowed to engage in more recreational activities outside of school than females. Sociocultural differences, especially in relation to language practices in Saudi Arabia, may provide males and females with unequal opportunities to experience language. It may be that the sociocultural differences in relation to language practices among males and females in Saudi Arabia provide males with richer language experiences than females. Surprisingly, it should be noted that an advantage for females was found in the Nonword Repetition Task only. Nonword repetition tasks are considered to be processing-based measure. Unlike knowledge-based measures, processing-based measures, such as nonword repetition tasks, reduce the role of cultural and language-specific experience (Kohnert, 2012). As a result, it may be that nonword repetition tasks are less sensitive to sex differences in language

development. It is clear that further investigation of the sex differences observed on language tests for Arabic speaking school age children in the present study is warranted.

With regards to the relationships among the language measures in the present study, correlations ranged from small to moderate in magnitude perhaps suggesting that the measures tap different but related language skills. When individuals exhibiting consistently low performance across the language measures were examined, there was considerable lack of agreement among tasks making interpretation difficult. For example, the agreement between the Arabic Receptive-Expressive Vocabulary Test and the Arabic Picture Vocabulary Task (Receptive only) was low, even though both tasks measured vocabulary skills in this population. It would appear that considerable work is needed to gain a better understanding of the characteristics of developmental language impairment in Arabic speakers, and to develop measures sensitive to both development and impairment.

Study Limitation

As mentioned previously, evaluating children's language skills in light of their dialectal background, culture orientation, and ethnicity is very important. However, dealing with dialects that have significant lexical and phonological variation and change, such as the Urban Hijazi dialect, can make language assessment challenging. In the present study, examiners matched the dialectal variations to the child's spoken output and accepted as correct commonly observed variations. It may be that future studies will assess the impact of dialectal variations in more detail.

Future Direction

This study shows that current Arabic language measures have significant limitations. Language measures aim to evaluate whether or not a child follows an expected level of language

development in order to address language concerns. It is clear from the description of all of the tasks in the abovementioned studies that the authors have worked to ensure that the design of stimuli was culturally appropriate for Arabic-speaking children. In addition, the stimuli were designed in light of aspects of Arabic phonology, morphology, syntax, and semantics.

Nevertheless, given the lack of normative data for typical and atypical language development in the majority of Arabic colloquial dialects, the design of the current Arabic tasks draws heavily on the findings of English-speaking children. The present findings suggest that currently available measures are not capturing crucial changes in child language development beyond about 7 years of age. There is a clear need to establish normative data across the ages studied in the current work in order to inform future Arabic language test design and develop measures sensitive to language development across a wider age range.

Conclusion

In this study, a large sample of school-aged Arabic-speaking children from Saudi Arabia completed individual measures of sentence comprehension, expressive language, sentence repetition, receptive and expressive vocabulary, and nonword repetition. Results revealed that available Arabic language measures are sensitive to developmental change in younger children only between the ages of 6 and 7. Only the Arabic Receptive-Expressive Vocabulary Test revealed significant performance growth with each increase in age band, although this test was the only measure to employ fixed (and different) start and stop points for each age group. The low performance on several language measures by females compared to their male peers may reflect sociocultural differences, especially in relation to language practices between males and females in Saudi Arabia. The results suggest that there is a need to develop language assessment

measures that evaluate a broad range of language abilities and more complex language skills for Arabic-speaking children, and that are based on the psycholinguistics of Arabic.

Chapter 3

English Language Learners (ELLs): Linguistic Gaps and Cognitive Strengths

Minority first language (L1) children who live in a majority English social context, and attend school where English is the language of instruction, are referred to as English Language Learners (ELLs). The ways in which ELLs' linguistic abilities develop and the potential factors affecting their development is of interest to both researchers and interventionists. Understanding the current level of language knowledge in ELLs, however, can present a challenge. Standardized language tests are commonly used in assessing language, however, such tests tend to tap prior knowledge and experience. ELLs may score poorly on such 'knowledge-based' measures because of the low levels of exposure to each of their languages. Measures of cognitive processing, on the other hand, are less dependent on ELLs' linguistic knowledge because they employ nonlinguistic or novel stimuli to tap skills considered to underlie language learning. Recent studies have investigated the extent to which such 'processing-dependent' measures, in particular working memory measures, support language learning. Nevertheless, the particular pattern of performance across knowledge- and processing-dependent measures is not well understood, but is likely influenced by several factors such as chronological age, age of first exposure, the richness of language environment outside of school, current use patterns, and mother's education. The purpose of this study was to investigate ELLs' performance on linguistic measures (Arabic L1 and English L2) and cognitive measures relative to age-level expectations for monolingual speakers, and to consider the influence on performance of a number of relevant experiential factors.

ELLs account for 16 to 36% of the elementary school population in large urban centers in Canada (Statistics Canada, <http://www.statcan.ca>), and 10% of the elementary school population in the United States (National Center for Education Statistics [NCES], 2006). ELLs who speak a minority language at home, and who are in the process of learning English as a second language (L2) through contact with the majority societal language, are L1 minority L2 learners. The developmental course of L1 and L2 acquisition amongst ELLs is highly variable and dependent on a number of factors. For example, the age of exposure to the languages is important (Paradis et al., 2011): ELLs who are exposed to two languages early in their life, at least before 3 years of age, have different degrees of familiarity with L1 and L2 than ELLs who begin the acquisition of L2 later in life or after establishing their L1 (Paradis et al., 2011). Broadly speaking, bilingual children who are exposed to two languages and who begin to learn them before 3 years of age are expected to be more fluent speakers of both languages than bilingual children who begin the acquisition of a second language after 3 years of age (Paradis et al., 2004). Another factor is related to ongoing exposure: ELLs from L1 minorities receive little community support for their L1, as such, their opportunities to hear and use their first language may decrease especially once ELLs start schooling (Anderson, 2012; Paradis et al., 2010). ELLs from minority L1, therefore, are likely to be at risk for loss and/or incomplete acquisition of their L1 (Anderson, 2012; Paradis, 2010). Understanding ELLs' language development and how ELLs acquire specific linguistic skills is important not only for understanding child development, but also for identifying factors that may influence the course of ELLs' development.

Language development is commonly measured using knowledge-based measures such as standardized measures of language and vocabulary. On such measures, a child's level of language development is determined relative to normative data from a largely (and possibly

exclusively) monolingual sample. As the exposure patterns of ELLs and the social contexts in which they are learning both languages differ from monolingual development, disadvantages for ELLs as compared to monolinguals might be expected on such measures. For example, in standardized testing of morphosyntax development, Paradis (2005, 2008) found that 1 in 24 ELLs achieved performance comparable to their monolingual peers after one year of exposure to English. Importantly, however, the number of ELLs who achieved performance comparable to their monolingual peers increased to approximately half of the ELLs after three years of exposure to English. Other studies have reported similar findings, that is, that it can take between two to three years of exposure to English for ELLs to accurately use early acquired morphemes, such as plural *-s* and progressive *-ing* (Jia, 2003; Jia & Fuse, 2007; Paradis, 2008). For late-acquired grammatical morphemes, such as verbal inflections *-s* and *-ed*, however, accurate use can take three to five years of exposure to English for ELLs. According to Paradis, Genesee, and Crago (2011), the pattern of morphosyntactic development in ELLs parallels that of younger monolingual children who acquire the same language. It would follow from these findings that the use of standardized tests of morphosyntax development with ELLs will be influenced by prior knowledge and experience, as well as actual language ability.

Similarly, ELLs have been found to perform at lower levels than their monolingual peers on standardized measures of vocabulary (Carlson & Meltzoff, 2008; Thordardottir, Rothenberg, Rivard, & Naves, 2006). The disadvantages for ELLs on vocabulary measures is in keeping with the *weaker link hypothesis* (Gollan, Montoya, Cera, & Sandoval, 2008). That is, lexical retrieval is influenced by the *word frequency effect*, in which more frequently used words are more easily accessed than relatively less frequently used words (Ellis, 2002). It has been suggested that with increased language practice, the links between semantic and phonology representations become

stronger, which places ELLs at a disadvantage due to their reduced practice in each language (Gollan et al., 2008; Pearson, Fernandez, Lewedeg, & Oller, 1997). As such, measuring vocabulary skills of ELLs is complicated by the fact that lexical-semantic knowledge of ELLs is distributed across both languages. As might be expected, measuring the vocabulary skills of ELLs in only one of their languages may underestimate their vocabulary knowledge (Carlson & Meltzoff, 2008; Thordardottir et al., 2006), even if the assessment is done in the child's dominant language (Kan & Kohnert, 2005). As a result, standardized measures of vocabulary may not reflect the full range of ELLs' vocabulary knowledge (Pearson et al., 1993).

It has been suggested that some language-related measures tap skills likely to be transferable from L1 to L2 resulting in higher performance of ELLs on such tasks (Paradis et al., 2011, 2013). Available evidence shows that perceptual-cognitive skills supporting written word decoding can be transferable from L1 to L2 learning (Paradis et al., 2011). For example, phonological awareness skills are strongly related to word reading (LaFrance & Gottardo, 2005), and have been found to be transfer across the two languages of ELLs (Chitiri, Sun, Willows, & Taylor, 1992; Wade-Woolley & Geva, 2000). One possible explanation is provided by the *Common Underlying Proficiency* theory of Cummins (1996), which proposes that underlying skills and metalinguistic knowledge acquired while learning one language supports the learning of other languages. Indeed, at least equivalent performance has been reported for ELLs when compared to monolingual peers for both phonological awareness (Bialystok, Majumder, & Martin, 2003; Bruck & Genesee, 1995; Campbell & Sais, 1995; Jackson, Holm, & Dodd, 1998; Kang, 2012), and basic word-decoding skills, especially when the language of testing is the same as the child's language of literacy instruction (Bialystok et al., 2003; Oller et al., 2007). In fact, a general advantage in phonological awareness abilities has been reported in several studies for

ELLs from different L1 backgrounds as compared to monolingual children (Bruck & Genesee, 1995; Campbell & Sais, 1995; Kang, 2012). It should be noted, however, that several studies have not found such an advantage, but rather reported no differences between ELLs and monolingual peers on phonological awareness tasks (Bialystok et al., 2003; Jackson et al., 1998). In sum, unlike language-specific lexical and morphosyntactic oral language abilities, perceptual-cognitive skills that are shared across languages, such as those supporting the decoding of written words, may be a source of strength in ELLs. Testing of such skills may help to characterize the language learning ability of ELLs.

Certain cognitive measures may also help to characterize the language learning ability of ELLs. The investigation of the cognitive processes underlying language impairment have implicated deficits in linguistic and nonlinguistic domains, such as deficits in processing speed, temporal integration, and working memory (Miller et al., 2001; Windsor & Kohnert, 2004). However, increasing evidence has suggested that performance in working memory in particular is strongly related to language learning in native and foreign languages (Majerus et al., 2006; Masoura & Gathercole, 2005). Immediate memory systems consist of the domain-specific short-term memory storage of information, and the domain-general executive processes of cognitive control that coordinate and direct those maintenance operations (Baddeley, 2000). Two types of information are involved in the ability to retain information over brief periods of time in short-term memory: the immediate recall of verbal information such as recall of word, letters or digits (verbal short-term memory) and visuospatial information such as the retention of either visual patterns or sequences of movements (visuospatial short-term memory) (Archibald & Gathercole, 2006b; Smyth & Scholey, 1996). Verbal short-term memory has been shown to make an

important contribution to new word learning (Majerus et al., 2006; Masoura & Gathercole, 2005), and vocabulary acquisition (Gathercole et al., 1992).

Working memory is assessed using complex memory span paradigms that imposed both temporary storage and significant processing demands. Examples of verbal complex span tasks are counting recall and backwards digit recall, in which a participant recalls numbers after counting or reversing the order, respectively. Examples of corresponding visuospatial tasks involve recalling locations or orientations after identifying a different shape or mentally rotating an image, respectively (Alloway et al., 2009). Executive processes of working memory have been associated with higher order linguistic abilities, such as language comprehension and word decoding (Cain et al., 2004; Engel de Abreu & Gathercole, 2012). Increasing evidence suggests that children with language impairment have consistent and substantial deficits on both short-term and working memory tasks that require the immediate memory of phonological forms (Archibald & Gathercole, 2006b; Henry et al., 2012). Age-appropriate performance, in contrast, has been reported for children with language impairment on visuospatial short-term and working memory measures (e.g., Archibald & Gathercole, 2006b). Consequently, a specific verbal deficit in immediate memory but not on visuospatial information may conceivably be the basis of some of the language learning difficulties experienced by children with language impairment (Archibald & Gathercole, 2006b; Henry et al., 2012).

Unlike knowledge-based measures that rely on acquired experience and skills, short-term and working memory measures emphasize the storage and processing of new information (Engel de Abreu et al., 2013). Given the emphasis on the processing of new information, such processing-based measures are inherently less biased because they pose similar demands on individuals regardless of their previous knowledge and experience. As a result, at least

equivalent performance would be expected for ELLs when compared to monolingual peers on such tasks. Indeed, several studies have indicated that bilingual children demonstrate equivalent performance as compared to monolingual peers on working memory measures (Engel de Abreu, 2011). In fact, some studies report that bilingual children may have some advantage in verbal working memory (Bialystok & Feng, 2009), and visuospatial working memory (Bialystok, Martin, & Viswanathan, 2005). It has been argued that a working memory advantage might arise in the bilingual context due to the need to switch between both languages on a regular basis (Bialystok, Craik, & Luk, 2008). Converging evidence comes from findings that bilingual children activate both of their languages (Costa, Roelstraete, & Hartsuiker, 2006; Jared & Kroll, 2001) giving rise to an ongoing need to resolve lexical conflict (Bialystok, 1999). Recent research has revealed that the constant use of cognitive control to resolve lexical conflict may boost performance on executive control tasks involving inhibition and shifting in bilingual children (Bialystok et al., 2008; Carlson & Meltzoff, 2008). Working memory is one of the cognitive mechanisms strongly linked to cognitive control (Baddeley & Hitch, 1974). As such, an ELLs advantage in tasks that rely on this mechanism may be expected (e.g., Bialystok et al., 2004), although null effects are also commonly reported (Bajo, Padilla, & Padilla, 2000; Bialystok et al., 2008).

One factor possibly accounting for the inconsistent reports of ELLs' performance on working memory tasks may be related to the nature of the stimuli involved, and in particular, the inclusion and type of verbal stimuli tested (Engel de Abreu et al., 2013). Consider, for example, the differences between tasks involving recall of highly familiar lexical stimuli such as digit recall compared to tasks that involve material not taught explicitly such as nonwords (Engel de Abreu et al., 2013; Gathercole & Adams, 1994). Engel de Abreu et al. (2013) reported an

advantage for monolingual over bilingual children on tasks with unfamiliar (i.e., nonword repetition) but not familiar verbal stimuli (i.e., digit recall). The lack of group differences on number-based immediate memory measures has been attributed to the familiarity of number stimuli: Children generally acquire number words at an early age making them familiar and easily recalled by all children regardless of individual differences in stored lexico-semantic knowledge in long-term memory (Engel de Abreu et al., 2013; Gathercole & Adams, 1994). Nonwords, on the other hand, have an unfamiliar phonological form, which might be expected to reduce any monolingual advantage in recall. Nevertheless, available evidence shows that even previous sublexical phonological knowledge can influence performance on nonword tasks (Thorn & Gathercole, 1999). Indeed, an advantage has been reported for monolingual English-speaking children as compared to ELLs on a nonword repetition test designed to follow the phonotactic rules of English (Thorn & Gathercole, 1999). Although nonword repetition may minimize the role of linguistic knowledge and experiences as suggested by Paradis et al. (2013), it is clear that nonword repetition does not entirely eliminate the effect of children's experience with the target language (Engel de Abreu et al., 2013).

It is clear from the results above that employing different measures of language and working memory might result in differing performance patterns among ELLs. In addition, a number of researchers have suggested that other factors influence ELLs' language development. Factors shown to affect ELLs' language acquisition include chronological age, age of first exposure (Birdsong, 2005; Hammer et al., 2012), the richness of language environment outside of school (Jia & Aarsonson, 2003; Jia & Fuse, 2007), current use patterns (Bedore et al., 2012), and mother's education (Bohman et al., 2010; Golberg et al., 2008). For age of exposure to language and length of exposure, research shows that age of exposure accounts for 65% of the

variance in pronunciation, vocabulary, and grammar of bilingual children (Birdsong, 2005). In addition, length of exposure was found to be a significant predictor of language performance among ELLs (Hammer et al., 2012). In terms of richness of language environment, Jia and Aaronson (2003) and Jia and Fuse (2007) found that richness of the environment around ELLs, such as the number of hours of English TV watched, the number of English native-speaking friends, and the number of English books read, is associated with faster English language acquisition. In regards to current use pattern, Bedore et al. (2012) found that among several factors, current use pattern is the most informative indicator accounting for more of the variance in language performance among ELLs. In terms of maternal education and socioeconomic status, research shows that higher parental education is associated with ELLs' vocabulary scores and ELLs' rate of English vocabulary growth (Bohman et al., 2010; Golberg et al., 2008). Importantly, some of these factors have also been reported to impact cognitive development generally including age (Case, Kurland, & Goldberg, 1982, Mueller Gathercole et al., 2010), experience (Bialystok, 2001), and mother's education (Bohman et al., 2010; Golberg et al., 2008). It is clear that different ELLs' experiences and levels of exposure across languages may result in varied profiles of language and cognitive performance.

Given the variation among ELLs, it is important to understand how best to measure language, language learning potential, and consider experiential factors in children engaged in the process of learning more than one language but who may not have mastered either language. The ways in which ELLs' linguistic abilities develop and resemble and differ from monolingual children is of interest to clinicians involved in assisting ELLs. The present study considered the case of Arabic-English ELLs. The purpose of this study was to investigate ELLs' performance on linguistic measures (Arabic L1 and English L2), and cognitive measures relative to age-level

expectations for monolingual speakers. Knowledge-dependent language measures included Arabic and English measures of vocabulary and language. As such, lower scores for the ELLs than the standardized mean scores of Arabic/English monolinguals would be expected. Processing-dependent measures included phonological awareness (word reading), and verbal and visuospatial short-term and working memory measures. Although nominally a language task, word reading in the children's language of literacy instruction was considered to tap phonological awareness. For verbally-mediated processing measures (phonological awareness, verbal short-term and working memory), we expected no difference from the standardized mean scores of monolinguals for ELLs, provided the stimuli employed were less affected by existing language knowledge. For the visuospatial immediate memory measures considered to be less sensitive to language differences, we expected equivalent performance by ELLs as compared to the standardized mean scores of monolinguals. The influence of a number of experiential factors potentially influencing the performance of the ELL was also evaluated including chronological age, age of first exposure, the richness of language environment outside of school, current use patterns, and mother's education.

Methods

Participants

The ELLs in the present study were bilingual speakers whose L1 was Arabic, who had been learning English as their language of instruction (English L2) in Canada, and were in grades 1 to 4 at elementary schools located in London, Ontario. A total of 59 children participated ($M_{age} = 7$ years; 11 months, $SD = 1.16$) including 29 males ($M_{age} = 8;3$, $SD = 1.26$) and 30 females ($M_{age} = 7;8$, $SD = 1.03$). All of the participants were recruited from a school providing instruction

in both English and Arabic ($n=27$), or from an extracurricular Arabic instruction class for children receiving regular schooling in English ($n=32$).

Procedure

Participants completed a battery of assessment measures in both Arabic (L1) and English (L2) individually in a quiet room in their school over 6 weekly sessions each of approximately 40 minutes. The battery included measures of vocabulary, (*Arabic Receptive-Expressive Vocabulary Test, AREVT*, El-Halees & Wiig, 1999; *Peabody Picture Vocabulary Test, PPVT*, Dunn & Dunn, 2006; *Expressive Vocabulary Test, EVT*, Williams, 2006), language (*Arabic Language Test, ALT*, Shaalan, 2010; *Clinical Evaluation of Language Fundamentals, CELF-4*, Semel, Wiig, & Secord, 2003), reading (*Arabic Word Reading Task, Test of Word Reading Efficiency, TOWRE*, Torgesen, Wagner, & Rashotte, 1999), short-term memory (*Arabic Nonword Repetition Task, A-NWR*, Shaalan, 2010; *Nonword Repetition Test, NWR*, Dollaghan & Campbell, 1998; relevant measures from the *Automated Working Memory Assessment, AWMA*, Alloway, 2007), working memory (relevant measures from the AWMA), and non-verbal intelligence (*Test of Non-Verbal Intelligence, TONI-3*, Brown, Sherbenou, & Johnsen, 1997), and other measures not reported here. The tests were administered in a fixed order so that each individual session involved the administration of particular tests. In session 1, the AREVT, A-NWR, and ALST were administered. In session 2, the ALT was delivered. Session 3 included the TONI-3, and Arabic Word Reading Task. In session 4, the AWMA and NWR were administered. Session 5 consisted of the CELF-4, SWE, and PDE. Lastly, session 6 included the PPVT, and EVT. A trained native Arabic speaker tested the children in the Arabic language battery tests, whereas a trained native English speaker tested the children in the English language battery tests.

Arabic Language Test Battery.

For all of the Arabic language tests administered, raw scores were converted to standard scores based on a monolingual Arabic-speaking unselected sample of 421 children ranging in age from 6;0 to 9;11 years and recruited from 10 schools in the region of Jeddah, Saudi Arabia (Balilah & Archibald, in press).

Vocabulary. The two subtests of the *Arabic Receptive-Expressive Vocabulary Test* (AREVT, El-Halees & Wiig, 1999) were completed by all participants. In the receptive subtest, children pointed to a picture corresponding to a given spoken word from a choice of four. In the expressive subtest, children named or described a picture with a single word or phrase. The total number of correct responses was counted for each subtest, with the maximum possible score for each subtest dependent on the child's age (6;0 to 6;11: $n=40$; 7;0 to 7;11: $n=45$; 8;0 to 8;1: $n=50$; 9;0 to 9;11: $n=55$).

Oral language. The *Arabic Language Test* (ALT, Shaalan, 2010) was administered. The ALT includes 3 subtests each divided into a first and second section (Section A and B). In the *Sentence Comprehension* subtest, children pointed to a picture that corresponded to the spoken sentence from a choice of three or four. In the *Expressive Language* subtest, children were given a sentence and were asked to generate a spoken word or phrase in reference to a picture cue. The *Sentence Repetition* subtest required immediate repetition of an audio recording of a sentence spoken by a native, adult male Arabic speaker. The total number of correct responses was counted for each subtest, with a maximum possible score of 40 for the Sentence Comprehension subtest (Section A: $n=22$; Section B: $n=18$), and 68 for the Expressive Language subtest (Section A: $n=24$; Section B: $n=44$). The 41 items of the Sentence Repetition subtest were scored on a 4-point scale (3=correct; 2=1 error; 1=2-3 errors; 0=4 or more errors, or no response), with a

maximum possible score of 123 (Section A: $n=18$; Section B: $n=23$). High test retest reliability has been reported for the three subtests of the ALT (the Sentence Comprehension Test = .95; the Expressive Language Test = .95, and the Sentence Repetition Test = .97) (Shaalán, 2014).

Reading efficiency. In the *Single Word Reading Task*, participants read aloud from a list as many printed words as possible within 45 seconds (maximum score = 104). The 104 Arabic sight words were selected from the single word list a list of commonly used words found in a popular reading series in Lebanon (Oweini & Hazoury, 2010), and progressed in difficulty based on frequency, and number of syllables (beginning with one syllable and progressing to up to six syllables). Words reflecting the unique properties of Lebanese culture were not selected for this task. As well, words that could not be pronounced correctly without diacritic marks, marks added to the original Arabic alphabet to modify word pronunciation (Lutf, You, Cheung, & Chen, 2013), were not selected for this task because standard Arabic text is written without these diacritic marks (Zayyan et al., 2016).

English Language Test Battery.

For all of the English language tests, raw scores were converted to the published norms for all standardized tests.

Vocabulary. Two vocabulary tests were administered, the *Peabody Picture Vocabulary Test (PPVT-4)*, Dunn & Dunn, 2006), and the *Expressive Vocabulary Test (EVT-2)*, Williams, 2006). In the PPVT, children were required to point to a picture corresponding to a given spoken word from a choice of four. In the EVT, children were provided a label or synonym for each item (picture and word) given.

Oral language. The four subtests required to compute the *Composite Language Score (CLS)* from the *Clinical Evaluation of Language Fundamentals (CELF- 4)*, Semel et al., 2003)

were administered. In the *Concepts and Following Directions* subtest, children pointed to a picture corresponding to given spoken directions. In the *Recalling Sentences* subtest, children imitated a sentence presented by the examiner. In the *Formulating Sentences* subtest, children were given visual stimuli and were asked to generate a sentence using target words or phrases. The final subtest was dependent on the child's age, *Word Knowledge* (under 9 years) or *Word Classes* (9 years or older). In the *Word Knowledge* and *Word Classes* subtests, children were asked to choose the two related words from a choice of three (*Word Knowledge*) or four (*Word Classes*), and describe their relationship.

Reading efficiency. The two subtests from the *Test of Word Reading Efficiency (TOWRE)*, Torgesen et al., 1999) were administered. In the *Sight Word Reading Efficiency (SWE)* subtest, participants read aloud as many printed words as possible within 45 seconds from a list that progressed in difficulty (maximum score=104). In the *Phoneme Decoding Efficiency (PDE)* subtest, participants read aloud as many pronounceable printed nonwords as possible within 45 seconds from a list that progressed in difficulty (maximum score=63).

Cognitive measures.

Short-term and working memory. Eight subtests from the *Automated Working Memory Assessment (AWMA)*, Alloway, 2007a) were administered. Measures of phonological short-term memory (*Digit Recall; Word Recall*) required the immediate repetition of numbers or word forms, and measures of visuospatial short-term memory (*Dot Matrix; Block Recall*) required the recall of location. Measures of verbal working memory (*Counting Recall; Backwards Digit Recall*) required the recall of numbers after counting or reversing the order, respectively, and measures of visuospatial working memory (*Odd One Out; Spatial Span*) required the recall of locations or orientations after identifying a different shape or mentally rotating an image,

respectively. The AWMA was administered to each child using the child's preferred language (Arabic or English). Of the participants, 70% preferred English, and 30% preferred Arabic. Given our findings of higher scores by Canadian children compared to the British normative sample of the AWMA (Nadler & Archibald, 2014), standardized scores for the AWMA subtests were based on unpublished norms from our monolingual Arabic-speaking sample (see also, Balilah & Archibald, in press).

Two additional verbal short-term memory tasks were administered: the *Arabic Nonword Repetition Task (A-NWR)*, Shalaan, 2010) and the *Nonword Repetition Task (NWR)*, Dollaghan & Campbell, 1998). In the A-NWR, children repeated the nonwords from an audio recording spoken by a native, adult male Arabic speaker. Items taken from Shalaan (2010) consisted of 48 nonwords varying in length (two to three syllables) and cluster type (no cluster, medial cluster, final cluster, and medial and final clusters). Each response was scored online as correct or incorrect by a trained research assistant, with a maximum possible score of 48. In the NWR, children repeated 16 nonwords, four each being one, two, three, or four syllables in length from Dollaghan and Campbell (1998). The nonwords excluded late-developing sounds, and were constructed from a set of 20 phonemes (11 consonants, 9 vowels). The stimuli were designed to alternate consonant-vowel structure, and none of the syllables corresponded to English words. In this study, the stimuli were presented in fixed order via a digital audio recording of a native English adult female speaker. Each response was scored online as correct or incorrect by a trained research assistant, with a maximum possible score of 16. Raw scores were converted to standardized scores based on unpublished norms from our monolingual Arabic sample for the A-NWR (see also, Balilah & Archibald, in press) and local norms for the NWR.

Nonverbal intelligence. The *Test of Non-Verbal Intelligence (TONI-3)* (Brown et al., 1997) was administered to measure general nonverbal cognitive abilities. In the TONI-3, children chose a picture to complete a visual pattern. Standardized scores for the TONI-3 task were based on published test norms.

Parent questionnaire. Parents completed an extensive questionnaire about their child's language background (Kaushanskaya, Gross, & Buac, 2010). The language background questionnaire included questions related to language immersion, history, and use, and the parent's rating of their child's current language abilities in each language. For language immersion, parents were asked to indicate how often their child participated in activities such as checking out library books, watching television, attending cultural events, and playing computer games in each of their languages. For language history and use, parents were asked to indicate when their child began to say single word, 2-word phrases, and complete sentences in each language, and also how many hours of the day during the week and weekend their child used their languages. In addition, parents were asked to rate their child's current speaking, understanding of spoken language, and reading abilities in each language on a scale from 0-10 (0=none, 1=very low, 2=low, 3=fair, 4=slightly less than adequate, 5=adequate, 6=slightly more than adequate, 7=good, 8=very good, 9=excellent, 10=perfect). Parents also provided information regarding maternal level of education for the child. The descriptors included some high school, completed high school, some college, completed college, some university, and completed university. Responses were transposed to a 3-point scale, with 1 corresponding to some/completed high school, 2 to some/completed college, and 3 to some/completed university. All parents of children reported that their children acquired Arabic as a first language from birth. Additionally, parents reported that 34 participants ($M_{age}=7.38$, $SD=1.07$; 17 males) were exposed

to English before 3 years of age on average, at 2;0 ($SD = 0;7$, range = 8 – 36 months), whereas 19 participants ($M_{age}=7.32$, $SD=1.33$; 9 males) were exposed to English after 3 years of age on average, at 5;5 ($SD = 1;5$, range = 40 – 96 months). Notably, the parents of 6 participants did not indicate the time when their child was first exposed to English. In addition, by parent report, approximately 80% of mothers had at least some college or university education.

Results

ELLs' Performance on Language Measures including Reading

Table 4 provides descriptive statistics for the age-adjusted standard scores of the ELLs on all language measures. For the Arabic language tasks, performance was significantly lower than the expected standardized mean for all measures, $t > 9.4$, $p < .001$, (all cases), with scores approximately 1 SD below the standardized mean for the vocabulary and oral language measures and more than 2 SD below for the single word reading task. Similarly, for the English language measures, performance was significantly lower than the expected standardized mean for the vocabulary and oral language measures, $t > 8.2$, $p < .001$, (all cases). For the reading measures, however, no significant differences were found for word reading (SWE) test, $t = -0.07$, $p > 0.05$, and significantly higher scores were observed for the nonword reading test (PDE), $t = 4.79$, $p < .001$.

Table 4
Descriptive statistics for the language measures

<i>Measure</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
<i>Arabic Language Measures</i>				
AREVT:				
Receptive	56	86.50*	24.35	6 - 127
Expressive	56	85.29*	21.71	42 - 132
ALT	56	82.46*	22.02	24 - 109
Single Word Reading	56	65.88*	30.94	0 - 105
<i>English Language Measures</i>				
EVT-2	53	88.26*	14.49	53 - 120
PPVT-4	53	87.06*	14.77	52 - 127

CLS	52	80.46*	16.84	50 - 120
SWE	59	102.92	14.69	66 - 129
PDE	55	109.31*	14.47	79 - 139

Note. AREVT = Arabic Receptive-Expressive Vocabulary Test; ALT = Arabic Language Test; EVT-2 = the Expressive Vocabulary Test; PPVT-4 = the Peabody Picture Vocabulary Test; CLS = Composite Language Score of the Clinical Evaluation of Language Fundamentals; SWE = the Sight Word Reading Efficiency; PDE = the Phoneme Decoding Efficiency.

* $p < .05$.

ELLs' Performance on Cognitive Measures

Table 5 presents descriptive statistics for ELLs on the short-term and working memory tests of the AWMA, both the Arabic (A-NWR) and English (NWR) nonword repetition tasks, and the nonverbal intelligence test (TONI-3). One-sample t -tests revealed no significant differences between the ELL scores and the expected standardized mean on Word Recall, Counting Recall, Dot Matrix, Odd One Out, Spatial Span, English nonword repetition, and nonverbal intelligence tasks, $t > 0.28$, $p > 0.05$, (all cases). Conversely, scores were significantly higher than expected on the Digit Recall, Backwards Digit Recall, and Block Recall subtests, $t > 2.3$, $p < 0.05$, (all cases). The only task for which significantly lower than expected scores were observed was for the Arabic nonword repetition test, $t = 3.7$, $p < 0.05$.

Table 5

Descriptive statistics for the cognitive measures

<i>Measure</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
AWMA:				
Digit Recall	59	108.27*	18.31	74 - 160
Word Recall	59	102.86	12.37	58 - 125
Counting Recall	59	100.75	12.86	77 - 131
Backwards digit Recall	59	103.29*	10.45	86 - 128
Dot Matrix	59	101.78	15.85	69 - 147
Block Recall	59	105.03*	10.88	78 - 129
Odd One Out	59	102.15	15.56	73 - 151
Spatial Span	59	100.44	12.01	68 - 122
A-NWR	56	90.96*	17.87	46 - 119
NWR	59	99.85	16.41	64 - 125
TONI-3	56	102.80	12.89	78 - 140

Note. AWMA = the Automated Working Memory Assessment; A-NWR = Arabic-Nonword Repetition; NWR = Nonword Repetition; TONI-3 = The Test of Non-Verbal Intelligence.
* $p < .05$.

Experience-based Predictors of Performance

The final set of analyses determined which of the demographic variables best predicted ELLs' performance on the Arabic L1 and English L2 languages measures and cognitive measures. Separate linear regressions were completed on the Arabic measures (the receptive and expressive vocabulary subtests of the AREVT, ALT, and Single Word Reading Task), with the following demographic variables entered as predictors: age in months (months of exposure to Arabic), richness of the Arabic environment outside school, current use patterns of Arabic during waking hours in a typical week, and maternal education. For the English language measures (EVT, PPVT, CLS, SWE, and PDE), the demographic variables included as predictors were the following: age in months, richness of the English environment outside school, current use patterns of English during waking hours in a typical week, age of first exposure to English in months, months of exposure to English, and maternal education. For the cognitive measures, all the demographic variables entered as predictors on the Arabic and English measures were entered.

A stepwise multiple-regression method was used to find the best fitting model that accounted for variation in each task's raw scores using separate models. As a first step, we examined correlations between the independent variables in each model, as strong correlations between the predictor variables could cause collinearity effect in the models. The correlation between age of first exposure to English in months and months of exposure to English was strong in magnitude, with $r = -.84$ ($p < .001$). As a result, we removed months of exposure to English, but not age of first exposure to English in months, from the regression models.

Following this, there were no independent variables correlated above .5. In addition, as Variance Inflation Factor (VIF) is common multicollinearity diagnostic, VIFs were calculated to measure the correlation among the independent variables. VIF values ranged from 1.0 to 1.3 in all cases, indicating no multicollinearity problems were present in the model. According to Mendenhall and Sincich (1996), VIF values greater than 10 indicate multicollinearity problems.

Table 6 summarizes the results of the stepwise regression models for the Arabic language measures. Age in months (months of exposure to Arabic) and richness of the Arabic environment outside school were the two variables that made a significant contribution to the models in all of the Arabic language tasks, with age in months explaining more variance as the strongest standardized beta coefficients were for age in months for all of the Arabic language measures.

Table 7 summarizes the results of the stepwise regression models for the English language measures. Age of first exposure to English in months, age in months, and richness of the English environment outside school were the three variables that made significant contributions to the models for the vocabulary measures (EVT and PPVT). Richness of the English environment outside school was the only variable that made a significant contribution to the models in the oral language measure (CLS). Age in months and maternal education were the two variables that made a significant contribution to the models in the SWE. Finally, age in months, richness of the English environment outside school, and maternal education were the three variables that made a significant contribution to the models in the PDE.

Table 8 summarizes the results of the stepwise regression models for the cognitive measures. Age explained significant variance for all cognitive measures. In addition, current use patterns of Arabic explained variance in the model for Digit Recall, age of first exposure to

English in months made a significant contribution to the model for the nonverbal intelligence test.

Table 6
Stepwise regression models results for Arabic language measures

Task	Factors	Unstandardized B	coefficients St.error	Standardized Beta	t	Coefficients Sig.	Adj. R ²
AREVT (Receptive)	Age	.45	.04	.88	9.93	<.001	.69
	Rich env.	.24	.09	.22	2.49	<0.05	
AREVT (Expressive)	Age	.37	.05	.75	6.83	<.001	.52
	Rich env.	.42	.11	.40	3.63	=.001	
ALT	Age	1.22	.42	.42	2.90	=.006	.16
	Rich env.	2.14	.91	.34	2.34	<0.05	
Single Word Reading	Age	.80	.10	.80	7.62	<.001	.57
	Rich env.	.80	.22	.36	3.51	=.001	

AREVT = the Arabic Receptive-Expressive Vocabulary Test; ALST = the Arabic Language Screening Test; and ALT = the Arabic Language Test. Age = age in months (months of exposure to Arabic); and Rich env. = richness of the Arabic environment outside school.

Table 7
Stepwise regression models results for English language measures

Task	Factors	Unstandardized B	coefficients St.error	Standardized Beta	t	Coefficients Sig.	Adj. R ²
EVT	Age of expo.	-.42	.09	-.47	-4.60	<.001	.68
	Age	.79	.12	.57	6.22	<.001	
	Rich env.	.91	.38	.24	2.38	<0.05	
PPVT	Age	1.25	.16	.67	7.83	<.001	.73
	Rich env.	1.82	.48	.36	3.80	=.001	
	Age of expo.	-.39	.11	-.31	-3.36	=.002	
CLS	Rich env.	.91	.32	.43	2.83	=.008	.16
SWE	Age	1.19	.13	.79	8.97	<.001	.66
	Maternal edu.	-7.74	2.81	-.24	-2.75	=.009	
PDE	Age	.74	.10	.71	6.99	<.001	.59

Rich env.	.77	.29	.27	2.62	<0.05
Maternal edu.	-5.06	2.28	-.22	-2.21	<0.05

EVT-2 = the Expressive Vocabulary Test; PPVT-4 = the Peabody Picture Vocabulary Test; CLS = Composite Language Score of the Clinical Evaluation of Language Fundamentals; SR = Sentence Recall; SWE = the Sight Word Reading Efficiency; PDE = the Phoneme Decoding Efficiency. Age = age in months; Rich env. = richness of the English environment outside school; Age of expo. = age of first exposure to English in months; and Maternal edu. = maternal education.

Table 8
Stepwise regression models results for cognitive measures

Task	Factors	Unstandardized B	coefficients St.error	Standardized Beta	t	Coefficients Sig.	Adj. R ²
AWMA:							
Digit Recall	Age	.18	.04	.51	4.44	<.001	.45
	Current use (Arabic)	-.12	.03	-.37	-3.21	=.003	
Word Recall	Age	.12	.03	.49	3.71	=.001	.23
Counting Recall	Age	.19	.03	.62	5.22	<.001	.38
Backwards digit Recall	Age	.11	.03	.46	3.41	=.001	.20
Dot Matrix	Age	.19	.03	.62	5.23	<.001	.38
Block Recall	Age	.14	.03	.57	4.50	<.001	.31
Odd One Out	Age	.11	.04	.36	2.54	=.015	.11
Spatial Span	Age	.18	.03	.62	5.15	<.001	.37
A-NWR	Age	.18	.08	.32	2.16	<0.05	.08
TONI-3	Age	.19	.06	.43	3.26	=.002	.31
	Age of expo. (English)	-.09	.04	-.29	-2.23	<0.05	

Note. AWMA = the Automated Working Memory Assessment; A-NWR = Arabic-Nonword Repetition; TONI-3 = The Test of Non-Verbal Intelligence. Age = age in months; Current use (Arabic) = current use patterns of Arabic during waking hours in a typical week; Age of expo. (English) = age of first exposure to English in months.

Discussion

The purpose of this study was to investigate ELLs' performance on linguistic measures (Arabic L1 and English L2) and cognitive measures relative to age-level expectations for monolingual speakers. The study also investigated the influence of a number of relevant experiential factors on ELLs' performance. Results of this study revealed that ELLs scored significantly lower than age-level expectations on the Arabic/English vocabulary and language tasks, and on the Arabic but not English reading tasks. ELLs demonstrated significantly higher performance on the English word and nonword reading tasks relative to normative data. Age and richness of environment were the two factors that consistently explained variance in the ELLs' performance in L1 Arabic. Among English language tasks, however, each language task had different factors that contributed to the ELLs' performance depending on the task. In general, age, richness of environment, and age of first exposure were the three informative predictors that explained ELLs' performance variance on the majority of the English language tasks. With regards to the immediate measures, ELLs scored at age-level expectations on the short-term and working memory and nonverbal intelligence tasks with the exception of the Arabic nonword repetition task. ELLs had significantly lower standardized scores on the Arabic nonword repetition task compared to the standardized mean scores of the monolingual group. Interestingly, the ELLs' scores were significantly higher on two number-based verbal working memory measures, and one visuospatial short-term memory task. Moreover, age alone was the variable that consistently contributed to the ELLs' performance on all of the cognitive measures.

In this study, Arabic(L1)-English(L2) ELLs' scores were lower than age-expectations based on standardized scores on knowledge-based language measures in both of their spoken languages. The finding of significantly lower standardized scores for the ELLs on Arabic and

English knowledge-based language tasks is consistent with many previous studies (e.g., Paradis, 2005; Paradis, Rice, Crago, & Marquis, 2008). Although one interpretation of significantly low language scores could be that a language learning deficit is present, there is reason to be cautious of such an interpretation in the case of ELLs. ELLs' performance on knowledge-based measures is strongly influenced by ELLs' acquired skills and experience with the target language (Jia, 2003; Jia & Fuse, 2007; Paradis, 2008). As a result, ELLs are at a disadvantage when compared to monolingual groups on knowledge-based measures such as those employed to estimate language abilities in the present study (Paradis et al., 2011). This suggestion is supported by the results of age-appropriate English nonword repetition by the ELL group in the present study. Nonword repetition has been found to discriminate on the basis of language impairment (Archibald & Gathercole, 2006a; Dollaghan & Campbell, 1998b), but not socioeconomic status (Campbell, Dollaghan, Needleman, & Janoksy, 1997), previous preschool experience (Campbell et al., 1997; Ellis Weismer et al., 2000), or bilingual status (Paradis et al., 2013). Consequently, nonword repetition has been suggested to be a less biased measure of language learning potential than knowledge-based language measures (Paradis et al. 2013). For the present findings, then, the age-appropriate nonword repetition scores might reflect unimpaired language learning potential in the ELL group despite scores in the deficit range on the knowledge-based language tests.

The ELLs showed comparable or higher than expected performance on the English word and nonword reading tasks, but lower than expected scores on the Arabic word reading task. It is possible that this pattern reflects a bilingual advantage, at least for the language of instruction. These findings may add to the growing evidence of a phonological awareness benefit from learning two different phonological systems (Kang, 2012; Marinova-Todd, Zhao, & Bernhardt,

2010; Oller et al., 2007). That is, ELLs' who are exposed to two phonologically different language systems may be at some advantage on phonological tasks (e.g., Kang, 2012; Marinova-Todd et al., 2010) especially when the language of testing is the same as the child's language of literacy instruction (Bialystok et al., 2003; Oller et al., 2007). In addition, the results suggest that English word and nonword reading tasks may be less sensitive to difference in language experience than traditional knowledge-based measures such as standardized measures of language and vocabulary. The significantly higher performance of ELLs on the English nonword reading task but not on the English word reading task is difficult to interpret. One possible explanation comes from the *dual route theory* (Coltheart, Curtis, Atkins, & Haller, 1993), which holds that words are read through a rapid memory route or a phonological decoding route. Nonwords are not represented in the child's lexicon, and as such, reading nonwords relies more heavily on phonological skills whereas high frequency words may be recalled through the rapid recognition route (Hagiliassis, Pratt, & Johnston, 2006). Given that ELLs may have better phonological skills to support phonological decoding, higher scores on the nonword reading task may be expected.

On all of the cognitive measures except the Arabic nonword repetition task, ELLs scored at or above age-level expectations. These results are consistent with previous evidence suggesting that measures of cognitive abilities in ELLs are less sensitive to difference in language experience than knowledge-based measures (Cockcroft, 2016; Engel de Abreu et al., 2013). Moreover, the present study's finding regarding the reduced performance of ELLs on Arabic nonword repetition but not in English nonword repetition adds to the growing evidence indicating that ELLs' performance on nonword repetition is affected by their previous sublexical phonological knowledge and experience in the target language (Thorn & Gathercole, 1999;

Kohnert et al., 2006; Windsor et al., 2010). That is, ELLs' performance on nonword repetition has been found to be highly affected by language-specific knowledge (Thorn & Gathercole, 1999). ELLs' better performance on English nonword repetition may have been affected by their relative strengths in English language-specific knowledge. In addition, as English is supported through schooling, ELLs may be exposed more frequently to lexical and sublexical analysis in English than in their L1 language.

Finally, the finding that ELLs scored significantly higher on two number-based verbal working memory measures and one visuospatial short-term memory task is consistent with previous evidence suggesting that bilingual children may be at some advantage on working memory abilities (Bialystok et al., 2004). However, comparing bilingual advantage in working memory abilities across studies is complicated by the fact that studies have used different tasks perhaps differing in the underlying cognitive processes (Engel de Abreu, 2011). Moreover, it is not clear why ELLs in this study demonstrated an advantage in these specific subtests but not on other subtests that may share the same underlying processes. It would appear that considerable work is needed to gain a better understanding of the ELLs' advantage on working memory abilities.

Regarding the results of the stepwise multiple-regression analyses, age and richness of environment were identified as the two factors that consistently contributed to the ELLs' performance in L1 Arabic. In fact, these results are not particularly surprising. In the present ELL sample, all children had been learning Arabic from birth, and so their age reflected their duration of exposure to Arabic. A large body of research indicates that age of exposure, or the amount of time that the child has been learning the target language, is a significant predictor of language performance (e.g., Hammer et al., 2012). Although significant, the richness of the

Arabic environment explained less variance than age of exposure in Arabic language performance in the ELL group, a finding consistent with previous research (Paradis, 2011). Among English language tasks, however, each language task had different factors that contributed to the ELLs' performance depending on the task. In general, age, richness of environment, and age of first exposure were the three informative predictors that contributed to ELLs' performance on the majority of the English language tasks. Unlike age in Arabic L1 that reflects the age of exposure to Arabic, age and exposure to English L2 are more independent of each other. The finding demonstrates that ELLs' performance on the majority of the English language tasks improves with age. In addition, age of first exposure to English was the most informative predictor that contributed to the model of the expressive vocabulary test, but it contributed less to the model of receptive vocabulary measures. That is, early age of first exposure to English has more predictive value on the ELLs' performance on the expressive vocabulary test, whereas age in months and richness of English environment outside school had more predictive value on ELLs' performance on the receptive vocabulary test. It could be that early age of first exposure reflects the depth of experience related to expressive vocabulary, whereas richness of English environment influences breadth of knowledge and is related more to receptive vocabulary. Another interesting result in the present study was that there was an inverse relationship between mother's level of education and ELLs' performance on the English reading tasks. This finding is difficult to interpret but may be related to the low variability observed in our maternal education scale. Finally, it should be noted that current language use was found to be a significant predictor of ELLs' performance on language in some other studies (Bedore et al., 2012); however, our findings show that current use patterns of Arabic/English did not account significantly for variation in any task of the Arabic L1 and English L2. The

inconsistent results among studies may be related to the different tasks that have been used, and the differences in the populations that have been studied.

Finally, among cognitive tasks, age was the variable that explained more variance in the ELLs' performance in all of the cognitive measures, and age alone was the variable that made significant contributions to the models for the majority of the cognitive measures. A large body of research findings indicate that a child's performance is expected to increase with age in cognitive tasks such as tasks that measure phonological short-term memory capacity (e.g., Case, Kurland, & Goldberg, 1982). Interestingly, a possible explanation for the overall results of the cognitive measures is that, unlike knowledge-based measures, cognitive measures were relatively minimally impacted by language experiences. Therefore, cognitive measures may provide an accurate indication of ELLs' language learning potential. Another interesting aspect of the cognitive measures in this study was the inverse relationship between the current use patterns of Arabic during waking hours in a typical week and ELLs' performance on the Digit Recall subtest of AWMA, and between age of first exposure to English and ELLs' performance on The Test of Non-Verbal Intelligence. These findings are difficult to interpret, and further investigation of the bilingual experience effects on the cognitive measures in this population would help to clarify this relationship. However, a possible explanation regarding the inverse relationship between the current use patterns of Arabic during waking hours in a typical week and ELLs' performance on the Digit Recall subtest of the AWMA is that more experience with English may lead to a better performance on the recall of English number words, as the majority of the ELLs in this study preferred to do the task in English. Moreover, a possible explanation of the relationship between age of first exposure to English and ELLs' performance on The Test of Non-Verbal Intelligence is that performance on such tasks may be affected by the ELLs' cultural experiences.

Study Limitation

Information about the Arabic colloquial dialects of ELLs was unavailable. Certainly, as Arabic-speaking children use the colloquial dialect in their daily oral communication, language assessment measures should address the acquisition of the colloquial dialect (Al-Tamimi, 2011). Unfortunately, there are no available assessment measures in the majority of Arabic colloquial dialects. In the present study, examiners matched the dialectical variations to the child's spoken output and accepted as correct commonly observed variations. It may be that future studies will assess the impact of dialectical variations in more detail.

Conclusion and Future Direction

The main objective of this study was to determine ELLs' performance on linguistic and cognitive measures relative to age-level expectations for monolingual speakers. In this study, 59 unselected children aged 6 to 9 years old, whose L1 was Arabic and who had been learning English as the language of instruction (English L2) in Canada, completed a battery of linguistic tests of vocabulary, language, and reading in both languages (Arabic L1 and English L2), and cognitive measures of short-term and working memory, and non-verbal intelligence. Significantly lower standardized scores were observed for ELLs as compared to the standardized mean scores of Arabic/English monolinguals on all the Arabic/English language tasks except on the single word reading tasks in the children's L2, or the language of literacy instruction. In comparison to the language measures, ELLs scored at or above age-level expectations on the cognitive measures of working memory and nonverbal intelligence except on the nonword repetition task in the children's L1. This study shows that employing different measures of language and cognition might result in differing performance among ELLs. That is, ELLs' standard scores varied among their Arabic L1 and English L2, as well as among different

language abilities relative to age-level expectations for monolingual speakers of each language. The results suggested that cognitive measures were relatively independent of language experiences and may provide an accurate indication of ELLs' language learning potential. In addition, the main finding of this study suggested that careful choice among tasks is required to ensure a better understanding of the current level of language knowledge in ELLs.

Chapter 4²

Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Arabic-Speaking English Language Learners (ELLs) Compared to Children with Developmental Language Disorder (DLD)

The number of English Language Learners (ELLs), or children whose first language (L1) is not English and who attend schools taught in English, is significantly increasing in Canada (Paradis et al., 2010) and the United States (Goldstein, 2004). Identifying children with language impairment (LI) in culturally and linguistically diverse communities, such as in the United States and Canada, is challenging. A large body of research indicates that English traditional assessment tools are not sensitive in identifying language impairment among ELLs who are in the process of learning English as a second language (e.g., Paradis, 2005). Many studies have found that knowledge-based assessment tools such as English standardized tests of language tap existing language knowledge and experience and may not distinguish those with a developmental language impairment from ELLs who have more limited language knowledge than their monolingual peers (Chu & Flores, 2011; Sandberg & Reschly, 2011). On the other hand, processing-dependent measures probing the abilities supporting language learning may be less dependent on ELLs' linguistic knowledge. Recent studies have investigated the utility of processing-dependent tasks such as measures of verbal short-term memory in distinguishing ELLs from children with underlying language impairment (Kohnert et al., 2006; Paradis et al., 2013). The purpose of the current study is to examine whether assessing ELLs in their L1 or

² A version of this chapter has been submitted for publication (Balilah, A., & Archibald, L. M. D. (under review). *Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Arabic-Speaking English Language Learners (ELLs)*. *Bilingualism: Language and Cognition*).

native language can provide meaningful information to inform diagnostic decisions. Of further interest was to determine whether the tests of verbal short-term and working memory differentiate ELLs from children with underlying language impairment.

Children with significant limitations in their language ability despite otherwise typical neurological and socioemotional development, as well as average educational and experiential opportunities are referred to as children with Developmental Language Disorder (DLD; also known as Specific Language Impairment; Bishop et al., 2016; Leonard, 2014). The language deficits in children with DLD can affect all areas of language (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). Although the profile of language deficits can be unique for each child with DLD, lexical, morphosyntactic, and to a lesser extent, pragmatic development are all expected to be impaired in children with DLD. Grammatical deficits, in particular, have been described as a hallmark deficit in DLD (Leonard et al., 1997). To identify children with DLD, speech-language pathologists (SLPs) commonly use norm-referenced standardized tests that have been normed with a monolingual population. In many research studies, DLD is diagnosed when two or more composite scores more than 1.25 *SDs* below the standardized mean are achieved based on five norm-referenced standardized tests of receptive and expressive language, in three domains of language (vocabulary, grammar, and narration) (Bishop et al., 2016; Tomblin et al., 1996).

Another group of children who may appear to have weak language skills at school is English Language Learners (ELLs), that is, those children who are receiving instruction in their second language (L2 English) or in a language other than their minority first language (L1). Research suggests that it can take 4 or 5 years for ELLs to gain English proficiency comparable to their monolingual peers (Hakuta et al., 2000). According to Paradis (2010), there is

considerable overlap in the linguistic features between typically developing ELLs (TD ELLs) who are in the early stage of developing their L2 (within the first two years in particular) and monolingual children with DLD, as both groups tend to have errors in vocabulary choice and grammatical morphemes (Tabors, 2008). Further, receiving instruction in English can also impact ELLs' L1 learning. Minority L1 children often receive minimal community support in their L1, and the opportunities to hear and use their L1 is diminished once they start schooling (Anderson, 2012; Paradis et al., 2010). As proficiency in ELLs' L2 grows, ELLs' skills in L1 often do not develop further or even reduce and diminish across time, a phenomenon termed *L1 incomplete acquisition* or *L1 loss* (Anderson, 2012; Paradis, 2010). L1 loss impacts lexical and grammatical systems (Anderson, 2012), two areas of language affected by DLD as described above.

As a result of being in the early stages of English acquisition and potential L1 loss, ELLs may have weak language skills in each of the languages they are learning, which poses challenges when concerns regarding language development and language learning arise. Several studies report that SLPs commonly use English norm-referenced standardized tests in order to assess ELLs' linguistic abilities (Caesar & Kohler, 2007; Gillam et al., 2013). Recent evidence suggests that administering knowledge-based assessment tools such as English standardized language tests and interpreting scores based on monolingual norms may lead to over-diagnosis of language impairment among ELLs (Donovan & Cross, 2002; Gutiérrez-Clellen, 1996; Klingner & Artiles, 2003). That is, linguistic measures tap existing linguistic knowledge, which, in ELLs, is influenced by their limited English language knowledge and experiences (Campbell, Dollaghan, Needleman, & Janosky, 1997). As an alternative to English testing of ELLs, some researchers have suggested that conducting assessments in ELLs' L1 or native language can

provide meaningful information regarding ELLs' language abilities (Chu & Flores, 2011; Wagner et al., 2005). Even assessing ELLs in their minority L1, however, may not be the best assessment approach to understanding the child's language skills. As will be shown below, ELLs tend to perform in a range similar to that of monolingual children with language impairment on knowledge-based measures of vocabulary and grammar leading to confusion as to whether the poor ELL scores are due to inexperience or impaired skills.

Vocabulary tests are commonly used by SLPs to assess language learning ability in children (Lezak, Howieson, & Loring, 2004). For ELLs, however, lexical-semantic knowledge is often distributed across languages with, for example, some vocabulary items being experienced mostly at school in English and other vocabulary items experienced mostly at home in the child's L1 (Gollan et al., 2008; Pearson et al., 1993; Umbel et al., 1992). The lower frequency of exposure and practice for individual words may result in weaker links between semantic and phonological representations in ELLs (Gollan et al., 2008). As a result, even TD ELLs have been found to score below their monolingual peers on vocabulary measures in both English (e.g., Bialystok, Luk, Peets, & Yang, 2010) and their L1 (Jackson, Schatschneider, & Leacox, 2014). Indeed, on single language vocabulary measures, TD ELLs often show performance comparable to monolingual children with language impairment (Umbel, Pearson, Fernandez, & Oller, 1992; Windsor & Kohnert, 2004). Therefore, using standardized vocabulary tests, which are affected by the amount of exposure to the two languages, may not reflect the full range of ELLs' vocabulary knowledge (Pearson et al., 1993). As a result, assessing single-language vocabulary knowledge in ELLs can increase the risk of overidentification of language impairment (Kohnert, 2010), even when testing is done in the child's stronger language (Kan & Kohnert, 2005). Completing assessments in each language to which the child has been exposed and considering

the ‘sum’ of lexical-semantic knowledge across both languages holds potential (Peña, Bedore, & Kester, 2016). However, given the diverse first-language background of ELLs, such an approach would place high demands on test development (Paradis et al., 2013).

Several studies have examined the utility of grammatical language tasks in discerning TD ELLs from monolingual children with DLD. For example, Paradis (2005) compared the accuracy of verb morphology in monolingual children with DLD and TD ELLs from multiple background languages in both spontaneous and elicited speech. The findings indicated that TD ELLs and monolingual children with DLD did not differ in the accuracy rate and error pattern of various grammatical morphemes, including: third person singular [-s], past tense [-ed], irregular past tense, BE as a copula and auxiliary verb, DO as an auxiliary verb, progressive [-ing], prepositions in and on, plural [-s], and determiners *a* and *the*. Similarly, Paradis, Rice, Crago, and Marquis, (2008) found significant overlap in the accuracy of verb morphology in TD ELLs and monolingual children with DLD. It is clear that ELLs’ performance on grammatical morphemes is affected by ELLs’ limited knowledge and experience with the target language. As ELLs are in the process of developing their morphosyntax, considerable overlap in the linguistic features between ELLs and monolingual children with DLD would be expected.

Although ELLs have been found to have weak vocabulary and grammatical skills possibly in both their L1 and L2, one language domain that may be positively impacted by bilingual learning is phonology. Phonological skills may benefit from cross-linguistic transfer, that is, phonological knowledge acquired in one language may support the learning of the phonological system of a newly acquired language (Chitiri et al., 1992; Wade-Woolley & Geva, 2000), especially when the phonological systems of the two language overlap (Gottardo, Yan, Siegel, & Wade-Woolley, 2001; Lindsey, Manis, & Bailey, 2003). One phonological skill

important to school learning is phonological awareness, the ability to manipulate sounds in the language. In particular, phonological awareness in early years is highly predictive of later word reading skills (Lafrance & Gottardo, 2005). Basic word reading tasks tap phonological awareness skills, and may be a less-biased linguistic measure for ELLs (Oller, Pearson, & Cobo-Lewis, 2007; Paradis, Genesee, & Crago, 2011). Indeed, bilingual children have been found to perform as well as monolingual children on basic word-reading and phonological awareness tasks. For example, Oller et al. (2007) found that TD ELLs do not differ from their TD monolingual peers on measures of basic word-decoding, especially when the language of testing is the same as the child's language of literacy instruction. In addition, Bialystok et al. (2003) found that bilingual children have at least equivalent performance when compared to their monolingual peers on phonological awareness. It should be noted that the cross-language influences of phonological awareness skills are most evident during the early stages of L2 development. However, in later stages of L2 development, more advanced oral language skills are necessary for reading achievement (Genesee et al., 2005). Moreover, well-developed phonics knowledge, such as knowledge of all phonemic elements and the phonemic-graphemic mapping that is essential for reading, are required to enhance cross-linguistic transfer (Oller et al., 2007). Nevertheless, measures of phonological processing skills hold promise for discriminating ELLs from those with DLD.

Given the need for effective language assessments of ELLs, attention has turned to the use of processing-dependent measures that are known to be highly sensitive to the language abilities that differentiate monolingual groups with and without language impairment. Investigations of the cognitive processes that underpin some of the language learning difficulties in children with DLD have implicated deficits in general cognitive processes such as processing

speed, temporal integration, and working memory (Miller et al., 2001; Windsor & Kohnert, 2004). However, deficits in immediate memory, in particular, have been reported consistently in children with DLD (Archibald & Gathercole, 2006b; Ellis Weismer, 1996; Montgomery, 1995, 2000). As might be expected, considerable research attention has focused on processing-dependent measures, with the idea that such tasks may be less dependent on ELLs' linguistic knowledge and, therefore, directly tap abilities underlying language learning (Kohnert et al., 2006; Paradis et al., 2013). Focusing on processing-dependent measures may then help discriminate between ELLs with and without language impairment.

A number of studies have reported deficits in two aspects of immediate memory in DLD: verbal short-term memory and working memory (Archibald & Gathercole, 2006b; Henry et al., 2012). Short-term memory tasks engage temporary storage; verbal versions generally impose serial recall of words, letters or digits, whereas visuospatial versions impose storage of visual patterns or sequences of movement (Baddeley, 2000; Conway et al., 2005). Verbal short-term memory has been found to be a key indicator of new word learning (Majerus et al., 2006; Masoura & Gathercole, 2005), and vocabulary acquisition (Gathercole, Willis, Emslie, & Baddeley, 1992; Gathercole et al., 1992). Working memory tasks, on the other hand, impose processing demands in addition to storage, and are generally assessed by complex memory span paradigms (Engel de Abreu, 2011). Examples of verbal complex span tasks are counting recall and backwards digit recall, in which a participant recalls numbers after counting or reversing the order, respectively. Examples of corresponding visuospatial tasks involve recalling locations or orientations after identifying a different shape or mentally rotating an image, respectively (Alloway et al., 2009). Working memory has been associated with complex cognitive activities, such as language comprehension and word decoding (Cain et al., 2004; Engel de Abreu &

Gathercole, 2012). It should be noted that previous research has reported comparable performance between monolingual children with DLD and TD peers on visuospatial short-term and working memory measures (e.g., Archibald & Gathercole, 2006b) suggesting that DLD deficits in immediate memory primarily involves the verbal domain (Archibald & Gathercole, 2006b).

Given that short-term and working memory measures emphasize the storage and processing of new information (Engel de Abreu et al., 2013), the influence of previous knowledge has been considered to be minimal. It has been suggested that processing-dependent measures such as verbal short-term memory and working memory measures may pose similar challenges and be equally familiar (or unfamiliar) to all children, regardless of the language they speak (Engel de Abreu & Gathercole, 2012). It should be noted that the majority of research comparing ELLs with monolingual children with DLD on processing-dependent measures have focused on nonword repetition measures, a task involving the immediate recall of made up or nonsense words and considered to tap verbal short-term memory. Given that the phonological forms of the nonwords are novel to all participants, this task is expected to minimize the role of prior language knowledge and experience. Accumulated evidence from ELL studies, however, shows that even previous sublexical phonological knowledge and experience can influence children's performance. For example, Kohnert and colleagues (Kohnert et al., 2006; Windsor et al., 2010) found that the performance of TD ELLs and monolingual English-speakers with language impairment did not differ on English nonword repetition tasks. Although nonword repetition tasks are considered a less biased form of assessment than knowledge-based measures (Paradis et al., 2013), nonword repetition does not completely eliminate the effect of children's experience with the target language (Kohnert et al., 2006).

To the best of our knowledge, there are no studies comparing ELLs with monolingual children with DLD on verbal complex memory measures. Evidence from research on typically developing individuals, however, indicate that using such tasks may minimize the role of linguistic knowledge and experiences in the ELL population. Indeed, equivalent performances have been reported for bilingual children when compared to monolingual peers in working memory measures (Cockcroft, 2016; Engel de Abreu, 2011; Engel de Abreu et al., 2013). In fact, other studies have reported better working memory abilities in bilingual children compared to their monolingual peers (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004). It has been suggested that the constant management of the two language systems in the bilingual mind may enhance executive functions such as working memory (Bialystok, 2001). According to Engel de Abreu (2011), the mixed findings regarding the bilingual advantage on working memory tasks across studies might arise from the fact that studies have used different tasks, which may additionally probe different underlying cognitive processes.

Importantly, the nature of the stimuli and the type of underlying cognitive processing that might be involved in different verbal short-term and working memory tasks is another factor that may account for the considerable differences observed in the ELLs' performance. That is, evidence from research on typically developing individuals indicates that verbal short-term and working memory involving highly familiar lexical stimuli may minimize the role of linguistic knowledge and experiences in ELLs. For example, Engel de Abreu et al. (2013) reported an advantage for monolingual over bilingual children on the nonword repetition task, but not on number-based verbal short-term and working memory measures. According to Engel de Abreu et al. (2013), as the number-based measures involve recall of highly familiar lexical stimuli, such tasks are considered to be equally familiar to all children and are generally acquired at an early

age. As a result, children's performance on these tasks are less affected by verbal long-term memory. This finding suggests that short-term and working memory tasks involving recall of highly familiar verbal materials such as number words may be an effective assessment tool that reduces assessment bias in the ELL population.

It should be noted that in any consideration of bilingual development, the specific languages being learned must be considered. The present study was concerned with the development of Arabic (L1) – English (L2) learners. Arabic is a Semitic language with a nonconcatenative morphology. The morphology, phonology, and orthography of Semitic languages are distinct from Indo-European languages such as English. Arabic has 28 consonants and 6 vowels. Arabic is a root and pattern language with complex interaction between syntax, morphology and phonology. Word roots mostly consist of three consonants that represent the lexical meaning (CCC; trilateral root; Beeston, 1970), and the pattern is primarily composed of vowels inserted between the root consonants. The roots carry a semantic meaning shared to various degrees by the derivative words associated with the same root (Bakalla, 1979). Moreover, the verbal inflection system of Arabic is relatively rich. Verbs are morphologically inflected for tense, and mood, and the verb should agree with the subject for aspects of person (first, second, and third), number (singular, dual, and plural), and gender (feminine and masculine) (Bakalla, 1979). Arabic is a transparent language, meaning it has a one-to-one relation between graphemes and phonemes. Interestingly, findings from a study that examined L1 and L2 reading skills in Arabic-speaking ELLs who were taught to read and write in Arabic and English found that phonological transfer can occur even between these two phonologically and orthographically distinct languages (Abu-Rabia & Siegel, 2002). However, it should be noted that only a few studies have focused on Arabic – English bilinguals, especially in regards

to their performance in relation to Arabic children with language impairments. Comparing Arabic – English bilinguals to Arabic children with language impairments is important in order to examine whether or not the well-established pattern of similar performance across these groups can occur in knowledge-based measures as compared to processing measures.

The present study compared the linguistic and cognitive performance of Arabic-speaking children (ELLs) to two monolingual peer groups: 1) typically developing Arabic-speaking children (A-TD), and 2) Arabic-speaking children with DLD (A-DLD). As the majority of the studies that use knowledge-based assessment tools have focused on assessing ELLs on their second language (English) (Paradis et al., 2013), it is relevant for clinical practice to examine whether conducting assessments in ELLs' L1 or native language can provide meaningful information to inform diagnostic decisions about ELLs. Moreover, given the shortcomings of knowledge-based measures in differentiating the language performance profiles of children with DLD and ELL, it is important to examine the diagnostic power of verbal short-term and working memory measures in differentiating ELLs from children with underlying language impairment. A bias in favour of the A-TD group was expected for the L1 Arabic measures of vocabulary and general language, and further, these knowledge-dependent language measures were not expected to differentiate the ELL and A-DLD groups. Performance on a basic word reading task, however, was expected to reveal a potential phonological strength in the ELL group relative to the A-DLD group, and potentially the A-TD group as well. For the processing-dependent immediate memory tasks, at least equivalent performance by ELL and A-TD groups was expected, and higher scores by the ELL than A-DLD groups. However, this latter predication was expected to be modified by the verbal demands of the task such that tasks with higher verbal demands (e.g., nonword repetition) would be less likely to differentiate the three groups than those with low verbal

demands (e.g., digit recall) or no verbal demands (e.g., visuospatial short-term or working memory tasks).

Methods

Participants

There were 480 children ($M_{\text{age}}=7;9$, $SD=1.12$; 187 males) participating in three groups in this study: (a) 59 unselected ELLs whose L1 was Arabic and who were learning English as the language of instruction (English L2) in Canada ($M_{\text{age}}=7;11$, $SD=1.16$; 29 males), (b) 369 typically developing monolingual Arabic-speaking children (A-TD) from Saudi Arabia ($M_{\text{age}}=7;11$, $SD=1.12$; 139 males), and (c) 52 monolingual Arabic-speaking children with DLD (A-DLD) from Saudi Arabia ($M_{\text{age}}=8;4$, $SD=1.00$; 19 males). All the children who participated in this study ranged from grade 1 to grade 4 (i.e., children 6-9 years of age). Children in the ELL group were recruited from a school providing instruction in both English and Arabic ($n=27$), or from an extra curricular Arabic instruction class for children receiving regular schooling in English ($n=32$). All of the ELLs were from homes in which Arabic was the first language. Children in the Arabic-speaking samples were recruited from 10 schools in Saudi Arabia (Jeddah) based on a study invitation sent home to all parents of children in the relevant grades. No group differences were found in sex distribution, $\chi^2(2) = 2.964$, $p > .05$, or age, $F(3, 476) = .608$, $p > .05$.

In order to identify which of the Arabic speaking children to include in the A-DLD group, the following criteria were applied based on the norms reported for this group in chapter 2 of this thesis: (1) Scores of at least 1 SD below the standardized mean on 2 of 4 language measures including the 3 subtests of the *Arabic Language Test (ALT)*; Shaalan, 2010) and the *Arabic Picture Vocabulary Test (APVT)*; Shaalan, 2010), and (2) a standard score not lower than

86 on the *Test of Non-verbal Intelligence (TONI-3; Brown, Sherbenou, & Johnson, 1997)*. Table 9 provides descriptive statistics for criterion measures for all groups. All of these tests were included as study measures and are described below. The Arabic measures were chosen to identify the children with DLD because these measures provide a description of the children's performance across receptive and expressive modality and content areas. Information regarding the children's expressive and receptive language abilities can promote a more sophisticated approach to evaluate the children's general language ability and identify children with DLD (Stark & Tallal, 1981).

Table 9.
Descriptive statistics for criterion measures for all groups.

Measure	Score	Participant group					
		ELLs		A-TD		A-DLD	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Arabic Language Test:							
Sentence Comprehension (A)	RS	19.30	2.89	20.23	1.87	18.83	2.13
	SS	94.86	20.84	101.66	13.75	88.19	17.70
Sentence Comprehension (B)	RS	13.27	3.51	13.85	2.65	11.44	2.37
	SS	98.66	18.87	102	14.14	85.80	12.70
Expressive Language (A)	RS	16.50	5.41	18.77	2.92	15.15	3.33
	SS	92.43	25.04	102.45	13.08	82.65	15.92
Expressive Language (B)	RS	25.34	10.86	35.92	5.18	30.44	4.49
	SS	72.09	31.08	102.52	13.63	82.13	11.22
Sentence Repetition (A)	RS	44.98	11.35	51.73	3.91	49.10	6.91
	SS	79.27	34.17	101.04	13.52	92.63	21.36
Sentence Repetition (B)	RS	18.23	12.76	47.64	9.96	34.63	8.29
	SS	57.45	22.21	102.78	13.23	80.26	11.02
Arabic Picture Vocabulary Test	RS	70.32	25.95	102.73	10.72	91.98	13.54
	SS	58.43	32.71	102.13	13.34	84.87	17
Nonverbal intelligence (TONI-3)	RS	15.57	6.39	12.80	5.74	11.29	3.54
	SS	102.80	12.89	95.66	12.61	89.65	10.45

Note. RS = raw score; SS = standard score ($M = 100$, $SD = 15$).

Materials and Procedure

All children completed a battery of assessment measures individually in a quiet room in their school over 4 weekly sessions each of approximately 40 minutes. The battery included measures of vocabulary, (*Arabic Receptive-Expressive Vocabulary Test, AREVT*, El-Halees & Wiig, 1999; *Arabic Picture Vocabulary Test, APVT*, Shaalan, 2010), language (*Arabic Language Screening Test, ALST*, El-Halees & Wiig, 1999; *Arabic Language Test, ALT*, Shaalan, 2010), reading (*Arabic Word Reading Task*), verbal short-term and working memory (*Arabic Nonword Repetition Task, A-NWR*, Shaalan, 2010; *Automated Working Memory Assessment, AWMA*, Alloway, 2007), and non-verbal intelligence (*Test of Non-Verbal Intelligence, TONI-3*, (Brown et al., 1997). The tests were administered in a fixed order so that each individual session involved the administration of particular tests. In session 1, the AREVT, A-NWR, and ALST were administered. In session 2, the ALT was administered. Session 3 included the APVT, TONI-3, and Arabic Word Reading Task. In session 4, the AWMA and other tasks not reported here for the ELL group were administered. A trained native Arabic speaker tested the children in the battery of assessment measures. Parents completed a questionnaire at the time of completing the study consent form.

Vocabulary tests. Two Arabic vocabulary tests were administered, the *Arabic Receptive-Expressive Vocabulary Test (AREVT)*, El-Halees & Wiig, 1999), and the *Arabic Picture Vocabulary Test (APVT)*, Shaalan, 2010). In the receptive subtest of the AREVT, the child was asked to point to a picture corresponding to a given spoken word from a choice of four. In the expressive AREVT subtest, the child named or described a picture with a single word or phrase. The total number of correct responses was counted for each subtest with the maximum possible score dependent on the child's age for each subtest (6;0 to 6;11: $n=40$; 7;0 to 7;11: $n=45$; 8;0 to

8;1: $n=50$; 9;0 to 9;11: $n=55$). In the APVT, which is a measure of receptive vocabulary, the child was asked to point to a picture corresponding to a given spoken word from a choice of four pictures with a maximum possible score of 132. High test retest reliability has been reported for the APVT = .97 (Shaalán, 2014).

Language tests. Two Arabic language tests were administered: *the Arabic Language Screening Test (ALST)*, El-Halees & Wiig, 1999), and *The Arabic Language Test (ALT)*, Shaalan, 2010). The tests of verbal abilities from the ALST involved spoken responses to 6 items, each assessing nouns and verbs, adjectives, morphology, understanding sentences, forming sentences, remembering instructions, and repeating sentences. The majority of subtests were scored with 1 point for each correct response with the exception of Repeating Sentences (2=correct; 1=1-2 errors of omission, addition, or substitution; 0=3 or more errors, or no response), Formulating Sentences (2=correct sentence; 1=few errors; 0=nonsense or unrelated sentence, or no response), and Adjectives (1=point correct pointing or correct naming; 0=incorrect pointing or naming, or no response). The highest possible overall raw score in the Verbal Abilities section was 60.

The Non Verbal Abilities subtest of the ALST consisted of 5 short sections of 5 or 6 items and required verbal responses, including the Missing Part activity, the Matching Words/Sentences, the Classification by Meaning, the Classification by Group Membership, and Arranging a Story. The majority of the subtests were scored with 1 point for each correct response, with the exceptions being the Classification by Group Membership (2=correct pointing to the related pictures and correct describing of the relationship among the group members; 1=correct pointing or describing; 0=incorrect pointing and describing, or no response) and Missing Part activity (2=correct pointing and naming; 1=correct pointing or correct naming; 0=incorrect pointing and naming, or no response). The highest possible overall raw score in the

Non Verbal Abilities section was 40 with a maximum possible score of 100 for the overall Arabic Language Screening Test.

For the Arabic Language Test (ALT), the task includes 3 subtests each divided into a first and second section (Section A and B). In the *Sentence Comprehension* subtest, the child had to point to a picture that corresponded to the spoken sentence from a choice of three or four pictures. In the *Expressive Language* subtest, the child was given a sentence and was asked to generate a spoken word or phrase in reference to a picture cue. In the *Sentence Repetition* subtest, the child heard an audio recording of sentences spoken by a native, adult male Arabic speaker and was asked to repeat them. The total number of correct responses was counted for each subtest, with a maximum possible score of 40 for the Sentence Comprehension subtest (Section A: $n=22$; Section B: $n=18$), and 68 for the Expressive Language subtest (Section A: $n=24$; Section B: $n=44$). The 41 items of the Sentence Repetition subtest were scored on a 4-point scale (3=correct; 2=1 error; 1=2-3 errors; 0=4 or more errors, or no response), with a maximum possible score of 123 (Section A: $n=18$; Section B: $n=23$). High test retest reliability has been reported for the three subtests of the ALT (the Sentence Comprehension Test = .95; the Expressive Language Test = .95, and the Sentence Repetition Test = .97) (Shaalán, 2014).

Reading efficiency. In the *Single Word Reading Task*, participants read aloud as many printed words as possible within 45-seconds (maximum score=104). The stimuli included in the task were selected from a commonly used single word list taken from a popular reading series in Lebanon (Oweini & Hazoury, 2010). The 104 Arabic sight words presented became progressively more difficult based on frequency (beginning with very frequently occurring words and progressing to less frequently occurring words), and number of syllables (beginning with one-syllable words and progressing up to six-syllable words). Since standard Arabic text is

currently written without diacritic marks (Zayyan et al., 2016), which are marks added to the original Arabic alphabet to modify word pronunciation (Lutf, You, Cheung, & Chen, 2013), words that could not be pronounced correctly without diacritic marks were not selected for this task. Moreover, words reflecting the unique properties of Lebanese culture were not selected for this task. For the AREVT, ALST, and single word reading measures, raw scores were converted to standard scores based on the normative data available reported in chapter 2 of this thesis.

Short-term and working memory. Eight subtests from the *Automated Working Memory Assessment* (AWMA, Alloway, 2007) were administered. Measures of verbal short-term memory (*Digit Recall; Word Recall*) required the immediate repetition of numbers or word forms. Measures of verbal working memory (*Counting Recall; Backwards Digit Recall*) required the recall of numbers after counting or reversing the order, respectively. In addition, four visuospatial short-term and working memory subtests from the AWMA were administered. Measures of visuospatial short-term memory (*Dot Matrix; Block Recall*) required the recall of locations. Measures of visuospatial working memory (*Odd One Out; Spatial Span*) required the recall of locations or orientations after identifying a different shape or mentally rotating an image, respectively. For the two monolingual Arabic groups, the AWMA was administered to each child using Arabic. For the ELL group, the AWMA was administered to each child using the child's preferred language (Arabic or English). Of the participants, 70% preferred English and 30% preferred Arabic.

One additional verbal short-term memory task was administered, the *Arabic Nonword Repetition Task* (A-NWR, Shaalan, 2010). In the A-NWR, the child repeated the nonwords presented via audio recording of a native, adult male Arabic speaker. Items taken from Shaalan (2010) consisted of 48 nonwords varying in length (two to three syllables) and cluster type (no

cluster, medial cluster, final cluster, and medial and final clusters). Each response was scored online as correct or incorrect by a trained research assistant with a maximum possible score of 48. Raw scores were converted to standard scores based on the Arabic normative data reported in chapter 2 of this thesis. For all of the subtests of AWMA, and the A-NWR task, raw scores were converted to standard scores based on the normative data reported in chapter 2.

Nonverbal intelligence. The *Test of Non-Verbal Intelligence (TONI-3, Brown, Sherbenou, & Johnsen, 1997)* was administered to measure general nonverbal cognitive abilities. In the TONI-3, children chose a picture to complete a visual pattern. Raw scores of the TONI-3 were converted to the standard scores based on published test norms.

Parent questionnaire. The parent questionnaire included questions related to maternal level of education. Parents were asked to check the highest level of education attained by the child's mother. The descriptors included some high school, completed high school, some college, completed college, some university, and completed university. Responses were transposed to a 3-point scale with 1 corresponding to some/completed high school, 2 to some/completed college, and 3 to some/completed university. By parent report, approximately 80% of mothers had at least some college or university education in the ELL group. In comparison, approximately 58% of mothers had at least some college or university education in the two monolingual groups.

In addition, parents in the ELL group only filled out a questionnaire about their child's language background (Kaushanskaya et al., 2010). Parents were asked to provide information related to their child's language immersion, history, use, and the parent's rating of their child's current language abilities in each language (on a scale from 0 = none to 10 = perfect). All parents reported that their children acquired Arabic as a first language from birth and began exposure to English, on average, at the age of 3;3 ($SD = 2;0$, range = 8 – 96 months). Notably, the parents of

6 participants did not indicate the time when their child was first exposed to English. As well, parents rated their child's current language abilities (speaking and understanding) as very good for both Arabic: ($M = 8.00$, $SD = 2.03$), and in English: ($M = 8.00$; $SD = 2.11$). None of the parents rated their child's current speaking and understanding abilities 3 (low) or lower in both Arabic and English languages. In addition, complete data were available for all but 3 children from the ELL group who did not complete all of the Arabic language tasks.

Results

Arabic Vocabulary, Language, and Reading Measures

Descriptive statistics comprised of the raw and standard scores on for the Arabic vocabulary, language, and word reading measures for the three groups, ELLs, A-TD, and A-DLD, are provided in Table 10. (Subtest scores for the vocabulary and language measures can be found in Table 9). The performance of the A-DLD and ELL groups were lower than the A-TD group on all measures whereas the performance of the ELL group was almost comparable to, or in some cases lower than, the A-DLD group.

Table 10.

Standard and raw scores on all language measures for the ELLs, A-TD, and A-DLD groups.

Measure	Score	Participant group					
		ELLs		A-TD		A-DLD	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Vocabulary (AREVT)	RS*	72.91	15.09	80.00	13.11	81.52	12.01
	SS	85.89	21.36	101.23	11.54	91.20	12.84
Language (ALST)	RS	49.38	16.88	63.14	12.12	57.12	11.66
	SS	86.16	16.73	101.24	12.74	91.15	13.41
Single Word Reading	RS	16.05	15.37	40.77	16.53	36.40	13.89
	SS	65.88	30.94	100.96	14.84	93.21	13.96

Note. RS = raw score; SS = standard score ($M = 100$, $SD = 15$). AREVT = the Arabic Receptive-Expressive Vocabulary Test; ALST = the Arabic Language Screening Test. * The maximum raw score of AREVT differed across age groups and was dependent on the child's age.

A multivariate analysis of variance (MANOVA) was completed to investigate group differences on the standard score of these language-related measures (AREVT, ALST, and Single Word Reading). Results revealed a significant group effect, Hotelling's T , $F(6, 942) = 42.54$, $P < .001$, $\eta^2_p = .213$. In addition, all of the univariate group comparisons were significant: AREVT, $F(2,474) = 41.41$, $p < .001$, $\eta^2_p = .149$, ALST, $F(2,474) = 39.57$, $p < .001$, $\eta^2_p = .143$, and Single Word Reading, $F(2,474) = 99.39$, $p < .001$, $\eta^2_p = .295$. Pairwise comparisons revealed significantly lower scores for the ELL and A-DLD groups than the A-TD group on AREVT, ALST, and Single Word Reading ($p < .001$). No significant differences were found between the ELL and A-DLD groups on either the AREVT or ALST ($p > .05$, all cases). The ELL group, however, had lower scores than the A-DLD group on the Single Word Reading Task only ($p < .05$). It should be noted that in a corresponding ANCOVA with maternal education as a covariate, the same pattern of results was observed for all the language tasks with one exception: the ELL group had significantly lower scores than the A-DLD group on the AREVT task ($p < .05$), after statistically controlling for the mothers' level of education.

Verbal Short-term and Working Memory

Table 11 provides descriptive statistics for the raw and standard scores on the verbal short-term and working memory subtests of AWMA (Digit Recall, Word Recall, Counting Recall, and Backwards Digit Recall) and the A-NWR task for the three groups, ELLs, A-TD, and A-DLD. The performance of the A-DLD group was lower than the A-TD and ELL groups on all measures whereas the performance of the ELL group was similar to, or numerically higher than, the A-TD group (except on the nonword repetition task, A-NWR).

Table 11.
Descriptive statistics of the verbal short-term and working memory subtests of AWMA and A-NWR task raw and standard scores for the ELL, A-TD, and A-DLD groups.

Measure	Score	Participant group					
		ELLs		A-TD		A-DLD	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Digit Recall	RS	23.09	5.22	21.37	4.08	19.60	3.89
	SS	107.64	18.21	101.01	14.70	92.83	14.82
Word Recall	RS	20.23	3.76	19.95	4.20	17.44	5.26
	SS	102.63	12.53	101.34	14.06	90.47	17.50
Counting Recall	RS	13.84	4.07	14.15	4.98	13.65	4.71
	SS	100.43	12.48	100.61	15.01	95.66	13.78
Backwards Digit Recall	RS	10.07	3.34	9.58	4.40	8.37	4.41
	SS	102.77	10.26	100.80	14.84	94.35	14.55
Arabic-Nonword Repetition	RS	34.20	8.62	39.23	6.65	34.81	7.88
	SS	90.96	17.87	101.32	13.97	90.61	18.14

Note. RS = raw score; SS = standard score ($M = 100$, $SD = 15$).

Results of the multivariate analysis of variance (MANOVA) completed on the standard scores of the verbal short-term and working memory measures (A-NWR, Digit Recall, Word Recall, Counting Recall, and Backwards Digit Recall) with group (ELL, A-TD, A-DLD) as a between-groups factor revealed a significant group effect: Hotelling's T , $F(10, 938) = 8.19$, $P < .001$, $\eta^2_p = .080$. Significant group effects were observed in the univariate comparisons for Digit Recall, $F(2,474) = 12.91$, $p < .001$, $\eta^2_p = .052$, Word Recall, $F(2,474) = 13.97$, $p < .001$, $\eta^2_p = .056$, Backwards Digit Recall, $F(2,474) = 5.51$, $p < .001$, $\eta^2_p = .023$, A-NWR, $F(2,474) = 20.67$, $p < .001$, $\eta^2_p = .080$, but not for Counting Recall, $F(2,468) = 2.63$, $p = 0.073$.

Pairwise comparisons of the verbal short-term and working memory measures (A-NWR, Digit Recall, Word Recall, Counting Recall, and Backwards Digit Recall) revealed significantly higher scores for the ELL group compared to the A-TD group on the Digit Recall subtest ($p = .007$), whereas no significant differences were found between both groups on the remaining AWMA subtests ($p > .05$ in all cases). For the A-NWR task, however, the ELL and A-DLD groups had significantly lower scores than the A-TD group ($p = .001$), and there was no

significant difference between the ELL and A-DLD groups ($p > .05$). The performance of the A-DLD group was significantly lower than the A-TD and ELL groups on Digit Recall, Word Recall, and Backwards Digit Recall ($p < .05$ in all cases), whereas no significant difference between the three groups in the Counting Recall subtest was found ($p > .05$). It should be noted that in the corresponding ANCOVA with maternal education as a covariate, the same pattern of results was observed for all of the verbal short-term and working memory subtests of AWMA (Digit Recall, Word Recall, Counting Recall, and Backwards Digit Recall), and the A-NWR task.

Visuospatial Short-term and Working Memory

Descriptive statistics comprised of the raw and standard scores on visuospatial short-term and working memory subtests of AWMA (Dot Matrix, Block Recall, Odd One Out, and Spatial Span) for the three groups, ELLs, A-TD, and A-DLD, are provided in Table 12. The three groups had almost identical performance on all visuospatial short-term and working memory subtests.

Table 12.

Descriptive statistics of the visuospatial short-term and working memory subtests of AWMA raw and standard scores for the ELL, A-TD, and A-DLD groups.

Measure	Score	Participant group					
		ELLs		A-TD		A-DLD	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Dot Matrix	RS	18.85	4.73	18.57	4.24	18.25	3.54
	SS	101.78	15.85	100.42	15.22	97.02	12.51
Block Recall	RS	18.32	3.87	16.99	5.09	17.52	5.70
	SS	105.03	10.88	100.15	14.91	98.90	15.27
Odd One Out	RS	15.66	5.07	14.89	4.91	16.71	3.99
	SS	102.15	15.56	99.56	15.45	103.11	10.28
Spatial Span	RS	13.00	4.34	13.11	4.98	12.81	4.02
	SS	100.44	12.01	100.29	15.32	97.97	11.84

Note. ZS = z-scores RS = raw score; SS = standard score ($M = 100$, $SD = 15$).

Results of the multivariate analysis of variance (MANOVA) investigating group differences on the standard score of the visuospatial short-term and working memory subtests of AWMA (Dot Matrix, Block Recall, Odd One Out, and Spatial Span) revealed no significant group effect: Hotelling's $T, F(8, 946) = 1.628, P > .05$. It should be noted that in a corresponding MANCOVA with maternal education as a covariate, the result was unchanged.

Discussion

This study compared the performance of Arabic-speaking ELLs on linguistic and cognitive measures to two monolingual peer groups: typically developing Arabic-speaking children (A-TD) and Arabic-speaking children with DLD (A-DLD). The primary objective of this study was to examine whether L1 measures (vocabulary, language, and reading), and cognitive measures (short-term and working memory measures) could discriminate between ELLs and monolingual peers with and without DLD. Results of this study revealed that the ELL group scored significantly more poorly on L1 measures (vocabulary, language, and reading) than the A-TD group. The performance of the ELL group was comparable to, or in some cases lower than, the A-DLD group on all of the Arabic language measures. With regards to the cognitive measures, however, no differences were found between the ELL and A-TD groups on the short-term and working memory measures, with the exception of the Arabic nonword repetition and counting recall tasks. The performance of the ELL group on the Arabic nonword repetition task was comparable to that of the A-DLD group and significantly lower than the A-TD group. Interestingly, the ELL group scored significantly higher than the A-TD and A-DLD groups on only one number-based verbal short-term memory measure.

The present study adds to a limited evidence base investigating L1 abilities in ELLs in comparison to monolingual children with DLD, and in particular, addresses the case of Arabic.

The finding of significantly lower scores on L1 knowledge-based language tasks (Arabic vocabulary and language) for the ELL group compared to the A-TD group is consistent with previous evidence suggesting that ELLs from minority ethnolinguistic communities, such as the ELL group in this study, are often at risk of losing and/or not fully acquiring their L1 (Paradis, 2010). The phenomena of L1 loss/incomplete acquisition in minority L1 children may result in an overlap in the language profiles of ELLs and children with DLD, particularly in lexical and grammatical skills (Anderson, 2012). Indeed, ELLs in the present study who were rated by parents as having at least adequate language skills in their L1 tended to score in the range considered to reflect language impairment in monolingual children on L1 vocabulary and language measures. Therefore, L1 knowledge-based measures that are strongly dependent on ELLs' opportunities or experience with the target language, such as the L1 vocabulary and language measures that were employed in the present study, do little to assist in differentiating language differences from language impairment. As a result, conducting assessments in ELLs' L1 or native language using knowledge-based measures to inform ELLs' diagnostic decision may not be clinically useful and may result in over diagnosis of language impairment in ELLs.

The ELL group in this study showed lower than expected performance on the Arabic Single Word Reading Task than the monolingual groups either with or without DLD. This finding appears to contradict other studies, where bilingual children performed as well as TD monolingual children on basic word-decoding tasks as a language-related measure that taps phonological awareness skills (e.g., Oller et al., 2007). It has been suggested that cross-linguistic transfer of phonological awareness, a skill strongly related to word reading, may support ELLs' performance on basic word-decoding tasks (Oller et al., 2007). However, well-developed phonics knowledge in a certain language is required to enhance the cross-linguistic transfer

(Oller et al., 2007). ELLs in this study attended school where English was the primary language of instruction. Knowledge of phonemic elements and phonemic-graphemic mapping, which are considered to be fundamental for reading in Arabic, may not be fully developed in the ELL group. The low performance by the ELL group on the Arabic reading task in this study may reflect the lack of specialized knowledge that is essential for reading in Arabic rather than reflecting an actual reading impairment in this population.

On all of the verbal memory tasks tapping short-term and working memory, except for the Arabic nonword repetition task, the performance of the ELL group was comparable to the A-TD group, whereas the performance of the A-DLD group was lower than the A-TD and ELL groups on the majority of these measures. These results, on the whole, are consistent with previous evidence suggesting that processing-dependent measures in ELLs are less sensitive to differences in language experience than knowledge-based measures (Engel de Abreu et al., 2013). The present findings regarding the reduced performance of the A-DLD group but not the ELL group on the majority of the verbal short-term and working memory subtests suggests that processing-dependent rather than knowledge-based measures may hold promise for differentiating between children with DLD and ELLs. To the best of our knowledge, there are no studies comparing ELLs to monolingual children with DLD on verbal complex memory measures. The current study, therefore, adds to the literature by showing that one verbal working memory subtest of the AWMA (Backwards Digit Recall), in addition to two verbal short-term subtests of the AWMA (Digit Recall and Word Recall), may be a viable option for reducing assessment bias in ELLs.

Importantly, the results of the verbal short-term and working memory measures in this study are consistent with previous evidence suggesting that the nature of the verbal stimuli

involved in verbal short-term and working memory tasks possibly account for the considerable difference observed in the ELLs' performance. Verbal short-term and working memory measures employed in this study that involve recall of highly familiar lexical stimuli, such as number words and basic words, appear to discriminate between ELLs and A-DLD groups. These tasks involve familiar lexical stimuli that are generally acquired at an early age by ELLs in both L1 and L2 may be equally familiar to all children and less affected by verbal long-term memory (Engel de Abreu et al., 2013). On the other hand, since nonword repetition tasks involve unfamiliar phonological form, it has been suggested that children's performance on this task relies on long-term phonological and lexico-semantic knowledge (Engel de Abreu et al., 2013). Indeed, the present study's finding adds to the growing body of evidence indicating that phonological structure and language experience impact ELLs' performance on nonword repetition tasks (Kohnert et al., 2006; Windsor et al., 2010). Unlike nonword repetition, therefore, verbal short-term and working memory tasks involving familiar lexical stimuli may be sensitive to the underlying differences between children with DLD and ELLs. Such measures may assist in differentiating language difference from language impairment. Moreover, the results indicated that not all processing-dependent measures are equally effective in reducing the role of prior knowledge or experience in ELLs. Searching for effective assessment measures in ELLs requires careful choice among verbal short-term and working memory measures.

The ELL group in this study scored significantly higher than the A-TD group on only the Digit Recall measure of verbal short-term memory. This finding is consistent with other studies that suggest that bilingual children may have some advantage on working memory tasks compared to monolingual children (Bialystok, Craik, Klein, & Viswanathan, 2004). This finding, however, conflicts with other studies that demonstrate no bilingual advantages in working

memory abilities (Bajo et al., 2000; Bialystok et al., 2008; Engel de Abreu, 2011). Moreover, the finding that the ELL group in this study demonstrated an advantage only on the Digit Recall subtest but not on other working memory measures that may share similar underlying processes is difficult to interpret. Given that there are only a few studies comparing working memory abilities between monolingual and bilingual children, additional research examining measures of verbal short-term and working memory in ELLs is warranted to gain a better understanding of the ELLs' advantage on working memory performance.

The A-DLD group in this study performed comparably with the ELL and A-TD groups on the Counting Recall measure of working memory. This finding is difficult to interpret, as previous research shows that children with DLD performed markedly lower on the Counting Recall measure of working memory (Archibald & Gathercole, 2006a). Particularly, the acquisition of counting knowledge and the counting strategies required by the Counting Recall subtest has been found to be sensitive to working memory (Archibald & Gathercole, 2006b). Given the lack of studies examining the language development and working memory abilities of Arabic-speaking children with DLD (Wiig & El-Halees, 2000), further investigation of the language and working memory deficits in this population would help to clarify this finding.

Finally, the ELL group in this study did not differ from their monolingual peers (A-TD and A-DLD) on all visuospatial short-term and working memory subtests. This finding is in line with previous studies that have shown no differences between TD and DLD groups on visuospatial short-term and working memory measures (Archibald & Gathercole, 2006c). This finding therefore provides substantial evidence that the immediate memory deficit in Arabic-speaking children with DLD primarily involves the verbal domain, a suggestion consistent with observations for monolingual English DLD speakers (Archibald & Gathercole, 2006b, 2006c).

Conclusion and Future Directions

In this study, the performance of 6-to-9 year old ELLs whose first language was Arabic and who had been learning English as the language of instruction in Canada was compared to two monolingual groups: typically developing Arabic-speaking children and Arabic-speaking children with DLD on linguistic measures of vocabulary, language, and reading, and cognitive measures of short-term and working memory. The primary objective of this study was to examine the diagnostic power of L1 linguistic measures (vocabulary, language, and reading), and cognitive measures (short-term and working memory measures) in discriminating between ELLs and monolingual peers with and without DLD. With the exception of the Arabic nonword repetition task, the performance of the ELL group was comparable to, or in some cases higher than, the A-TD group on all of the phonological short-term and working memory subtests, whereas the performance of the A-DLD group was lower than the A-TD and ELL groups on the majority of the phonological short-term and working memory subtests.

The significant overlap between the A-DLD and ELL groups on L1 knowledge-based measures indicates that using such measures may increase the risk of over diagnosis of language impairment among ELLs. This finding suggests that tasks that focus more on the cognitive processes that underlie language learning, rather than children's opportunities or experiences with the test language may provide a more accurate representation of ELLs' linguistic abilities. However, it is clear from the verbal short-term and working memory results in this study that not all processing-dependent measures are equally effective in reducing the role of prior knowledge or experience in ELLs. For example, the present study's findings add to the growing body of evidence that indicates that ELLs' performance on nonword repetition is affected by their previous sublexical phonological knowledge and experience in the target language (Thorn &

Gathercole, 1999; Kohnert et al., 2006; Windsor et al., 2010). Furthermore, the present study indicates that verbal short-term and working memory tasks involving familiar lexical stimuli may be valid assessment tools that minimize the role of linguistic knowledge and experiences and help distinguish between ELLs from children with underlying language impairment.

Chapter 5

General Discussion

This final chapter reviews the main findings of the three studies presented in this thesis and the implication of the findings. The first section considers the measurement of language ability and impairment in Arabic-speaking children. The second section discusses linguistic and cognitive measures in Arabic-speaking English Language Learners (ELLs). The third section investigates linguistic and cognitive measures in Arabic-speaking ELLs and monolingual Arabic-speaking children with and without Developmental Language Disorder (DLD). The last section summarizes the practical implications of the findings and presents the final conclusion.

The Measurement of Language Ability and Impairment in Arabic-Speaking Children

This thesis considers the case of Arabic-speaking children including monolingual Arabic-speaking children and Arabic-speaking ELLs. As reviewed in study 1, the majority of available language assessments have been developed for English-speaking children. The available Arabic measures have been largely based on English measures. There is a lack of well-documented normative data for different ages and for the range of ability in language development in Arabic-speaking children. Only a few studies have focused on the measurement of language ability and impairment in Arabic-speaking children (Wiig & El-Halees, 2000). One purpose of study 1 was to investigate one psychometric property of available Arabic tests, sensitivity to developmental change. Performance differences across age bands studied would indicate that the existing measures capture developmental language growth, whereas a lack of such age-related differences would suggest that the tests fail to capture crucial aspects of Arabic language development. A second aim of the study was to examine the consistency with which individuals are identified with low language skills across tests. Low performance across a number of measures would

increase the confidence with which individual participants could be considered to have a language impairment. The final goal of study 1 was to explore the possible impact of sex differences on the measures. Sex differences on language performance would signal the need for further investigation of this factor in language acquisition and consideration in test design.

Summary of findings. Study 1 investigated the performance of 421 monolingual Arabic-speaking children from 6 to 9 years of age on a battery of assessment measures including individual measures of sentence comprehension, expressive language, sentence repetition, receptive and expressive vocabulary, and nonword repetition. This is the first investigation of these available Arabic tests with Arabic-speaking children from Saudi Arabia. Age effects were observed for all measures except the nonword repetition task. The findings suggest that nonword repetition tasks may not be sufficiently challenging for the full age range of Arabic-speaking children. The length of all stimuli in the task are two- to three-syllable words, but this length may not be sufficiently challenging for our full age range of Arabic-speaking children given that some Arabic dialects can have up to seven-syllable words. In fact, many studies have shown that stimuli length affects children's performance on nonword repetition with older children reaching the ceiling for short nonwords (e.g., Archibald & Gathercole, 2006). This study also shows that the majority of available Arabic language measures are sensitive to developmental change in younger children only between the ages of 6 and 7, suggesting that these language measures became less challenging with age. In addition, when individuals exhibiting consistently low performance across the language measures were examined, there was a considerable lack of agreement among tasks making interpretation difficult. Sex differences in favour of male participants were observed for several of the language subtests raising questions about the impact

of sociocultural differences between males and females on language practice and development in Saudi Arabia.

Implications of findings. The results of study 1 shows that current Arabic language measures have significant limitations. The results suggest that there is a need to develop language assessment measures that include more complex language skills and a broader range of language abilities to capture the level of language development in older Arabic children, and that are based on the psycholinguistics of Arabic. There is a clear need to establish normative data across the ages studied in the current work in order to inform future Arabic language test design and to develop measures sensitive to language development across a wider age range. The low performance on several language measures by females compared to their male peers requires further investigation, not only for understanding child development in Saudi Arabia but also for consideration of such sex differences in test design. The lack of agreement among individuals exhibiting low performance across the language measures tasks suggests that considerable work is needed to gain a better understanding of the characteristics of developmental language impairment in Arabic speakers, and to develop measures sensitive to both development and impairment.

Linguistic and Cognitive Measures in Arabic-Speaking ELLs

As reviewed in chapter 1, ELLs' performance of standardized measures of vocabulary and morphosyntax development is influenced by their prior knowledge and experience. ELLs may score poorly on such language-related measures because of the low levels of exposure to each of their languages (Carlson & Meltzoff, 2008; Paradis, 2005). It was suggested that perceptual-cognitive skills that are shared across languages, such as those supporting the decoding of written words, may help to characterize the language learning ability of ELLs (Oller

et al., 2007). Similarly, it has been argued that short-term and working memory measures are inherently less biased than language-related measures. Such processing-based measures emphasize the storage and processing of new information and pose similar demands on individuals regardless of ELLs' previous knowledge and experience (Engel de Abreu et al., 2013). Nevertheless, the particular pattern of performance across knowledge- and processing-dependent measures is not well understood, but is likely influenced by a number of relevant experiential factors. The purpose of study 2 was to investigate ELLs' performance on linguistic tests (vocabulary, language, and reading) in Arabic L1 and English L2, and cognitive measures of short-term and working memory and non-verbal intelligence relative to age-level expectations for monolingual speakers. A second goal was to investigate the influence of several factors on ELLs' performance such as chronological age, age of first exposure, the richness of language environment outside of school, current use patterns, and mother's education.

Summary of findings. The finding of significantly lower standardized scores for the ELLs on Arabic and English knowledge-based language tasks is consistent with many previous studies (e.g., Paradis, 2005; Paradis, Rice, Crago, & Marquis, 2008). Moreover, the comparable or above age-level expectations scores by ELLs on the English word and nonword reading tasks, but not on the Arabic word reading task, adds to the growing evidence of a phonological awareness benefit from learning two different phonological systems, at least for the language of instruction (Kang, 2012; Marinova-Todd et al., 2010; Oller et al., 2007). The comparable or above age-level expectations scores by ELLs on all of the cognitive measures except the Arabic nonword repetition task is consistent with previous evidence suggesting that measures of cognitive abilities in ELLs are less sensitive to difference in language experience than knowledge-based measures (Cockcroft, 2016; Engel de Abreu et al., 2013). Moreover, the

finding regarding the reduced performance of ELLs on Arabic nonword repetition, but not on English nonword repetition, adds to the growing evidence indicating that ELLs' performance on nonword repetition is affected by their previous sublexical phonological knowledge and experience in the target language (Thorn & Gathercole, 1999; Kohnert et al., 2006; Windsor et al., 2010). The finding that ELLs scored significantly higher on two number-based verbal working memory measures and one visuospatial short-term memory task is consistent with previous evidence suggesting that bilingual children may be at some advantage on working memory abilities (Bialystok et al., 2004). Finally, a possible explanation for the overall results of the stepwise multiple-regression analyses is that, unlike knowledge-based measures, cognitive measures were relatively minimally impacted by language experiences.

Implications of findings. The results of the English word and nonword reading tasks suggest that such tasks may be less sensitive to difference in language experience than traditional knowledge-based measures such as standardized measures of language and vocabulary. Moreover, ELLs' standard scores varied among their Arabic L1 and English L2, as well as among different language abilities, relative to age-level expectations for monolingual speakers of each language. These results suggest that to ensure a better understanding of the current level of language knowledge in ELLs, careful choice among tasks is required. The overall results of study 2 suggested that cognitive measures were relatively independent of language experiences and may provide an accurate indication of ELLs' language learning potential.

Linguistic and Cognitive Measures in Arabic-speaking ELLs and Monolingual Arabic-Speaking Children With and Without DLD

As reviewed in chapter 1, administering knowledge-based assessment tools such as English standardized language tests and interpreting scores based on monolingual norms may

lead to over-diagnosis of language impairment among ELLs (Donovan & Cross, 2002; Gutiérrez-Clellen, 1996; Klingner & Artiles, 2003). The findings of study 2 suggested that unlike knowledge-based measures, cognitive measures were relatively minimally impacted by language experiences. The aim of study 3 was to investigate the utility of knowledge- and processing-dependent measures in distinguishing ELLs from children with underlying language impairment. Processing-dependent measures such as verbal short-term memory and working memory measures may pose similar challenges and be equally familiar (or unfamiliar) to all children, regardless of the language they speak (Engel de Abreu & Gathercole, 2012). Particularly, verbal short-term and working memory measures that involve recall of highly familiar lexical stimuli, such as number words and basic words, have been found to minimize the role of linguistic knowledge and experiences in ELLs (Engel de Abreu et al., 2013). In fact, other studies have reported better working memory abilities in bilingual children compared to their monolingual peers (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004). A second goal of study 3 was to examine whether assessing ELLs on their L1 or native language can provide meaningful information to inform diagnostic decisions. It has been suggested that conducting assessments in ELLs' L1 or native language rather than using English testing may provide an accurate indication of ELLs' language abilities (Chu & Flores, 2011; Wagner et al., 2005). Evidence from several studies, however, shows that as proficiency in ELLs' L2 grows, ELLs' skills in L1 often do not develop further or even reduce and diminish across time (Anderson, 2012; Paradis, 2010). Such a pattern poses challenges when concerns regarding language development and language learning arise.

Summary of findings. Study 3 compared the performance of Arabic-speaking ELLs on linguistic and cognitive measures to two monolingual peer groups: typically developing Arabic-

speaking children (A-TD) and Arabic-speaking children with DLD (A-DLD). The finding of significantly lower scores on L1 knowledge-based language tasks (Arabic vocabulary and language) for the ELL group compared to the A-TD group is consistent with previous evidence suggesting that ELLs are often at risk of losing and/or not fully acquiring their L1 (Paradis, 2010). Moreover, the lower than expected performance by ELLs on the Arabic Single Word Reading Task as compared to the monolingual groups either with or without DLD may reflect ELLs' lack of specialized knowledge that is essential for reading in Arabic rather than reflecting an actual reading impairment in this population. The performance of the ELL group was comparable to the A-TD group on all of the verbal memory tasks tapping short-term and working memory, except for the Arabic nonword repetition task. These results, on the whole, are consistent with previous evidence suggesting that processing-dependent measures in ELLs are less sensitive to differences in language experience than knowledge-based measures (Engel de Abreu et al., 2013). In addition, the results are consistent with previous evidence suggesting that verbal short-term and working memory measures involving the recall of highly familiar lexical stimuli, such as number words and basic words, may be equally familiar to all children and less affected by verbal long-term memory (Engel de Abreu et al., 2013). The finding of significantly lower scores on the Arabic nonword repetition task for the ELL group compared to the A-TD group adds to the growing body of evidence indicating that phonological structure and language experience impact ELLs' performance on nonword repetition tasks (Kohnert et al., 2006; Windsor et al., 2010). The ELL group in this study scored significantly higher than the A-TD group on only the Digit Recall measure of verbal short-term memory. Given that there are only a few studies comparing working memory abilities between monolingual and bilingual children, additional research examining measures of verbal short-term memory and working memory in

ELLs is warranted to gain a better understanding of the ELLs' advantage on working memory performance. Finally, the ELL group in this study did not differ from their monolingual peers (A-TD and A-DLD) on all visuospatial short-term and working memory subtests. This finding is in line with previous studies that have shown no differences between TD and DLD groups on visuospatial short-term and working memory measures (Archibald & Gathercole, 2006c).

Implications of findings. The significant overlap between the ELL group and the monolingual Arabic-speaking children with DLD on L1 knowledge-based measures indicates that using such measures may increase the risk of the over-diagnosis of language impairment among ELLs. The overall finding of study 3 suggests that tasks that focus more on the cognitive processes that underlie language learning, rather than children's opportunities or experiences with the test language, may provide a more accurate representation of ELLs' linguistic abilities. Moreover, the verbal short-term and working memory results in this study highlighted that not all processing-dependent measures are equally effective in reducing the role of prior knowledge or experience in ELLs. For example, ELLs' performance on nonword repetition in this study is affected by their previous sublexical phonological knowledge and experience in the target language. Importantly, the present study indicates that verbal short-term and working memory tasks involving familiar lexical stimuli may be valid assessment tools that minimize the role of linguistic knowledge and experiences and help distinguish between ELLs and children with underlying language impairment.

Practical Implications

The ELLs in the present studies exhibited low performance on Arabic and English knowledge-based language tasks such as the vocabulary and language measures. This finding indicates that administering knowledge-based assessment tools, such as the measures employed

in this study, may increase the risk of over-diagnosis of language impairment among ELLs. Word and nonword reading measures that tap phonological awareness skills, on the other hand, may provide a more accurate representation of ELLs' linguistic abilities if the language of testing is the same as the child's language of literacy instruction. Cognitive measures such as verbal short-term and working memory involving familiar lexical stimuli may minimize the role of linguistic knowledge and experiences in ELLs, and such tasks may be valid assessment tools that help distinguish between ELLs and children with underlying language impairment. Even though nonword repetition tasks, tasks considered to tap verbal short-term memory, have been found to be a less biased form of assessment than knowledge-based measures, such tasks do not entirely eliminate the effect of children's experience with the target language.

Conclusions

There are three studies presented in this thesis. The first study assessed sensitivity to developmental change and sex differences of several available language measures in school age Arabic-speaking children. The second study investigated ELLs' performance on linguistic measures (Arabic L1 and English L2) and cognitive measures relative to age-level expectations for monolingual speakers, and considers the influence on performance of a number of relevant experiential factors. The third study examined cognitive and linguistic markers that may differentiate Arabic-speaking ELLs from age-matched monolingual children with and without DLD. A key finding of the first study was that available Arabic language tests are not sensitive to age-related differences across the 6-9 year age range. Tests tapping more complex language skills for older children need to be developed. The findings of studies 2 and 3 suggested that using knowledge-based measures may underestimate the language learning ability of ELLs, and may increase the risk of over-identification of language impairment among ELLs. Moreover,

careful choice among tasks is required to ensure a better understanding of the current level of language knowledge in ELLs. The overall results of studies 2 and 3 also suggest that unlike knowledge-based measures, cognitive measures may be valid assessment tools that minimize the role of linguistic knowledge and experiences and help distinguish between ELLs and children with underlying language impairment.

References

- Abdalla, F. (2002). *Specific language impairment in Arabic-speaking children: deficits in morphosyntax*. Ph.D. dissertation, McGill University.
- Abdalla, F., & Crago, M. (2008). Verb morphology defects in Arabic-speaking children with specific language impairment. *Applied Psycholinguistics*, 29(2), 315–340.
- Abu-Rabia, S., & Siegel, L. S. (2002). Reading, Syntactic, Orthographic, and Working Memory Skills of Bilingual Arabic-English Speaking Canadian Children. *Journal of Psycholinguistic Research*, 31(6), 661–678.
- Al-Akeel, A. (1998). *The acquisition of Arabic language comprehension by Saudi children*. Unpublished Ph.D. dissertation, University of Newcastle upon Tyne, UK.
- Al-Essa, A. (2006). When Najd meets Hijaz: dialect contact in Jeddah. *A Paper Presented at the 7th International Conference of AIDA, Vienna (Austria)*.
- Al-Tamimi, F. Y. (2011). *Assessment of language development in Arabic*. Encyclopedia of Language and Literacy Development (PP. 1-9). London, ON: Western University.
- Alahmadi, S. D. (2015). Loanwords in the Urban Meccan Hijazi Dialect: An Analysis of Lexical Variation according to Speakers' Sex, Age and Education. *International Journal of English Linguistics*, 5(6), 34–58.
- Aljenaie, K. (2001). *The emergence of tense and agreement in Kuwaiti Arabic children*. Unpublished Ph.D. dissertation, University of Reading.
- Alloway, T. P. (2007). *The Automated Working Memory Assessment*. London: Pearson Assessment.
- Alloway, T. P., Rajendran, G., & Archibald, L. M. D. (2009). Working Memory in Children With Developmental Disorders. *Journal of Learning Disabilities Volume*, 42(4), 372–383.

doi: 10.1177/0022219409335214

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Anderson, R. T. (2012). First language loss in Spanish-speaking children: Patterns of loss and implications for clinical practice. In B. A. Goldstein (Ed.), *In Bilingual Language Development and Disorders in Spanish-English Speakers* (p. 193–212.). Baltimore: MD 21285.
- Archibald, L. M. D., & Gathercole, S. E. (2006a). Nonword Repetition: A Comparison of Tests. *Journal of Speech Language and Hearing Research, 49*(5), 970–983. doi: 10.1044/1092-4388(2006/070)
- Archibald, L. M. D., & Gathercole, S. E. (2006b). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*(6), 675–693. doi: 10.1080/13682820500442602
- Archibald, L. M. D., & Gathercole, S. E. (2006c). Visuospatial immediate memory in specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(2), 265–278.
- Archibald, L. M. D., & Joanisse, M. F. (2009). On the Sensitivity and Specificity of Nonword Repetition and Sentence Recall to Language and Memory Impairments in Children. *Journal of Speech Language and Hearing Research, 52*(4), 899. doi: 10.1044/1092-4388(2009/08-0099)
- August, D., Calderón, M., & Carlo, M. (2002). *Transfer of skills from Spanish to English: A study of young learners*. Washington, DC: Center for Applied Linguistics. (Rep. No. ED-98-CO- 0071). Washington, DC: Center for Applied Linguistics.

- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A. D., & Hitch, G. (1974). *Working memory*. In B. G. (Ed.), *The psychology of learning and motivation*.
- Bajo, M., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador, & Y. Gambier (Eds.), In *Translation in context* (pp. 127_142). Amsterdam: Benjamins.
- Bakalla, M. (1979). *The morphological and phonological components of the Arabic verb (Meccan Arabic)*. London: Longman.
- Balilah, A., & Archibald, L. M. D. (under review). Sentence Recall and Single Word Reading in Monolingual Children and Same-age English Language Learners (ELLs) with and without Parental Concerns about Language Development. *Manuscript Submitted for Publication*.
- Balilah, A., & Archibald, L. M. D. (in press). The Measurement of Language Ability and Impairment in Arabic-speaking Children. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice*. Springer.
- Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience*, 10, 136–148.
- Bedore, L. M., & Leonard, L. B. (1998). Specific Language Impairment and Grammatical Morphology: A Discriminant Function Analysis. *Journal of Speech Language and Hearing Research*, 41(5), 1185. doi: 10.1044/jslhr.4105.1185
- Bedore, L. M., & Peña, E. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, 11(1), 1–29.

- Bedore, L. M., Peña, E., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T., Gillam, R. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition, 15*(3), 616–629. doi: 10.1017/S1366728912000090
- Beeston, A. F. L. (1970). *The Arabic language today*. London: Hutchinson University Library.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636–644. doi: 10.1111/1467-8624.00046/full
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. New York, NY: Cambridge University Press.
- Bialystok, E., Craik, F., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging, 19*, 290–303.
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition, 34*(4), 859–873.
- Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language, 109*(2), 93–100.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531. doi: 10.1017/S1366728909990423
- Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24*(1), 27–44.
- Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The

- rise and fall of inhibitory control. *International Journal of Bilingualism*, 9(1), 103–119. doi: 10.1177/13670069050090010701
- Birdsong, D. (2005). *Determining age effects in second language acquisition*. Handbook of bilingualism: psycholinguistic approaches. (2nd ed.). 109-27. New York: Oxford University Press.
- Bishop, D. V. M. (1992). The Underlying Nature of Specific Language Impairment. *Journal of Child Psychology and Psychiatry*, 33(1), 3–66. doi: 10.1111/j.1469-7610.1992.tb00858.x
- Bishop, D. V. M. (1994). Is Specific Language Impairment a Valid Diagnostic Category? Genetic and Psycholinguistic Evidence. *Philosophical Transactions: Biological Sciences*. Royal Society. doi: 10.2307/56025
- Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. East Sussex, UK: Psychology Press.
- Bishop, D. V. M. (2000). Pragmatic Language Impairment: A Correlate of SLI a Distinct Subgroup, or Part of the Autistic Continuum? In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome*. Psychology Press, Hove, East Sussex.
- Bishop, D. V. M. (2003). Genetic and environmental risks for specific language impairment in children. *International Journal of Pediatric Otorhinolaryngology*, 67S1:S143-S157. doi: 10.1098/rstb.2000.0770
- Bishop, D. V. M. (2014). *Uncommon understanding: development and disorders of language comprehension in children* (Classic ed). Psychology Press.
- Bishop, D. V. M., & Clarkson, B. (2003). Written Language as a Window in to Residual Language Deficits: A Study of Children With Persistent and Residual Speech and Language

- Impairments. *Cortex*, 39(2), 215–237. doi: 10.1016/S0010-9452(08)70106-0
- Bishop, D. V. M., & Edmundson, A. (1987). Language impaired 4-year olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Research*, 52, 156–173.
- Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders*, 44(5), 600–615. doi: 10.1080/13682820802259662
- Bishop, D. V. M., North, T., & Donlan, C. (1996). Pergamon Nonword Repetition as a Behavioural Marker for Inherited Language Impairment: Evidence From a Twin Study. *Journal of Child Psychology and Psychiatry*, 37(4), 91–403.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE : A Multinational and Multidisciplinary Delphi Consensus Study . Identifying Language Impairments in. *PloS One*, 11(7), e0158753. doi: 10.1371/journal.pone.0158753
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*. doi: 10.1111/jcpp.12721
- Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 13(3), 325–344. doi: 10.1080/13670050903342019
- Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. (2010). Arabic Morphology in the Neural Language System. *Journal of Cognitive Neuroscience*, 22(5),

998–1010. doi: 10.1162/jocn.2009.21273

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics*, 32(1), 221–241. doi: 10.1017/S0142716410000354

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, 301, 419–421.

Briscoe, J., Bishop, D. V. M., & Frazier Norbury, C. (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*, 42(3), 329–340. doi: 10.1017/S0021963001007041

Brooks-Gunn, J., Han, W., & Waldfogel, J. (2002). Maternal Employment and Child Cognitive Outcomes in the First Three Years of Life: The NICHD Study of Early Child Care. *Child Development*. WileySociety for Research in Child Development. doi: 10.2307/3696270

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence-3 (TONI-3): A language-free measure of intelligence, aptitude and reasoning*.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, 22(2), 307–324.

Bukshaisha, F. (1985). *An experimental phonetic study of some aspect of Qatari Arabic*.

Unpublished doctoral dissertation, University of Edinburgh, Edinburgh.

Bulos, A. (1965). *The Arabic trilateral verb: a comparative study of grammatical concepts and progress*. Beirut: KHAYATS.

Caesar, L. G., & Kohler, P. D. (2007). The State of School-Based Bilingual Assessment: Actual Practice Versus Recommended Guidelines. *Language Speech and Hearing Services in Schools*, 38(3), 190. doi: 10.1044/0161-1461(2007/020)

- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31–42.
- Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology, 13*(1), 61–68.
- Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing Bias in Language Assessment. *Journal of Speech Language and Hearing Research, 40*(3), 519. doi: 10.1044/jslhr.4003.519
- Carlson, S., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282–298.
- Case, R., Kurland, D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*(3), 386–404.
- Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A Longitudinal Investigation of Reading Outcomes in Children With Language Impairments. *Journal of Speech Language and Hearing Research, 45*(6), 1142. doi: 10.1044/1092-4388(2002/093)
- Catts, H. W., & Jensen, P. J. (1983). Speech Timing of Phonologically Disordered Children: Voicing Contrast of Initial and Final Stop Consonants. *Journal of Speech Language and Hearing Research, 26*(4), 501–510. doi: 10.1044/jshr.2604.501
- Chitiri, H., Sun, Y., Willows, D., & Taylor, I. (1992). Word recognition in second-language reading. *Advances in Psychology, 83*, 283–297.
- Chu, S., & Flores, S. (2011). Assessment of English Language Learners with Learning Disabilities. *Clearing House, 84*(6), 244–248. doi: 10.1080/00098655.2011.590550
- Cobo-Lewis, A. B., Eilers, E. R., Pearson, B. Z., & Umbel, V. C. (2002). Interdependence of

- Spanish and English knowledge in language and literacy among bilingual children. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children*. Clevedon, England: Multilingual Matters.
- Cockcroft, K. (2016). A comparison between verbal working memory and vocabulary in bilingual and monolingual South African school beginners: implications for bilingual language assessment. *International Journal of Bilingual Education and Bilingualism*, 19(1), 74–88. doi: 10.1080/13670050.2014.964172
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100(4), 589–608. doi: 10.1037//0033-295X.100.4.589
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic Markers for Specific Language Impairment (SLI). *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(6), 741–748.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. doi: 10.3758/BF03196772
- Costa, A., Roelstraete, B., & Hartsuiker, R. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, 13(6), 612–617. doi: 10.3758/BF03213911
- Craig, H. K., & Washington, J. A. (1993). Access Behaviors of Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research*, 36(2), 322. doi: 10.1044/jshr.3602.322
- Cummins, J. (1996). *Negotiating identities: Education for empowerment in a diverse society*.

Ontario: California Association for Bilingual Education.

Cummins, J. (2000). Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. *Retrieved April.*

Daher, J. (1998). *Linguistic variation in Damascus Arabic: A quantitative analysis of men's and women's speech*. Unpublished doctoral dissertation, New York University, New York.

Davison, M. D., & Hammer, C. S. (2012). Development of 14 English grammatical morphemes in Spanish–English preschoolers. *Clinical Linguistics and Phonetics*, 26(8), 728–742. doi: 10.3109/02699206.2012.700679

Dollaghan, C., & Campbell, T. (1998a). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing*, 41(5), 1136–1146.

Dollaghan, C., & Campbell, T. (1998b). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146.

Donovan, S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Dulay, H. C., & Burt, M. K. (1973). Should we Teach Children Syntax? *Language Learning*, 23(2), 245–258. doi: 10.1111/j.1467-1770.1973.tb00659.x

Dulay, H. C., & Burt, M. K. (1974). Natural Sequences in Child Second Language Acquisition. *Language Learning*, 24(1), 37–53. doi: 10.1111/j.1467-1770.1974.tb00234.x

Dunn, L. M., & Dunn, D. M. (2006). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson, Inc.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale* (2nd ed.). Windsor: NFER-Nelson Publishing Company Ltd.

Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D., & Calderón,

- M. (2007). The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics*, 28(1), 171–190.
- El-Halees, Y., & Wiig, E. H. (1999a). *Arabic Language Screening Tests: Preschool and School-Age*. Arlington, Schema Press.
- El-Halees, Y., & Wiig, E. H. (1999b). *Arabic Receptive-Expressive Vocabulary Test: Preschool and School-Age*. Arlington, Schema Press.
- Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis Weismer, S. (1996). Capacity limitations in working memory: The impact on lexical and morphological learning by children with language impairment. *Topics in Language Disorders*, 17, 33–44.
- Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-aged children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43(4), 865–878.
- Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory (Hove, England)*, 19(5), 529–537. doi: 10.1080/09658211.2011.590504
- Engel de Abreu, P. M. J., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, 56(2), 630–642.
- Engel de Abreu, P. M. J., & Gathercole, S. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, 104(4), 974–986.
- ErDOS, C., Genesee, F., & Savage, R. (2011). Individual differences in second language reading

- outcomes. *International Journal of Bilingualism*, 15(1), 3–25. doi: 10.1177/1367006910371022
- Farwell, C. (1972). A Note on the Production of Fricatives in Linguistically Deviant Children. *Papers and Reports on Child Language Development*, 4, 93–101.
- Fee, E. J. (1995). The phonological system of a specifically language-impaired population. *Clinical Linguistics and Phonetics*, 9(3), 189–209. doi: 10.3109/02699209508985332
- Ferguson, C. (1959). Diglossia. *Word*, 15, 325–340.
- Fox, B., & Routh, D. K. (1980). Phonemic analysis and severe reading disability in children. *Journal of Psycholinguistic Research*, 9(2), 115–119. doi: 10.1007/BF01067466
- Fujiki, M., Brinton, B., & Sonnenberg, E. A. (1990). Repair of overlapping speech in the conversations of specifically language-impaired and normally developing children. *Applied Psycholinguistics*, 11(2), 201. doi: 10.1017/S0142716400008778
- Gathercole, S. E., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, 33(5), 672–688.
- Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28(5), 887–898.
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English Language Learners in U.S. Schools: An Overview of Research Findings. *Journal of Education for Students Placed at Risk (JESPAR)*, 10(4), 363–385. doi: 10.1207/s15327671espr1004_2
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. New York: Cambridge University Press.

- Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development & disorders: A handbook on bilingualism & second language learning, Vol. 11*. Paul H Brookes Publishing.
- German, D. J., & Simon, E. (1991). Analysis of Children's Word-Finding Skills in Discourse. *Journal of Speech Language and Hearing Research, 34*(2), 309–316. doi: 10.1044/jshr.3402.309
- Gibson, T., Peña, E., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism, 17*(1), 90–110. doi: 10.1080/13670050.2012.743960
- Gilhool, A., Burrows, L., Goldstein, B., & Paradis, J. (2009). English phonological skills of English language learners. In *In Poster presented at the Annual American Speech Language Hearing Association Convention, New Orleans, LA*.
- Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of Specific Language Impairment in Bilingual Children: Assessment in English. *Journal of Speech Language and Hearing Research, 56*(6), 1813. doi: 10.1044/1092-4388(2013/12-0056)
- Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*(1), 41–65.
- Goldstein, B. (2004). *Bilingual Language Development and Disorders in Spanish-English Speakers*. Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.
- Goldstein, B. (2012). *Bilingual Language Development and Disorders in Spanish-English Speakers*. Paul H. Brookes Pub. Co.

- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.
- Gottardo, A., Yan, B., Siegel, L. S., & Wade-Woolley, L. (2001). Factors Related to English Reading Performance in Children With Chinese as a First Language: More Evidence of Cross-Language Transfer of Phonological Processing. *Journal of Educational Psychology*, *93*(3), 530–542. doi: 10.1037//0022-0663.93.3.530
- 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. doi: 10.1044/1092-4388(2007/015)
- Gutiérrez-Clellen, V. F. (1996). Language diversity: Implications for assessment. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of communication and language*, Baltimore, MD: Paul Brookes.
- Hagiliassis, N., Pratt, C., & Johnston, M. (2006). Orthographic and Phonological Processes in Reading. *Reading and Writing*, *19*(3), 235–263. doi: 10.1007/s11145-005-4123-9
- Hakuta, K., Butler, Y., & Witt, D. (2000). How Long Does It Take English Learners to Attain Proficiency? *The University of California Linguistic Minority Research Institute*.
- Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish–English bilingual children’s language abilities. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1251–1264.
- Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008). Exposure to English Before and After Entry into Head Start: Bilingual Children’s Receptive Language Growth in Spanish and

- English. *International Journal of Bilingual Education and Bilingualism*, 11(1), 30–56. doi: 10.2167/beb376.0
- Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, 53(1), 37–45. doi: 10.1111/j.1469-7610.2011.02430.x
- Hodson, B., & Paden, E. (1981). Phonological Processes Which Characterize Unintelligible and Intelligible Speech in Early Childhood. *Journal of Speech and Hearing Disorders*, 46(4), 369–373. doi: 10.1044/jshd.4604.369
- Hoff, E. (2005). *Language development*. Belmont, CA: Wadsworth/Thomson Learning.
- Ibrahim, M. H. (1989). Communicating in Arabic: Problems and prospects. In F. Coulmas (Ed.), *Language adaptation* (pp. 39–58). Cambridge, United Kingdom: Cambridge University Press.
- Jackson, C. W., Schatschneider, C., & Leacox, L. (2014). Longitudinal Analysis of Receptive Vocabulary Growth in Young Spanish English–Speaking Children From Migrant Families. *Language Speech and Hearing Services in Schools*, 45(1), 40–51. doi: 10.1044/2013_LSHSS-12-0104
- Jackson, N., Holm, A., & Dodd, B. (1998). Phonological awareness and spelling abilities of Cantonese–English bilingual children. *Asia Pacific Journal of Speech, Language and Hearing*, 3(2), 79–96.
- Jacobson, P. F., & Cairns, H. S. (2008). Considering Linguistic Input in a Bilingual Situation: Implications for Acquisition. *Language Speech and Hearing Services in Schools*, 39(3), 352–364. doi: 10.1044/0161-1461(2008/033)
- Jared, D., & Kroll, J. (2001). Do bilinguals activate phonological representations in one or both

- of their languages when naming words? *Journal of Memory and Language*, 44(1), 2–31.
- Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research*, 46(6), 1297–1311.
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, 24(1), 131–161.
- Jia, G., Aaronson, D., & Wu, Y. H. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, 23(4), 599–621.
- Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research*, 50(5), 1280–1299.
- Johnston, J. R., & Smith, L. B. (1989). Dimensional Thinking in Language Impaired Children. *Journal of Speech Language and Hearing Research*, 32(1), 33. doi: 10.1044/jshr.3201.33
- Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research*, 37(2), 418. doi: 10.1044/jshr.3702.418
- Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical–semantic skills in L1 and L2. *Journal of Speech Language and Hearing Research*, 48(2), 372–383. doi: 10.1044/1092-4388(2005/026)
- Kang, J. (2012). Do bilingual children possess better phonological awareness? Investigation of Korean monolingual and Korean-English bilingual children. *Reading and Writing*, 25(2), 411–431.
- Kaushanskaya, M., Gross, M., & Buac, M. (2010). *Unpublished Parent Interview Questionnaire:*

Developed in the Language Acquisition and Bilingualism Lab, Waisman Center, University of Wisconsin-Madison.

Khalaf, A., Ekblom, O., Kowalsk, J., Berggren, V., Westergren, A., & Al-Hazzaa, H. (2013).

Female university students' physical activity levels and associated factors-a cross-sectional study in southwestern Saudi Arabia. *International Journal of Environmental Research and Public Health*, *10*, 3502–3517.

Klingner, J., & Artiles, A. (2003). When should bilingual students be in special education?

Educational Leadership, *61*(2), 66–71.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and

implications for clinical actions. *Journal of Communication Disorders*, *43*(6), 456–473. doi: 10.1016/j.jcomdis.2010.02.002

Kohnert, K. (2012). Processing skills in early sequential bilinguals. In B. Goldstein (Ed.),

Bilingual language development & disorders in Spanish-English speakers. Baltimore, MD: Brookes.

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate

children with language impairment from typical bilinguals? *Learning Disabilities Research and Practice*, *21*(1), 19–29. doi: 10.1111/j.1540-5826.2006.00204.x

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. M., & Plomin, R.

(2005). Genetic Influences in Different Aspects of Language Development: The Etiology of Language Skills in 4.5-Year-Old Twins. *Child Development*, *76*(3), 632–651. doi: 10.1111/j.1467-8624.2005.00868.x

Lafrance, A., & Gottardo, A. (2005). A longitudinal study of phonological processing skills and

reading in bilingual children. *Applied Psycholinguistics*, *26*(4), 559–578.

- Leonard, L. B. (1982). Phonological deficits in children with developmental language impairment. *Brain and Language*, *16*(1), 73–86. doi: 10.1016/0093-934X(82)90073-6
- Leonard, L. B. (2014). *Children with specific language impairment*. MIT press.
- Leonard, L. B., Eyer, J., Bedore, L. M., & Grela, B. (1997). 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.
- Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical Morphology and the Lexicon in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research*, *42*(3), 678. doi: 10.1044/jslhr.4203.678
- Lezak, M. D., Howieson, D. B., & Loring., D. W. (2004). *Neuropsychological Assessment* (4th ed.). Oxford: Oxford University Press.
- Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspect of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge, UK: Cambridge University Press.
- Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, *95*(3), 482.
- Loeb, D. F., & Leonard, L. B. (1988). Specific language impairment and parameter theory. *Clinical Linguistics & Phonetics*, *2*(4), 317–327. doi: 10.3109/02699208808985263
- Lorge, I., & Chall, J. (1963). Estimating the Size of Vocabularies of Children and Adults: An analysis of methodological issues. *The Journal of Experimental Education*, *32*(2), 147–157. doi: 10.1080/00220973.1963.11010819
- Lutf, M., You, X., Cheung, Y.-M., & Chen, C. L. P. (2013). Arabic font recognition based on

- diacritics features. *Pattern Recognition*, 47(2), 672–684. doi: 10.1016/j.patcog.2013.07.015
- Majerus, S., Poncelet, M., Greffe, C., & van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, 93(2), 95–119.
- Marchman, V. A., & Martinez-Sussmann, C. (2002). Concurrent validity of caregiver/ parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, 45(5), 983–998.
- Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics & Phonetics*, 24(4–5), 387–400. doi: 10.3109/02699200903532508
- Masoura, E., & Gathercole, S. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 13(3–4), 422–429.
- Mendenhall, W., & Sincich, T. (1996). *A second course in statistics: regression analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Miller, C., Kail, R., Leonard, L. B., & Tomblin, J. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44(2), 416–433.
- Montgomery, J. (1995). Examination of phonological working memory in specifically language impaired children. *Applied Psycholinguistics*, 16, 355–378.
- Montgomery, J. (2000). Relation of working memory to off-line and real-time sentence processing in children with specific language impairment. *Applied Psycholinguistics*, 21,

117–148.

- Nadler, R. T., & Archibald, L. M. D. (2014). The Assessment of Verbal and Visuospatial Working Memory With School Age Canadian Children. *Canadian Journal of Speech-Language Pathology and Audiology*, 38(3).
- Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–35). Hillsdale, NJ: Erlbaum.
- Newman, D. (2002). The phonetic status of Arabic within the world's language: The uniqueness of the lughat al-daad. *Antwerp Paper in Linguistic*, 100, 65–75.
- Nippold, M. A. (1998). *Later language development: The school-age and adolescent years* (2nd ed.). Austin, TX: PRO-ED.
- Oller, D. K., & Eilers, E. R. (2002). *Language and literacy in bilingual children*. (Vol. 2). Multilingual Matters.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191–230.
- Omar, M. (1973). *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.
- Oweini, A., & Hazoury, K. (2010). Towards a sight word list in Arabic. *International Review of Education*, 56(4), 457–478. doi: 10.1007/s11159-010-9170-z
- Paradis, J. (2005). Grammatical Morphology in Children Learning English as a Second Language. *Language Speech and Hearing Services in Schools*, 36(3), 172. doi: 10.1044/0161-1461(2005/019)
- Paradis, J. (2007). Second language acquisition in childhood. In *Blackwell handbook of language development* (pp. 387–405). doi: 10.1002/9780470757833.ch19/summary

- Paradis, J. (2008). Tense as a clinical marker in English L2 acquisition with language delay/impairment. In E. Gavrusseva & B. Haznedar (Eds.), *Current Trends in Child Second Language Acquisition: A Generative Perspective*, (pp. 337–356). Amsterdam: John Benjamins.
- Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics*, *31*(2), 227–252.
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. doi: 10.1075/lab.1.3.01par
- Paradis, J., Emmerzael, K., & Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*(6), 474–497.
- Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders : a handbook on bilingualism and second language learning*. Paul H. Brookes Pub. Co.
- Paradis, J., Rice, M. L., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, *29*(4), 689–722. doi: 10.1017/S0142716408080296
- Paradis, J., Schneider, P., & Duncan, T. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*(3), 971–981.
- Paul, R., & Shriberg, L. D. (1982). Associations between Phonology and Syntax in Speech-Delayed Children. *Journal of Speech Language and Hearing Research*, *25*(4), 536–547. doi: 10.1044/jshr.2504.536

- Pearson, B., Fernandez, S., Lewedeg, V., & Oller, D. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58.
- Pearson, B., Fernández, S., & Oller, D. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x/full
- Peña, E., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: two languages classify better than one. *International Journal of Language & Communication Disorders*, *51*(2), 192–202. doi: 10.1111/1460-6984.12199
- Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing Test Bias Through Dynamic Assessment of Children's Word Learning Ability. *American Journal of Speech-Language Pathology*, *10*(2), 138. doi: 10.1044/1058-0360(2001/014)
- Pence, K. L., & Justice, L. M. (2008). *Language development from theory to practice*. Upper Saddle River, NJ: Pearson Education.
- Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press.
- Prochazka, S. (2006). Arabic. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2ed ed.). Amsterdam, the Netherlands: Elsevier Ltd.
- 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. doi: 10.1207/s15327817la0604_1
- Rice, M. L., Sell, M. A., & Hadley, P. A. (1991). Social Interactions of Speech, and Language-Impaired Children. *Journal of Speech Language and Hearing Research*, *34*(6), 1299. doi: 10.1044/jshr.3406.1299

- Rice, M. L., & Wexler, K. (1996a). Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *Journal of Speech Language and Hearing Research, 39*(6), 1239. doi: 10.1044/jshr.3906.1239
- Rice, M. L., & Wexler, K. (1996b). Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *Journal of Speech Language and Hearing Research, 39*(6), 1239. doi: 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. doi: 10.1044/jshr.3804.850
- Ruscello, D. M., St. Louis, K. O., & Mason, N. (1991). School-Aged Children With Phonologic Disorders: Co-existence with other Speech/Language Disorders. *Journal of Speech Language and Hearing Research, 34*(2), 236–242. doi: 10.1044/jshr.3402.236
- Sandberg, K. L., & Reschly, A. L. (2011). English Learners: Challenges in Assessment and the Promise of Curriculum-Based Measurement. *Remedial and Special Education, 32*(2), 144–154. doi: 10.1177/0741932510361260
- Schelleter, C. (1990). Pronoun use in normal and language-impaired children. *Clinical Linguistics & Phonetics, 4*(1), 63–75. doi: 10.3109/02699209008985471
- Schwartz, R. G., Leonard, L. B., Folger, M. K., & Wilcox, M. J. (1980). Early Phonological Behavior in Normal-Speaking and Language Disordered Children: Evidence for a Synergistic View of Linguistic Disorders. *Journal of Speech and Hearing Disorders, 45*(3), 357–377. doi: 10.1044/jshd.4503.357
- Semel, E. M., Wiig, E. H., & Secord, W. A. (1989a). *CELF-R Screening Test*. San Antonio, Psychological Corporation.

- Semel, E. M., Wiig, E. H., & Secord, W. A. (1989b). *Clinical Evaluation of Language Fundamentals (revised)*. San Antonio, Psychological Corporation.
- Semel, E. M., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals* (3rd ed.). San Antonio, Psychological Corporation.
- Semel, E. M., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, (CELF-4)*. Inc. San Antonio, USA. The Psychological Corporation. A Harcourt Assessment.
- Shaalan, S. (2010). *Investigating Grammatical Complexity in Gulf Arabic Speaking Children with Specific Language Impairment (SLI)*. Doctoral dissertation, UCL (University College London).
- Shaalan, S. (2014). Reliability and Validity of Four Arabic Language Tests : A comparison of performance of Qatari School-aged children with and without language impairment. *Arab Journal of Applied Linguistics*, 2(1), 2490–4198.
- Shahin, K. (2010). *Fostering Language Development in Arabic*. *Encyclopedia of Language and Literacy Development*. London, ON: Canadian Language and Literacy Research Network.
- Sheng, L., & McGregor, K. K. (2010). Object and Action Naming in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research*, 53(6), 1704. doi: 10.1044/1092-4388(2010/09-0180)
- Sienny, M. E. (1978). *The syntax of urban Hijazi Arabic*. Beirut, Lebanon: Librairie du Liban.
- Smyth, M. M., & Scholey, K. A. (1996). The relationship between articulation time and memory performance in verbal and visuospatial tasks. *British Journal of Psychology*, 87(2), 179–191. doi: 10.1111/j.2044-8295.1996.tb02584.x
- Snowling, M., Bishop, D. V. M., & Stothard, S. E. (2000). Is Preschool Language Impairment a

- Risk Factor for Dyslexia in Adolescence? *Journal of Child Psychology and Psychiatry*, 41(5), 587–600. doi: 10.1111/1469-7610.00651
- Stark, R., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, 46(2), 114–122.
- Steckol, K. F., & Leonard, L. B. (1979). The use of grammatical morphemes by normal and language-impaired children. *Journal of Communication Disorders*, 12(4), 291–301. doi: 10.1016/0021-9924(79)90027-3
- Stothard, S. E., Snowling, M., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, 41(2), 407–418.
- Stow, C., & Dodd, B. (2003). Providing an equitable service to bilingual children in the UK: a review. *International Journal of Language & Communication Disorders*, 38(4), 351–377. doi: 10.1080/1368282031000156888
- Tabors, P. (2008). *One Child, Two Languages* (2nd ed.). Baltimore, MD: Paul Brookes Publishing.
- Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Golinkoff & R. M. Hirsh-Pasek (Eds.), *Action meets word: How children learn verbs* (pp. 477–498). New York: Oxford University Press.
- Thompson-Panos, K., & Thomas-Ruzic, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quartely*, 17(4), 609–623.
- Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1–21. doi: 10.1080/14769670500215647

- Thorn, A., & Gathercole, S. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology: Section A*, 52(2), 303–324.
- Tomblin, J. B., Records, N., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1284–1294.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The Stability of Primary Language Disorder. *Journal of Speech Language and Hearing Research*, 46(6), 1283–1296. doi: 10.1044/1092-4388(2003/100)
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of word reading efficiency*. Psychological Corporation.
- Trauner, D., Wulfeck, B., Tallal, P., & Hesselink, J. (1995). *Neurological and MRI profiles of children with developmental language impairment. Technical Report CND-9513, Center for Research in Language, University of California at San Diego*. doi: 10.1017/S0012162200000876
- Umbel, V. M., Pearson, B. Z., Fernandez, M. C., & Oller, D. K. (1992). Measuring Bilingual Children's Receptive Vocabularies. *Child Development*, 63(4), 1012–1020. doi: 10.1111/j.1467-8624.1992.tb01678.x
- Wade-Woolley, L., & Geva, E. (2000). Processing novel phonemic contrasts in the acquisition of L2 word reading. *Scientific Studies of Reading*, 4(4), 295–311.
- Wagner, R. K., Francis, D. J., & Morris, R. D. (2005). Identifying English Language Learners with Learning Disabilities: Key Challenges and Possible Approaches. *Learning Disabilities Research & Practice*, 20(1), 6–15.

- Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., Jones, M., & M., K.-L. (2000). Nonword Repetition Performance in School-Age Children With and Without Language Impairment. *Journal of Speech Language and Hearing Research*, 43(4), 865. doi: 10.1044/jslhr.4304.865
- Wiig, E. H., & El-Halees, Y. (2000). Developing a language screening test for Arabic-speaking children. *Folia Phoniatrica et Logopaedica*, 52(6), 260–274.
- Wiig, E. H., & Secord, W. A. (1992). *Test of Word Knowledge*. San Antonio, Psychological Corporation.
- Wiig, E. H., Secord, W. A., & Semel, E. M. (1992). *Clinical Evaluation of Language Fundamentals-Preschool*. San Antonio, Psychological Corporation.
- Williams, K. T. (2006). *Expressive Vocabulary Test* (2nd ed.). Minneapolis, MN: Pearson, Inc.
- Windsor, J., & Kohnert, K. (2004). The Search for Common Ground: Part I. Lexical performance by linguistically diverse learners. *Journal of Speech Language and Hearing Research*, 47(4), 877. doi: 10.1044/1092-4388(2004/065)
- Windsor, J., Kohnert, K., Lobitz, K., & Pham, G. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, 19(4), 298–310.
- World Health Organization. (1993). *The ICD-10 classification for mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Switzerland: WHO.
- Wulfeck, B., & Bates, E. (1991). Differential Sensitivity to Errors of Agreement and Word Order in Broca's Aphasia. *Journal of Cognitive Neuroscience*, 3(3), 258–272. doi: 10.1162/jocn.1991.3.3.258
- Zayyan, A. A., Elmahdy, M., binti Husni, H., & Al Ja'am, J. M. (2016). Automatic diacritics

restoration for modern standard arabic text. In *2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 221–225). IEEE. doi:

10.1109/ISCAIE.2016.7575067

Zdorenko, T., & Paradis, J. (2012). Articles in child L2 English: When L1 and L2 acquisition meet at the interface. *First Language*, *32*(1–2), 38–62.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale (PLS-3)*. San Antonio, TX: Psychological Corp.

Appendix A

Descriptive statistic for all Arabic language tasks scores as a function of age band: 6-6;11 (n=106), 7-7;11 (n=104), 8-8;11 (n=106), and 9-9;11 (n=105) for the monolingual Arabic-speaking children

<i>Age</i>	<i>M</i>	<i>SD</i>	<i>Range.</i>	<i>Skew.</i>	<i>Kurt.</i>
Arabic Receptive- Expressive Vocabulary Test ^a :					
Arabic Receptive Vocabulary Test					
6-6;11	34.25 (-1.27)	2.66 (.40)	28-40 (-2.22 / -.47)	-.27	-.50
7-7;11	40.33 (-.36)	2.34 (.35)	34-45 (-1.31 / .35)	-.33	-.51
8-8;11	45.35 (.40)	2.48 (.37)	38-50 (-.71 / 1.10)	-.74	.26
9-9;11	50.77 (1.23)	2.72 (.35)	35-55 (.19 / 1.86)	-2.23	10.09
Arabic Expressive Vocabulary Test					
6-6;11	29.90 (-1.10)	3.60 (.52)	23-38 (-2.10 / .07)	.37	-.28
7-7;11	34.70 (-.41)	3.84 (.55)	22-43 (-2.24 / .79)	-.24	.12
8-8;11	40.23 (.39)	3.66 (.52)	27-49 (-1.38 / 1.66)	-.05	.77
9-9;11	45.25 (1.12)	4.11 (.59)	35-55 (-.36 / -2.53)	.102	-.09
Total-Arabic Receptive- Expressive Vocabulary Test					
6-6;11	64.15 (-1.23)	5.02 (.39)	53-77 (-2.09 / -.24)	.04	-.44
7-7;11	75.03 (-.39)	5.02 (.39)	58-88 (-1.71 / .60)	-.47	.36
8-8;11	85.58 (.41)	5.06 (.39)	68-99 (-.94 / 1.45)	-.35	.54
9-9;11	96.02 (1.22)	5.71 (.44)	78-109 (-.17 / 2.22)	-.47	1.26
Arabic Picture Vocabulary Test					
6-6;11	95.47	10.61	72-116	-.23	-.62
7-7;11	100.04	11.75	58-119	-1.11	1.93
8-8;11	103.85	9.61	72-119	-.89	.89
9-9;11	106.12	11.92	65-123	-1.01	1.05
Arabic Language Screening Test:					
Arabic Language Screening Test (verbal abilities)					
6-6;11	34.03	8.33	14-53	.08	-.57
7-7;11	38.05	7.45	19-52	-.31	-.54
8-8;11	40.23	6.89	22-54	-.51	-.13

9-9;11	41.62	6.64	22-56	-.32	.01
Arabic Language Screening Test (non-verbal abilities)					
6-6;11	20.92	6.41	4-34	.11	-.48
7-7;11	23.87	5.22	14-34	.24	-.92
8-8;11	25.55	4.26	16-36	-.14	-.16
9-9;11	25.37	5.11	9-35	-.71	1.26
Total-Arabic Language Screening Test					
6-6;11	54.94	13.54	22-81	.07	-.91
7-7;11	61.91	11.31	39-84	-.18	-.78
8-8;11	65.77	9.77	40-87	-.52	-.06
9-9;11	66.99	10.21	37-89	-.35	.37
Arabic Language Test:					
Sentence Comprehension (A)					
6-6;11	19.42	2.39	8-22	-2.69	10.33
7-7;11	19.83	2.51	9-22	-2.29	5.97
8-8;11	20.32	1.75	10-22	-2.41	10.58
9-9;11	20.32	2.42	7-22	-3.09	11.83
Sentence Comprehension (B)					
6-6;11	12.34	2.52	4-18	-.69	.64
7-7;11	13.19	2.79	5-18	-.63	.39
8-8;11	14.22	2.30	6-18	-.76	1.02
9-9;11	14.47	2.87	4-18	-1.26	2.28
Total- Sentence Comprehension					
6-6;11	31.76	4.12	14-39	-1.73	5.02
7-7;11	33.02	4.36	16-40	-1.61	3.81
8-8;11	34.54	3.46	16-40	-1.81	7.02
9-9;11	34.79	4.75	12-40	-2.15	6.53
Expressive Language (A)					
6-6;11	16.91	3.58	7-24	-.195	-.108
7-7;11	18.65	2.98	10-24	-.24	-.21
8-8;11	18.93	3.01	11-24	-.48	-.25
9-9;11	18.82	2.77	11-24	-.69	.49
Expressive Language (B)					
6-6;11	32.04	6.47	19-44	-.27	-.76
7-7;11	35.43	4.89	23-44	-.40	-.13
8-8;11	36.75	4.49	25-44	-.33	-.69
9-9;11	36.76	4.07	27-44	-.34	-.31
Total- Expressive Language					
6-6;11	48.94	9.41	26-68	1.13	-.61
7-7;11	54.09	7.15	33-68	-.35	.02
8-8;11	55.68	6.86	40-67	-.31	-.94
9-9;11	55.58	6.23	42-68	-.42	-.33
Sentence Repetition (A)					
6-6;11	50.30	4.53	34-54	-2.03	4.09
7-7;11	51.40	4.69	27-54	-2.56	7.65
8-8;11	51.84	3.96	30-54	-2.61	8.61

9-9;11	51.68	6.10	22-54	-3.30	10.49
Sentence Repetition (B)					
6-6;11	39.93	11.59	15-64	-.09	-.76
7-7;11	46.58	9.07	26-67	-.09	-.51
8-8;11	47.49	8.92	30-65	-.10	-.1
9-9;11	50.18	10.19	19-69	-.71	-.11
Total- Sentence Repetition					
6-6;11	90.24	12.30	61-118	-.22	-.27
7-7;11	97.98	8.98	75-115	-.21	-.37
8-8;11	99.33	9.54	71-116	-.29	-.56
9-9;11	101.86	13.86	55-123	-1.42	1.73
Total-Arabic Language Test					
6-6;11	170.94	20.90	115-224	-.16	.01
7-7;11	185.09	15.36	147-217	-.28	-.22
8-8;11	189.55	15.30	149-216	-.30	-.49
9-9;11	192.23	20.63	134-230	-1.13	.73
Nonword Repetition task					
6-6;11	37.23	7.98	7-48	-1.33	2.43
7-7;11	39.23	6.62	16-48	-1.27	2.09
8-8;11	38.89	6.66	13-48	-1.18	2.02
9-9;11	39.26	7.04	5-48	-1.52	4.84

a-raw scores (z-scores)

Appendix B

Ethics Approval (1)



Research Ethics

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Lisa Archibald
 File Number: 103912
 Review Level: Full Board
 Approved Local Adult Participants: 0
 Approved Local Minor Participants: 800
 Protocol Title: Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Children Learning English as an Additional Language
 Department & Institution: Health Sciences/Communication Sciences & Disorders, Western University
 Sponsor: Natural Sciences and Engineering Research Council

Ethics Approval Date: August 16, 2013 Expiry Date: September 30, 2016

Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Instruments	Arabic Language Test (Shaalaa, 2010)	
Instruments	Arabic Language Screening Test (El-Halees and Wiig, 1999)	
Instruments	Arabic Receptive- Expressive Vocabulary Test (El-Halees and Wiig, 1999)	
Instruments	Arabic Picture Vocabulary Test (APVT) (Shaalaa, 2010)	
Instruments	Arabic word reading task of the Arabic Lists of Sight Words (Oweini and Hazoury, 2010)	
Instruments	Arabic Nonword Repetition Test (Shaalaa, 2010)	
Instruments	Invented Language Rule Task (morpheme learning) (Connell, 1987)	
Instruments	Receptive Vocabulary (Peabody Picture Vocabulary Test-4 (Dunn and Dunn, 2006)	
Instruments	Expressive Vocabulary Test-2 (Williams, 2006) WASI Vocabulary (Wechsler, 1999)	
Instruments	Test of Word Reading Efficiency (Torgesen et al., 1999)	
Instruments	English Sentence Recall Task (Archibald & Joannis, 2009)	
Instruments	Test of Non-verbal Intelligence (TONI-3) (Brown et al., 1997)	
Instruments	English Nonword Repetition Test (Dollaghan & Campbell, 1998)	
Instruments	(Arabic-English) The Intelligibility in Context Scale (ICS) (McLeod, Harrison, & McCormack, 2012)	
Instruments	ELL's Parent Interview Questionnaire (Waisman Center, University of Wisconsin-Madison, Kaushanskaya, Gross, & Buac, 2010)	
Western University Protocol		2013/05/24
Letter of Information	LOI_monolingual arabic	2013/07/22
Other	Response letter describing changes in response to ethics review.	
Letter of Information	LOI_monolingual English	2013/09/09
Other	Consent Form (to be used with all LOIs)	
Letter of Information	LOI_ELL	2013/09/09

This is to notify you that The University of Western Ontario Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the applicable laws and regulations of Ontario has granted approval to the above named research study on the approval date noted above.

This approval shall remain valid until the expiry date noted above assuming timely and acceptable responses to the NMREB's periodic requests for surveillance and monitoring information.

Members of the NMREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussions related to, nor vote on, such studies when they are presented to the NMREB.

The Chair of the NMREB is Dr. Riley Hinson. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Ethics Officer to Contact for Further Information

<input checked="" type="checkbox"/> Grace Kelly (gkelly@uwo.ca)	<input type="checkbox"/> Vikki Tran (vikki.tran@uwo.ca)	<input type="checkbox"/> Shamel Walcott (swalcott@uwo.ca)
--	--	--

This is an official document. Please retain the original in your files.

Appendix C

Ethics Approval (2)



Research Ethics

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Lisa Archibald
File Number: 103912
Review Level: Delegated
Protocol Title: Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Children Learning English as an Additional Language
Department & Institution: Health Sciences/Communication Sciences & Disorders, Western University
Sponsor: Natural Sciences and Engineering Research Council

Ethics Approval Date: January 31, 2014 **Expiry Date:** September 30, 2016

Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Amendment	Jan 9, 2014. Rev to Protocol_AWMA Test	
Instruments	AWMA Scoresheet	

This is to notify you that The University of Western Ontario Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the applicable laws and regulations of Ontario has granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above.

This approval shall remain valid until the expiry date noted above assuming timely and acceptable responses to the NMREB's periodic requests for surveillance and monitoring information.

Members of the NMREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussions related to, nor vote on, such studies when they are presented to the NMREB.

The Chair of the NMREB is Dr. Riley Hinson. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 0000941.

Ethics Officer to Contact for Further Information

<input checked="" type="checkbox"/> Grace Kelly <small>(grace.kelly@uwo.ca)</small>	<input type="checkbox"/> Vikki Tran <small>(vikki.tran@uwo.ca)</small>	<input type="checkbox"/> Mina Mekhail <small>(mmekhail@uwo.ca)</small>	<input type="checkbox"/> Erika Basile <small>(ebasile@uwo.ca)</small>
--	---	---	--

This is an official document. Please retain the original in your files.

CURRICULUM VITA

Name: Areej Mazin Balilah

- Education:**
- Bachelor's degree with First Honors, from the Childhood Studies field, King Abdul Aziz University (2006).
 - Master's degree in Speech and Language Science from Western University (2012).

Honours and Awards: Book Award, contest of The Ministry of Education, National Commission for Childhood. Saudi Arabia (2010).

Publications:

Accepted manuscripts:

- Balilah, A., Archibald, L.M.D. (in press). The Measurement of Language Ability and Impairment in Arabic-speaking Children. Revisiting the assessment of second language abilities: From theory to practice. Springer.
- Balilah, A., Rafat, Y., Archibald, L.M.D. (in press). Domain-Specific and Domain-General Processing Accounts in Children with Specific Language Impairment (SLI): Contribution of Cross Linguistic Evidence. Arab Journal of Applied Linguistics.

Manuscripts under review:

- Balilah, A., Archibald, L.M.D. (under review). Sentence Recall and Single Word Reading in Monolingual Children and Same-age English Language Learners (ELLs) with and without Parental Concerns about Language Development. Canadian Journal of Speech Language Pathology and Audiology.

- Balilah, A., & Archibald, L. M. D. (under review). Linguistic and Cognitive Factors Sensitive to Language Performance Differences in Arabic-Speaking English Language Learners (ELLs). *Bilingualism: Language and Cognition*

Other Publications

- 12 children short stories about children rights, under the supervision of the Ministry of Culture and Media-Saudi Arabia, and in association with UNICEF (2009).
- Balilah, A. (2010). *My Child and I, and the Computer*. Saudi Arabia, The Ministry of Education, National Commission for Childhood, 978-9960-48-824-0.

Peer-Reviewed Posters

- Kwok, Balilah, Albakri, Elbadri, Wang, Oram Cardy, & Archibald, Quantifying Intratest Performance Variability on Standardized Language Measures in Arabic-speaking English Language Learners (ELLs), Children with Developmental Language Disorder, and Typical Controls, Madison, Wisconsin, June, 2017.
- Balilah & Archibald, The Measurement of Language Ability and Impairment in Arabic speaking Children, Madison, Wisconsin, June, 2016.
- Balilah, & Archibald, Investigating a working memory advantage in bilingual Arabic-English children, Workshop on Bilingualism and Executive Function: An Interdisciplinary Approach, New York, May, 2015.
- Balilah, & Archibald, Linguistic and Cognitive

Factors Sensitive to Language Performance
Differences in English Language Learners (ELLs),
Madison, Wisconsin, June, 2015.

- Balilah, & Archibald, Linguistic and Cognitive Factors Differentiating Arabic-speaking Children with and without Specific Language Impairment (SLI), Madison, Wisconsin, June, 2014.
- Balilah & Archibald, The Language Characteristics of Arabic Speakers with Specific Language Impairment (SLI), Symposium on Research in Child Language Disorders, Madison, Wisconsin, June, 2013.
- Balilah, & Archibald, Sentence Recall and Single Word Reading in Monolingual and ELL with and without Parental Concerns about Language Development, Child Language Seminar, The University of Manchester, June, 2013.
- **Teaching Assistant.** CSD 9515- Speech-Language Pathology for Audiology / Winter, 2013
- **Member of the Teaching Council.** (2006-Present), Childhood Studies Section – Faculty of Home Economics-King Abdul Aziz University.

**Related Work
Experience:**