ACTA UNIVERSITATIS UPSALIENSIS
Studia Linguistica Upsaliensia
27

Words and non-words

Vocabulary and phonological working memory in Arabic-Swedish-speaking 4–7-year-olds with and without a diagnosis of Developmental Language Disorder

Linnéa Öberg

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Humanistiska teatern, Engelska parken, Thunbergsvägen 3H, Uppsala, Saturday, 12 December 2020 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Laurice Tuller (University of Tours).

Abstract
Öberg, L. 2020. Words and non-words. Vocabulary and phonological working memory in Arabic-Swedish-speaking 4–7-year-olds with and without a diagnosis of Developmental Language Disorder. *Studia Linguistica Upsaliensia* 27. 269 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1046-6.

This thesis investigates the vocabulary skills and the non-word repetition (NWR) performance of 99 typically developing (TD) 4 –7-year-old Arabic-Swedish-speaking children and 11 Arabic-Swedish-speaking children with a diagnosis of Developmental Language Disorder (DLD). The children's early language development, family backgrounds and language exposure patterns are explored through parental questionnaires, and for the DLD children also via interviews with parents, teachers and speech-language pathologists regarding their developmental history, language skills and communicative behaviour. Vocabulary comprehension and production is assessed with Cross-linguistic Lexical Tasks (CLT; Haman et al. 2015) in Arabic and Swedish. Phonological working memory is assessed with four different NWR tasks with varying item length, phonological complexity and language-likeness (Radeborg et al. 2006, Chiat 2015, Abou Melhem et al. 2011). For vocabulary, differences between the two languages (Arabic and Swedish) and differences between comprehension and production are explored, as well as effects of age, exposure and socioeconomic status (SES). For NWR, effects of age, task, item length and phonological complexity are investigated, as well as effects of vocabulary and exposure.

Results: Vocabulary comprehension and production scores were found to increase with age in both Arabic and Swedish. Daily language exposure predicted comprehension and production scores in Arabic, but only production scores in Swedish. Length of exposure to Swedish was the most important predictor of Swedish vocabulary scores. SES (parental education) did not predict vocabulary scores in either language. For NWR, scores increased with age on all tasks. There were also task and item effects. Factors related to NWR performance were type of task, item length, phonological complexity and vocabulary skills.

At group level, the DLD children scored below their TD peers on both vocabulary and NWR tasks. Many DLD children had particularly low vocabulary scores in their first language (Arabic), despite extensive and continuous exposure from birth. There was substantial overlap between the TD and the DLD groups on NWR performance, and not all DLD children scored low on NWR. Having a history of language delay or language difficulties in the family was more common among the DLD children than the TD children. The study underscores the importance of considering patterns of language exposure and developmental history when assessing the language skills of bilingual children with potential DLD.

Keywords: bilingualism, Arabic, Swedish, preschool children, vocabulary, phonological working memory, CLT, NWR, language exposure, developmental language disorder, DLD

Linnéa Öberg, Department of Linguistics and Philology, Box 635, Uppsala University, SE-75126 Uppsala, Sweden.

ISSN 1652-1366
ISBN 978-91-513-1046-6
urn:nbn:se:uu:diva-421590 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-421590)

Acknowledgements

The past four and a half years have certainly been challenging and stressful at times, but they have also been immensely fun, stimulating and rewarding. First of all, I would like to thank my supervisor Ute Bohnacker. I am very grateful to have had the opportunity to pursuit a PhD in the BiLI-TAS project. Throughout this process, you have always been available and encouraging. You have generously shared your time and expertise whenever I needed your advice. Thank you for taking the time to read and comment on my work so thoroughly, it has substantially improved the quality of this thesis.

I would also like to thank Anette Månsson, who was my second supervisor during the initial stages, for sharing your knowledge about the Arabic language. Furthermore, I would like to express my gratitude to Carolyn Letts, my 'mock defense' opponent – your valuable feedback was very helpful when writing the final chapter.

I am also grateful to the following grant foundations whose generous support enabled me to attend conferences, workshops and winter schools during my time as a PhD student: *Anna-Maria Lundins stipendiefond*, *Jerringfonden* and *Stiftelsen Lundberska Idofonden*. Furthermore, I am thankful to the Swedish Research Council, *Vetenskapsrådet*, for funding the BiLI-TAS-project (VR 421-2013-1309, Bohnacker).

I would also like to express my gratitude to all the children and their parents who agreed to participate in the project, and for serving me Arabic coffee when I came to visit for data collection (*tack!* and *shukran!*). Furthermore, I would like to thank all (pre)schools, associations and congregations who helped recruiting children and who allowed us to visit them for data collection.

I was extremely fortunate to be part of the BiLI-TAS team. This thesis could not have been written without your support! I have many fond memories from team meetings, lunches, conferences and dinners. My warmest thanks to Ute Bohnacker (again!), to Rima Haddad (for the many fun and sometimes frustrating moments we shared during seemingly endless hours of data collection, and for transcribing and scoring the Arabic data), to Josefin Lindgren (for giving me advice about (PhD) life, statistics and teaching, and for helping me with data wrangling, coding and analyses), and to Buket Öztekin (for cheering me up when I needed it the most). I would also like to thank Karin Koltay (for always being so enthusiastic and supportive) and Sibylle Dillström (for bringing semlor and other treats to the team meetings), who were part of the BiLI-TAS-team in the beginning of my PhD journey.

Furthermore, I would like to thank Pascale Wehbe, Zeinab Shareef, Alaa Arifeh, Amal Choumar and Frauke Jonsson, who helped with recruitment, data collection and transcription.

I was also very lucky to have been surrounded by so many great people at the Department of Linguistics and Philology: Aaron Smith, Ali Basirat, Armin Chiocchetti, Artur Kulmizev, Ellinor Lindqvist, Emil Lundin, Erik Mo Welin, Fabienne Cap, Fredrik Sixtensson, Fredrik Wahlberg, Gongbo Tang, Harald Hammarström, Jenny Rahbek, Klara Bertils, Mahmut Agbaht, Maja Robbers, Marc Tang, Marie Dubremetz, Miryam de Lhoneux, Patrick Hällzon, Phillip Rönchen, Pär Eliasson, Rune Rattenborg, Sara Stymne, Simon Magnusson and Yan Shao. I have many fond memories from lunch time discussions, LingSing rehearsals and Fridays at William's. I would particularly like to thank *Klara* for (if only temporarily) taking my mind off the thesis with conversations about the ideal foot strike and choir singing, and for making sure that I ate healthy meals even when things were very stressful towards the end. Additionally, I want to thank Anna Ingves, Fredrika Nyström, Karin Senter and Maria Håkansson Ramberg, fellow PhD students at the Faculty of Languages, for good company during lunches and activities with Språk & Lärande and courses we took together.

Turning to the world outside university, I would like to thank the Janglerz: *Isa*, *Ellinor* (again!) and *Ylva* for your encouragement and for all the entertaining messages and updates, and *Elsa* for taking the time to meet up whenever you were in Uppsala (or I in Skåne) and for always being so enthusiastic about my thesis project.

Last but not least, I would like to thank my family. I am thankful to my parents *Margaretha* and *Åke* for always standing by me and believing in me. *Mamma*: for as long as I can remember you have had a great thirst for knowledge; your curiosity has always inspired me to learn new things, and to find joy in asking *how* and *why* things are the way they are. *Pappa*: even though I never became any kind of engineer, I trust that you are happy with the fact that my career choice at least has a solid foundation in your favourite guiding principles: science and proven experience. Furthermore, I would like to thank my sisters, my big sister *Lisa*, and my little sister *Veronika* – you are both inspirations to me. *Lisa*: like a true big sister, you have always led the way; thank you for always being there whenever I need some big-sisterly advice. *Veronika*: you work so tirelessly and hard to achieve your goals – next time we will celebrate your degree!

And finally, *Daniel* – you have been by my side through all the ups and downs related to the PhD project. Thank you for always being so optimistic, supportive and for constantly pushing me outside my comfort zone. I truly could not have done this without you.

Uppsala, October 2020
Linnéa Öberg

Contents

Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
BESA	Bilingual English Spanish Assessment
BNT	Boston Naming Test
BPVS	British Picture Vocabulary Scale
CDI(s)	(MacArthur-Bates) Communicative Development Inventories
CELF	Clinical Evaluation of Language Fundamentals
CLT(s)	Cross-linguistic Lexical Tasks(s)
CNRep	Children's Non-word Repetition test
DLD	Developmental Language Disorder
EOWPVT	Expressive One Word Picture Vocabulary Test
EVIP	Échelle de Vocabulaire en Images de Peabody
ID	Intellectual Disability
L1	First Language
L2	Second Language
LS	Language-Specific
NRT	the Non-word Repetition Test
NVIQ	Non-Verbal Intelligence Quotient
NWR	Non-Word Repetition
PPVT	Peabody Picture Vocabulary Test
PRT	the Preschool Repetition Test
QU	Quasi-Universal
(S)LI	(Specific) Language Impairment
SLP	Speech-Language Pathologist
TD	Typically Developing
WM	Working Memory

1 Introduction

A substantial proportion of children who grow up in Sweden today are regularly exposed to more than one language. Some of them are being exposed to two languages from birth, while others start to receive regular exposure to an additional language sometime during childhood. Although Sweden does not keep official statistics on the number of speakers of different languages, such numbers can be inferred by consulting related information. For instance, the proportion of children age 0–18 with a so-called 'foreign background' amounted to 38% in 2019 (Statistics Sweden, 2020), and it can be assumed that the vast majority of them are being exposed to another language in addition to Swedish.[1] Further estimations regarding the number of bilingual children in Sweden can be made by investigating the number of children who are entitled to so-called mother tongue instruction, MTI (also referred to as heritage language education). During the academic year of 2018/2019, 28% of the pupils in grade 1–9 (corresponding to age 7–15) were entitled to MTI instruction (Swedish National Agency for Education, 2019).[2] During the past decade, the number of Arabic speakers have increased rapidly in Sweden, and Arabic is now considered to be the second largest language in Sweden after Swedish.[3]

Expanding our knowledge about bilingual language development is important for several reasons. Bilingual children are at a greater risk of falling behind academically compared to their monolingual peers, if they have not developed sufficient language skills in the language of schooling when starting school (Ardasheva, Tretter, & Kinny, 2012). Second, confusion among educational and health care staff about what should be considered 'normal' language development in bilingual children may lead to both over- and underidentification of developmental disorders of language,

[1] In the official statistics, 'foreign background' (swe: 'utländsk bakgrund') is defined as either being born abroad, or as having two parents born abroad. Thus, children with only one parent born in a foreign country are not considered to have a foreign background (Statistics Sweden, 2020).

[2] According to the Swedish Education Act, pupils are entitled to mother tongue instruction if at least one parent speaks the language, if it is used on a daily basis in the home, and if the pupil has basic knowledge in the language in question (SFS 2010:800, chapter 10, §7).

[3] However, as Parkvall (2015, p. 154) points out, it can be questioned whether Arabic should be considered as one and the same language, due to diglossia in the Arabic-speaking parts of the world, and the vast variation within as well as between regions regarding spoken varieties.

communication and literacy among bilinguals (Dollaghan & Horner, 2011; Grimm & Schulz, 2014).

Developmental Language Disorder (DLD) is a condition that typically emerges in early childhood, and manifests as a pronounced deficit in the development of linguistic skills that cannot better be explained by hearing impairment, intellectual disability, medical syndromes, or neurological problems (Bishop, 1997, pp. 21–23; Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 consortium, 2017; Leonard, 2014, p. 3). Other terms are also used to refer to this condition, for instance (Specific) Language Impairment ((S)LI) and Primary Language Impairment (PLI). It is estimated that DLD affects around 7–10% of the population (Norbury et al., 2016; Tomblin et al., 1997).

Salameh, Nettelbladt, Håkansson and Gullberg (2002) investigated all referrals of suspected DLD during one year in the late 1990's in a national health care region in southern Sweden.[4] They found that bilingual children were generally referred at a later age (after age five) compared to monolinguals, and the bilingual children were also overrepresented in the group of children who were determined to have severe DLD. Over a decade later, Nayeb, Wallby, Westerlund, Salameh and Sarkadi (2015) found that a high proportion (82%) of Swedish child healthcare nurses believed that bilingual children show slower language development than monolinguals, and as a result, they were more inclined to simplify language screening for bilinguals and delay referrals for speech-language pathologist (SLP) assessment. In 2015, a survey was conducted in five national health service regions in Sweden, investigating the prevalence of severe DLD among children in grade 1–3 (corresponding to age 7–9). Although there was much variation between the regions, bilingual children were overall heavily overrepresented (51%) in the group of children who were determined to have severe DLD. What is more, among the group of children where the participating SLPs could not determine whether the diagnosis should be considered as severe, around half of the children were bilingual or had another mother tongue than Swedish (SOU 2016:46, pp. 222–223). Further evidence of the overrepresentation of bilinguals among children with severe DLD lies in the fact that during the autumn of 2015, around 47% of the pupils attending the national school for children with severe DLD, Hällsboskolan, were eligible for MTI (SOU 2016:46, pp. 591–592).

What is the reason for this confusion regarding 'typical' vs. 'atypical' language development in bilingual children? The problem is linked to the fact that there is substantial overlap in the linguistic features that characterise second language acquisition and DLD (Boerma, Wijnen, Leseman, & Blom, 2017; Goral & Conner, 2013; Håkansson, 2001; Kohnert, 2010; Paradis, 2005; Paradis & Crago, 2000). Bilingual children may score below monolingual

[4] The term used by Salameh et al. was 'LI'.

norms on standardised language tests, leading to possible overidentification of DLD (Andersson et al., 2019; Peña, Bedore, & Kester, 2016). However, the opposite scenario is also true; if healthcare professionals think that delayed language development is associated with bilingualism, it may lead to underidentification of DLD. Armon-Lotem (2012) underscores that this 'diagnostic dilemma' can be resolved by integrating research evidence concerning bilingualism and developmental language disorders.

Apart from performing low on formal language tests, some common characteristics have been found to be associated with DLD. Although they do not apply to all children with DLD, three main *risk factors* will be described below.

First, children with DLD are often delayed in their early language development. For instance, children with a late debut of first words or first word combinations, so called *late talkers*, run an increased risk of developing persistent language disorders (Paradis, Emmerzael, & Sorenson Duncan, 2010; Trauner, Wulfeck, Tallal, & Hesselink, 2000). Contrary to popular belief, late emergence of first words and multi-word utterances are not characteristic of bilingual language development. However, bilingual children may not reach these developmental milestones at the same time in *both* languages (Hoff et al., 2012).

Second, DLD is subject to *heredity*, and there seems to be a genetic basis for this finding (Bishop, North, & Donlan, 1995; DeThorne et al., 2006; Tomblin & Buckwalter, 1998). Children with DLD are more likely to have a close relative with language, communication and/or literacy problems compared to their typically developing peers (Kalnak, Peyrard-Janvid, Sahlén, & Forssberg, 2012).

Third, at group level, parental evaluations of their children's language skills have been found to differentiate between bilingual children with a DLD diagnosis and typically developing bilingual children (Restrepo, 1998). Also, *parental concern* about their children's language development has been identified as a predictor of the presence of DLD, and parental concern seems to increase with the level of severity of the disorder (Salameh, Nettelbladt, Håkansson, et al., 2002).

Within the research area of childhood bilingualism, several aspects are known to influence the developmental trajectories of the minority language and the majority language.[5] These factors include age of onset, acquisition setting, status in society (including socioeconomic status, SES), input quantity and quality. Additionally, it is important to consider which language

[5] In this thesis, different terms are being used interchangeably to refer to the language spoken in the family (in this case, Arabic), namely *minority language*, *first language (L1)*, *home language* and *mother tongue*). The language spoken in the wider society (in this case, Swedish) is referred to as the *majority language* or the *second language (L2)*.

combination the child is acquiring. Six major influencing factors will be briefly described in the following.

First, it is important to underscore that not all bilingual children become bilingual at the same age. Some children grow up with two languages from birth (simultaneous bilinguals), while others start to learn a second language later on during childhood (sequential bilinguals). *Age of acquisition* or *length of exposure* to a language is an important predictor of language proficiency (DeKeyser, 2000; Johnson & Newport, 1989; Paradis, 2011; Unsworth, 2013b).

Second, the *acquisition setting* is an influential predictor of language development. For instance, some children grow up with two languages, where all the input in one language comes almost exclusively from one parent. For other bilingual children, the minority language may be prominent in several environments (such as in preschool/school, at the playground, at religious service centres or in their local community), and the child may have a wide range of interlocutors to speak the language with (such as extended family and friends) (Kohnert, 2010; Pearson, 2007).

Third, aspects related to *social status* are known to influence language development in bilinguals. For instance, the *status of the language in society* plays a role. As Pearson (2007, p. 402) argues, achieving the same proficiency level in the minority language as in the majority language may take more conscious effort, since input in the majority language is present everywhere in the majority language community and the media, and children are naturally drawn to it, for instance as a result of schooling. In a number of studies, the *socioeconomic status* (SES) of the family has been shown to influence children's language acquisition, in particular vocabulary. Bilingual children from high-SES homes often outperform their bilingual peers from lower-SES backgrounds on vocabulary tests in the majority language (Buac, Gross, & Kaushanskaya, 2014; Calvo & Bialystok, 2014; Cobo-Lewis, Pearson, Eilers, & Umbel, 2002a; Prevoo et al., 2014). With respect to the minority language, some evidence seems to suggest that bilinguals from lower-SES households may perform higher than high-SES children at least in some vocabulary tasks (Cobo-Lewis, Pearson, Eilers, & Umbel, 2002b), while other studies have not found such an effect (Buac et al., 2014; Prevoo et al., 2014).

Fourth, the *quantity and quality of language input* is known to strongly influence bilingual language development. The relative amount of exposure to each language may vary substantially between individual bilingual children.[6] The relative amount of exposure to each language (input quantity) has been found to be directly related to the performance on measures of vocabulary size (V. C. M. Gathercole, Thomas, Roberts, Hughes, & Hughes,

[6] Among the children participating in the present study, the relative amount of daily exposure ranges from 5% Arabic and 95% Swedish to 95% Arabic and 5% Swedish according to parental reports.

2013; Prevoo et al., 2014; Thordardottir, 2011) as well as certain aspects of morphosyntax (V. C. M. Gathercole & Thomas, 2009). Input quality refers to the diversity, richness and complexity of the language input. Receiving high quality input from parents (for instance in the form of decontextualized language and lexical diversity) has a positive effect on the vocabulary development in children (Hart & Risley, 1995; Rowe, 2012). Parental use of facilitative communication strategies may increase the vocabulary size in young children (Cartmill et al., 2013; Hart & Risley, 1995; Rowe, 2012).

Fifth, it is important to consider the specific *language combination* a bilingual child is acquiring. If the languages are closely related, they may share morphosyntactic patterns and have a large proportion of cognates, which may serve as a facilitating factor, boosting vocabulary knowledge in both languages, but especially in the weaker language (Lindgren & Bohnacker, 2020).

Finally, bilingual children's *linguistic knowledge is distributed* over two (or more) languages (Pearson, Fernández, & Oller, 1993). Some children may have similar proficiency in both languages while others may have very unbalanced proficiency levels, which is related to variability in bilingual children's language input. Because of this distribution of language knowledge, it is essential to assess both languages in order to obtain a fair representation of the child's full language competence.

If healthcare professionals are aware of which characteristics of DLD hold for both monolinguals and bilinguals, and are knowledgeable about factors influencing bilingual language development, they may be more confident in making accurate assessments of bilinguals with suspected DLD. Assessing both languages at the same time as collecting information about early language development and input patterns may give insights into whether there is an imbalance in proficiency levels, as well as potential DLD. On the surface, assessing both languages may seem like a straightforward endeavour, yet, there are many questions that remain. The first challenge is that there is often a lack of suitable assessment materials, and what is more, in many languages very little is known about what characterises DLD in that language. Language disorders are often claimed to manifest in both languages of a bilingual child (Kohnert, 2010; Salameh, Nettelbladt, Håkansson, et al., 2002). However, due to differences in proficiency level and typological differences between the languages DLD may not manifest in the same way in both languages. Second, it may be difficult to identify a suitable group for comparison. Ideally, the child should be compared against a group of children speaking the same language combination and growing up in the same country. Even if there is a sizeable minority in the country in question, it is rarely the case that norming or reference data exists for language proficiency in that population. What is more, when such norming samples exist, it is not clear which the suitable cut-off score for a particular language test is for identifying DLD. Several researchers suggest that exposure should be taken into account when making

such decisions (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013; Thordardottir, 2015).

1.1 The present study

This study is part of the BiLI-TAS project (Bohnacker, 2013), a six-year research project funded by the Swedish Research Council (Vetenskapsrådet, VR), with Ute Bohnacker as the PI.[7] The project explores typical and atypical language development in bilingual Turkish-Swedish-speaking and Arabic-Swedish-speaking children growing up in Sweden, and develops new methods for the linguistic assessment of these groups. The project encompasses three sub-studies for each language group: first, a large cross-sectional study including ca 100 typically developing children aged 4–7, second, a longitudinal follow-up study, and finally a clinical study of children with a DLD diagnosis. Here, *typically developing* refers to children who do not have a DLD diagnosis, and have (supposedly) typical language development. In all sub-studies, data is collected in both languages by native or near-native speakers of the respective languages, using assessment materials that are comparable between the languages. In the longitudinal follow-up, ca ten of the youngest children in the cross-sectional study are seen again two years later and assessed with the same materials as before, which allows for an investigation of language development over time. In the clinical study, children who have received a DLD diagnosis are assessed using the same materials as in the cross-sectional study, allowing for a comparison between the children in the clinical group and their typically developing peers in the cross-sectional sample. Background information is obtained via parental questionnaires in all studies, and via interviews with parents in the longitudinal and clinical studies. Additionally, teachers and speech-language pathologists (SLPs) are interviewed in the clinical study.

In the BiLI-TAS project, three main areas of language abilities and language processing are being investigated: non-word repetition (NWR, targeting phonological working memory), vocabulary (comprehension and production), and narratives (comprehension and production). A majority of the assessment tasks utilised in the project were developed within the COST Action IS0804 'Language Impairment in a multilingual society: Linguistic patterns and the road to assessment', a research network running 2009–2013, with the objective to explore the linguistic and cognitive abilities of bilingual children with SLI in different migrant communities.[8]

[7] BiLI-TAS is an acronym for 'Bilingual Language Impairment, Turkish, Arabic, Swedish', and is used colloquially to refer to the research project 'Language Impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden'.
[8] Several of these tools are described in Armon-Lotem, de Jong, & Meir (2015).

In the present study, the language skills of 99 Arabic-Swedish-speaking children without a DLD diagnosis and 11 Arabic-Swedish-speaking children with a DLD diagnosis aged 4–7 are investigated. Language assessment materials that allow for a comparison between proficiency in the two languages (Arabic and Swedish) are used. Additionally, the linguistic environment and family background is explored through parental questionnaires with questions about input patterns, early language development and family history. It is the first study to use the COST-Action material Cross-Linguistic Lexical Task (CLT; Haman, Łuniewska, & Pomiechowska, 2015), to study both vocabulary comprehension and production in a large group of Arabic-Swedish-speaking bilinguals. Furthermore, it is the first large-scale study systematically investigating the performance on different non-word repetition (NWR) tasks with different phonological setups for Arabic-Swedish-speaking bilinguals aged 4–7. Several of the NWR tasks are COST-Action materials, including a Swedish and an Arabic version of the quasi-universal non-word repetition task (Chiat, 2015) and the NWRT-Lebanese (Abou Melhem, Kouba Hreich, & dos Santos, 2011). Additionally, a Swedish NWR task is used (Radeborg, Barthelom, Sjöberg, & Sahlén, 2006).

There were several reasons for investigating these specific language skills using these specific tasks. First, vocabulary is an important cornerstone of language proficiency. Vocabulary is also a language domain that is heavily influenced by language input. Thus, if there is an imbalance in proficiency between the languages, it is likely to be reflected in unevenly sized vocabularies (V. C. M. Gathercole et al., 2013; Hoff et al., 2012; Hoff, Rumiche, Burridge, Ribot, & Welsh, 2014; Thordardottir, 2011). By contrast, NWR is a task targeting phonological working memory, and it is much less affected by language input than vocabulary (Meir & Armon-Lotem, 2017; Thordardottir & Brandeker, 2013). At the same time, NWR has been identified as a clinical marker of DLD in both monolingual populations (Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Kalnak, Peyrard-Janvid, Forssberg, & Sahlén, 2014; Saiegh-Haddad & Ghawi-Dakwar, 2017; Topbaş, Kaçar-Kütükçü, & Kopkalli-Yavuz, 2014) and bilingual populations (Abed Ibrahim & Hamann, 2017; Boerma et al., 2015; dos Santos & Ferré, 2018; Thordardottir & Brandeker, 2013).

To summarise, this thesis aims to shed light on how vocabulary skills and NWR performance develop in Arabic-Swedish-speaking 4–7-year-olds, using a cross-sectional design. Another aim is to characterise these children in terms of their language input, early language development and (family) background history, and to investigate how these aspects are connected to the children's language proficiency in both languages. Finally, the thesis investigates similarities and differences in the linguistic, communicative and social characteristics of Arabic-Swedish-speaking bilinguals with and without a DLD diagnosis. Although one of the aims of the study is to explore how DLD

manifests in Arabic-Swedish-speaking 4–7-year olds, the main focus will be on the children in the cross-sectional study who (supposedly) have typical language development (TD). The data and conclusions presented in this thesis thus form a knowledge base for how vocabulary and NWR skills develop in Arabic-Swedish-speaking 4–7-year-olds (the important age span that constitutes the transition from preschool to formal schooling), an age where assessment for suspected DLD is often conducted. This study may provide insights for healthcare and educational professionals, so that they can make well-informed judgments regarding the language abilities of bilingual children. Moreover, the conclusions drawn from this study form a basis for further research beyond the scope of this thesis.

1.2 The Arabic-speaking population in Sweden

The following subsection contains a brief description of the background characteristics of the Arabic-speaking population in Sweden. It is based on an unpublished internal report by Ute Bohnacker, PI of the BiLI-TAS project (Bohnacker, 2017).

As already mentioned, there are no official statistics on the number of speakers of different languages in Sweden. Therefore, the number of speakers of a language can only be inferred, for instance by investigating the number of inhabitants in Sweden who are born in a certain country where the language is spoken. Arabic is spoken in various countries in the Middle East. The linguistic situation in many of these countries is complex; the Arabic varieties spoken differ both within and between national borders, and there is diglossia where Modern Standard Arabic is used in writing and spoken in certain formal contexts. In addition to this, other languages are also prominent in the region, for instance Kurdish, Aramaic/Syriac, and Armenian. Thus, one cannot assume that people originating from a country where Arabic is the main language will have Arabic as their mother tongue.

The main influx of immigration to Sweden from Arabic-speaking countries has occurred from 1980 until today, from three countries in the Middle East: Iraq, Lebanon and Syria. The immigrants have mainly come to Sweden as refugees following wars and conflicts in the home countries. The earliest large wave of migration to Sweden from an Arabic-speaking country came from Lebanon following the civil war in 1975. A large group of Iraqis followed after the Gulf war in the early 1990's, and later the Iraq war in the early 2000's. The latest large wave of Arabic-speaking immigrants were Syrians, fleeing the civil war starting in 2011. During recent years (2009–2016), the vast majority of the people from an Arabic-speaking country who were granted a residence permit in Sweden came from Syria (56 807) and Iraq (3 719). In 2016, the majority of the individuals from Syria and Iraq who were granted a residence permit received them due to asylum (Syria: 80%, Iraq: 50%). The

second most common reason for being granted a residence permit was family reunification (Syria: 18%, Iraq: 39%) (Swedish Migration Agency, 2017).

According to Statistics Sweden, there were 362 597 people originating from nineteen countries where Arabic is the main language who resided in Sweden as of 31 December 2016 (Statistics Sweden, 2017a).[9] The vast majority of these people originated from Syria (41%), Iraq (37%) and Lebanon (7%). In addition to the people residing in Sweden who were born in an Arabic-speaking country, there are also people who were born in Sweden with one or both parents from an Arabic-speaking country. In 2016, their number amounted to roughly 180 000. Thus, although not all people residing in Sweden who were born in an Arabic-speaking country, or who have at least one parent from an Arabic-speaking country speak Arabic, the total number of Arabic speakers in Sweden is likely to exceed 400 000.

Today, most people with a background in an Arabic-speaking country reside in the three large metropolitan areas of Stockholm, Malmö and Göteborg, and there is also a sizeable Arabic-speaking minority in Södertälje, a town adjacent to Stockholm (Statistics Sweden, 2017b; Swedish National Agency for Education, 2017).

In Sweden, mother tongue instruction, MTI, has been offered since the 1970's. Of all the languages that are offered in Swedish MTI Arabic has the highest number of eligible pupils nation-wide, and it has been so for the past ten years. During the academic year of 2016/2017, 23% (64 261) of the total number of children (275 329) who applied for MTI requested instruction in Arabic. In comparison to other languages, Arabic has a high attendance rate, with 66% of eligible children attending MTI.[10] Additionally 25% (7 948) of all children who attended *förskoleklass* (preparatory year between preschool and school) and were eligible for MTI instruction, spoke Arabic (Swedish National Agency for Education, 2017).

In sum, most people residing in Sweden with roots in an Arabic-speaking country have connections to Syria, Iraq or Lebanon. The exact number of Arabic speakers in Sweden is not known, but likely exceeds 400 000. Arabic is by far the most common language among children who are eligible for MTI instruction.

1.2.1 Language skills of Arabic-Swedish-speaking children

There are a handful of studies in a Swedish context that investigate Arabic-Swedish-speaking children's language skills in *both languages*. Common denominators for these studies are that they compare small groups of children,

[9] These countries are Syria, Iraq, Lebanon, Morocco, Egypt, Palestine, Tunisia, Saudi Arabia, Sudan, Jordan, Libya, Algeria, Kuwait, United Arab Emirates, Yemen, Qatar, Bahrain, Mauritania and Oman.

[10] There is also an unknown number of Arabic-speaking children who fulfil the criteria, but do not apply for MTI.

one group with typical language development (TD) and another group with DLD, on some language tasks.

In connection with Salameh's doctoral thesis from 2003, three articles were published comparing the Arabic and Swedish language skills of ten children with typical language development, and ten children with a diagnosis of DLD (or in their terms SLI), all aged between 3;10–6;7 (Salameh, 2003). Håkansson, Salameh and Nettelbladt (2003) investigated these children's language comprehension and expressive morphosyntax. They found that the TD group scored higher than the DLD group on a standardised test of language comprehension in Swedish.[11] Furthermore, the TD group generally had higher scores on a task probing morphosyntactic structures in both languages. A follow-up of the same children after six and twelve months showed that the TD group consistently scored higher than the DLD group on the morphosyntactic structures that were probed in both languages at all three sessions (Salameh, Håkansson, & Nettelbladt, 2004). The children's phonological development was investigated in both languages using picture naming tasks. At group level, the DLD children made more errors overall than the children in the TD group. What is more, the DLD group had more types of errors (e.g. segment deletions and syllable reductions), particularly in Arabic, compared to the TD children (Salameh, Nettelbladt, & Norlin, 2003).

In Holmström's doctoral thesis from 2015, vocabulary was investigated in 44 monolingual and 44 bilingual Arabic-Swedish-speaking children aged 5;11–9;3 with and without a DLD diagnosis (or in her terms, LI) (Holmström, 2015). In the first sub-study, 15 bilingual children with a DLD diagnosis were compared to 15 bilingual children without a DLD diagnosis on tasks targeting comprehension of school-related words and production of nouns (the Arabic versions were translated from Swedish). The children with a DLD diagnosis scored significantly below the TD children on all tasks, except from the Arabic production task. Another sub-study in the thesis explored the performance on a word association task (an adaptation of the Kent and Rosanoff list; Kent & Rosanoff, 1910), conducted in both languages in a longitudinal design, where children were seen twice (the second session was 12 months after the first). The task entails hearing a word and uttering the first associated word one comes to think of. Holmström classified the responses as sound-based, syntagmatic, paradigmatic or 'other', where syntagmatic associations were considered to be more advanced than sound-based, and paradigmatic associations were considered the most advanced. Holmström found that the children in the TD group had more associations that were classified as advanced at both sessions compared to the DLD group.

Salameh (2011) investigated vocabulary size and word associations in both languages in two groups of Arabic-Swedish-speaking bilinguals, one group

[11] Similar results were found for a standardised test targeting expressive morphosyntax, but the difference in scores was not significant between the groups.

who attended bilingual Arabic-Swedish instruction since first grade (N=16) and one group who attended monolingual instruction in Swedish (N=33). The children came from the same ethnically diverse area, and were in fourth grade (corresponding to age ten) at the time of testing. Salameh used a Swedish and an Arabic translation of the Peabody Picture Vocabulary Test (PPVT-III) to assess vocabulary comprehension. She found that the children who attended bilingual instruction received significantly higher scores in Arabic than the children in the monolingual instructional setting. Conversely, the children who received bilingual schooling performed lower on Swedish comprehension than the children in monolingual schooling.[12] The Kent and Rosanoff word association task was used to explore word associations in both languages. The same method of coding and classifying the associations was utilised as in the study by Holmström (2015), described above. Salameh found that the children in bilingual schooling had a larger proportion of associations that were classified as advanced in Arabic compared to the children in monolingual schooling.

In sum, although there are previous studies investigating the language skills of Arabic-Swedish-speaking children, there are no studies of larger groups of children with typical language development that would enable investigations of age development and age group comparisons. Furthermore, none of the previous studies have systematically investigated the influence of background factors such as SES, daily language exposure and length of exposure on the performance on language tasks.

1.3 Bilingualism and DLD

Previous research investigating the manifestations of DLD in monolinguals has largely focused on morphosyntactic features (often describes as 'clinical markers') in expressive language. However, clinical manifestations of DLD are language specific, i.e. a structure that is affected in one language may not be affected in another language (see Leonard (2014, Chapter 4) for an extensive overview of the manifestations of DLD across languages). Although the clinical manifestations of DLD are well studied in some languages (e.g. English), there is little to no research about either typical language development or clinical manifestations of DLD in many other languages. Many assessment materials have been developed for English, and they are often translated into other languages.[13] Letts (2013) points out that this may be

[12] Approaching significance ($p = .062$).

[13] These versions may be official translations provided by the publisher (e.g. the Swedish version of Test for Reception of Grammar-2 (TROG-2) (Bishop, 2009b), or unofficial translations used in clinical or research contexts (e.g. the Swedish and Somali versions of the Peabody Picture Vocabulary Test-IV (Dunn & Dunn, 2007) used by Ganuza and Hedman (2019)).

precarious for several reasons. First, since the manifestations of DLD vary between languages, translating test items may lead to probing structures that are not associated with DLD in that language. Additionally, some test items may be culturally unsuitable or have a different level of difficulty in one language compared to another language. Moreover, and importantly for a multilingual setting, morphosyntactic correlates of DLD in monolinguals often overlap with common features of second language acquisition (Boerma, Wijnen, et al., 2017; Håkansson, 2001; Håkansson & Nettelbladt, 1993; Paradis, 2005; Paradis & Crago, 2000). Typically developing bilingual children who are only assessed in the majority language may perform significantly below the norm compared to monolinguals on standardised language tests targeting vocabulary (Boerma, Leseman, Wijnen, & Blom, 2017; Peña et al., 2016) as well as general language skills (Andersson et al., 2019; Barragan, Castilla-Earls, Martinez-Nieto, Restrepo, & Gray, 2018). Thus, it is precarious to compare the language performance of a bilingual child against monolingual norms, as emphasised by Letts (2013) and Kohnert (2010). Ideally, bilingual children should be compared to children speaking the same language combination, and growing up in the same country as them. Unfortunately, relevant assessment materials, norms or reference data rarely exist for the bilingual population in question.[14]

1.3.1 Methodological considerations

In light of the above-mentioned difficulties in assessing language skills and correctly identifying DLD in bilinguals, the following considerations were taken into account when choosing the assessment materials and assessment procedure for the present study.

First, it was decided to include parental reports via questionnaires and interviews. Parental reports allow the researcher or clinician to receive information about factors that are influential in language development. Such information may provide indications on whether the slow or unexpected language development of a bilingual child may be due to a language disorder, or whether it can be explained by environmental factors such as the quantity or quality of language input.

Second, it was decided to include a language measure that is known to be highly dependent on exposure factors. If there is an imbalance in language exposure it is likely to be reflected in uneven proficiency levels in the two languages. In the current study, vocabulary tasks were utilised. Vocabulary is a linguistic domain that is known to be heavily influenced by language

[14] Notable exceptions include assessment materials and reference data available for the Spanish-English-speaking population in the USA (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013; Peña, Bedore, & Kester, 2016), and for French-English-speaking children in Canada (Dunn, Theriault-Whalen, & Dunn, 1993; Thordardottir, 2011).

exposure. Furthermore, measures of both vocabulary comprehension and vocabulary production are included, since exposure may influence comprehension and production differently (Thordardottir, 2011).

Third, it was decided to include a measure that is known to be a clinical marker associated with DLD, but at the same time has found to be minimally influenced by earlier language experience. Non-word repetition is such a measure. In the current study, phonological working memory is assessed using four different NWR tasks.

Finally, since very little is known about what typical language development looks like in Arabic-Swedish-speaking bilinguals, it was decided that the main focus of the study should be on children with typical language development. However, a smaller group of children with a DLD diagnosis are also assessed with the same assessment tools, and their performance is compared to that of the children with typical language development.

1.3.2 Assessment tools

This section contains a brief description of the tools that are used in the current study. The tools will be characterised in terms of what is being assessed, and the rationale for using them in identifying DLD in bilingual children. Detailed descriptions of the tools and how they are used are provided in the Methods chapter (section 3.1, section 3.2.1 and section 3.2.2). The assessment tools are *parental questionnaires* for assessing family history, early language development, linguistic environment and language use, *Cross-Linguistic Lexical Tasks* (CLTs; Haman et al., 2015), and *non-word repetition tasks* (Abou Melhem et al., 2011; Chiat, 2015; Radeborg et al., 2006) to assess phonological short term memory. An overview of the areas of assessment and the tools used is given in Table 1.1.

Table 1.1. Area of investigation and the tools used in the current study.

Area of investigation	Description of the tools
Background information	*Parental questionnaire* Key questions: • Early language development • Family history of language difficulties or literacy difficulties • Estimations of proficiency in both languages • Reported daily language input and language use
Vocabulary development	*Cross-Linguistic Lexical Tasks, CLTs* Assessment of vocabulary size in Arabic and Swedish: • comprehension and production • nouns and verbs
Phonological working memory	*Non-word Repetition Tasks*: • LS-Swe: 2–5 syllables, items complex syllable structure, matching Swedish phonotactics • QU-Swe: 2–5 syllables, items with simple syllable structure • QU-Ara: 2–5 syllables, items with simple syllable structure • NWRT-Leb: 1–3 syllables, items with complex syllable structure

1.3.2.1 Background information

Parental questionnaires and interviews have been proven to be particularly informative when assessing bilingual children with suspected DLD. Tuller (2015) highlights two major reasons for this. First, in situations where the clinician cannot directly assess the child's language skills in the minority language, parental reports may provide information about the proficiency in the minority language. Second, parental reports on the appearance of early language milestones, heredity for language disorders and estimations of language proficiency have proven to be useful in identifying those children who have a language disorder from those who have typical language development (Paradis et al., 2010; Restrepo, 1998). Four key areas of inquiry should be included in such a questionnaire.

First, the *early language development* should be outlined, particularly the appearance of the first word and the first word combinations. Many researchers suggest that late language emergence could be defined as the first word appearing after 18 months, the first word combinations appearing after 24 months, or having an expressive vocabulary of less than 50 words at 24 months (Fenson et al., 1994; Nettelbladt, 2007; Tuller, 2015). Investigations of both monolingual and bilingual populations have found that on average, the

first word commonly emerges at around 10–12 months and the first word-combination at around 17–20 months in children with typical language development. By contrast, these developmental milestones often appear later in children with a DLD diagnosis (Paradis et al., 2010; Trauner et al., 2000).

Second, parents should be asked whether there is a *family history of speech, language or literacy difficulties*. Having close family members (siblings, parents or grandparents) with such difficulties has been found to be more common in children with DLD than in TD children (Kalnak et al., 2012). Additionally, Restrepo (1998) found that having a family history of speech, language or literacy difficulties was a reliable identifier of DLD. However, Paradis, Emmerzael and Sorenson Duncan (2010) emphasise that the topic can be sensitive, and that there may be cultural as well as individual variation of the willingness to share such information.

Third, *parents' estimations of their child's language proficiency* have been found to reliably distinguish between children with DLD and their non-impaired peers at group level (Paradis et al., 2010; Restrepo, 1998). Parental evaluations of their children's language skills have also been shown to correlate with language skills as measured by standardised language tests, but there is not a perfect relationship (Patterson, 2000). Roberts, Burchinal and Durham (1999) suggest that parents from low-socioeconomic status/minority backgrounds may be particularly prone to over- or underestimate their children's emergent language skills. Additionally, parents may not be able to reliably evaluate their child's proficiency in the majority language if their own proficiency is low.

Finally, parental questionnaires should ask about *language input in both languages*. As previously mentioned, the relative amount of exposure to each language has been shown to directly influence proficiency, mainly for vocabulary but also certain aspects of grammar (V. C. M. Gathercole et al., 2013; Thordardottir, 2011; Unsworth, 2013b). Estimations of the relative amount of exposure to each language can be made by asking the parents how much the child hears language A and language B, respectively, during the day. Information on the age of acquisition (or length of exposure) can be obtained by asking the parents how old their child was when they started to receive regular exposure to each language. Additionally, further information about the quality of the language input can be obtained by asking whether the parents engage in book-reading activities with their child, and in which language.

The rationale for focusing on these background questions will be discussed further in the background section of the clinical chapter (Chapter 6).

1.3.2.2 Vocabulary development

Children with DLD often have deficits in the vocabulary domain, with a slower rate of vocabulary growth (Rice & Hoffman, 2015), and smaller vocabularies than their typically developing peers (Gray, Plante, Vance, & Henrichsen, 1999; Kapalková & Slančová, 2017; Khoury Aouad Saliby, dos

Santos, Kouba Hreich, & Messarra, 2017; Peña et al., 2016; Spaulding, Hosmer, & Schechtman, 2013; Thordardottir & Brandeker, 2013). At the same time, bilingual children with typical language development may have smaller vocabularies compared to monolinguals in one of their languages, or both, depending on the relative amount of exposure to each language (Thordardottir, 2011). Thus, when a bilingual child has poor vocabulary skills in one or both languages, it can be difficult to tell whether this should be attributed to exposure patterns, or whether it is due to a language disorder, or both.

In the current study, the Cross-Linguistic Lexical Task (CLT; Haman et al., 2015) is employed to investigate vocabulary comprehension and production in both languages. This task was developed in order to measure vocabulary comprehension and production in bilingual children in both languages with comparable assessment materials. The task is not translated, but each language version has been developed to be culturally and linguistically suitable for each specific language, and with comparable level of difficulty across the language versions. Currently, there are 29 language versions available. The task is explained in more detail in the Methods chapter (section 3.2.1).

1.3.2.3 Phonological working memory

Apart from having deficits in knowledge-based language skills (such as vocabulary), children with DLD often have deficits in language processing. Poor performance in non-word repetition, a task that entails repeating a series of phonological nonsense forms, is associated with DLD in monolinguals (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Topbaş et al., 2014). One advantage of NWR tasks is that they are much less affected by previous experience compared to for instance vocabulary tasks (Engel, Santos, & Gathercole, 2008; S. E. Gathercole, Willis, Emslie, & Baddeley, 1991). Although poor performance on NWR tasks has also has been established in bilinguals with DLD, many studies find that NWR tasks do not provide as accurate indications of presence or absence of DLD as for monolinguals (Boerma et al., 2015; Thordardottir & Brandeker, 2013). Typically developing bilinguals may be disadvantaged compared to monolinguals on certain NWR tasks, if the non-word items have a phonological setup that is similar to the phonotactical patterns of the majority language (Boerma et al., 2015; Thordardottir & Brandeker, 2013).

In the present study, four different NWR tasks are used to explore the performance across NWR items with different phonological setup and complexity. Three of the tasks were developed with multilingual assessment in mind. Two of them consist of NWR items that do not adhere to the phonological setup of a specific language, but contain items of 2–5 syllables with a simple consonant-vowel syllable structure (Chiat, 2015). Additionally, one task consists of items with a restricted phoneme inventory and few syllables (1–3) of mixed simple and complex setup (Abou Melhem et al., 2011). Finally, a language-specific task, matching the rules of Swedish lexical

phonology, is also used (Radeborg et al., 2006). The task contains items with 2–5 syllables, a wide phoneme inventory and mixed simple and complex syllable structures. More information about the non-word repetition tasks used in this study can be found in the Methods chapter (section 3.2.2).

1.4 Thesis structure

This thesis contains eight chapters. The current chapter (Chapter 1) establishes the rationale for the study, by introducing the topic of Developmental Language Disorder in the context of bilingualism, describing the motivation for selecting the methodology, as well as giving an introduction to the bilingual population that is investigated (Arabic-Swedish-speaking 4–7-year-olds). Chapter 2 specifies the aims and overarching research questions of the study. After this follows the Methods chapter (Chapter 3), which contains a detailed description of the participants in the cross-sectional (section 3.1.1) and the clinical (section 3.1.2) studies, the materials used (section 3.2), and the data collection procedure (section 3.3). The results of the cross-sectional study are reported in Chapter 4 (Vocabulary) and Chapter 5 (Phonological working memory). Both chapters commence with an overview of the previous literature, continue with the results of the current study, and conclude with a discussion of the results in light of the previous literature. In Chapter 6, the results from the clinical study are reported. The structure of the clinical chapter deviates slightly from the cross-sectional chapters. It starts with a review of the background literature (section 6.1). Next, the children in the clinical study are described in terms of their background characteristics (early language development, exposure patterns, etc., section 6.2.2). After that, the vocabulary (section 6.2.3) and NWR (section 6.2.4) performance of the children in the clinical study is compared to the performance of the children in the cross-sectional study. The chapter ends with a discussion of the results in light of the background literature (section 6.3), including a discussion of individual children (section 6.3.4) (the children with a DLD diagnosis and some of the low-performing children from the cross-sectional study). In Chapter 7, the results of the study are summarised (section 7.1), followed by a general discussion (section 7.2). Finally, the thesis is summarised in Swedish in Chapter 8.

2 Aims and research questions

The present study has three main aims. The first aim is to understand the vocabulary skills of 4–7-year old Arabic-Swedish-speaking children with typical language development growing up in Sweden. For this purpose, vocabulary comprehension and production is investigated in both the minority language (Arabic) and the majority language (Swedish). Developmental trajectories are established by comparing the performance across age. Part of this first aim is to explore to which extent external factors related to language exposure (reported daily exposure and length of exposure) and family background (SES) influence the children's vocabulary abilities in the majority and the minority language.

The second aim is to understand how non-word repetition (NWR) performance develops with age in Arabic-Swedish-speaking children. Part of this second aim is to examine how performance is affected by the properties of the NWR tasks or by children's language experience. Two major questions are being posed: (1) how does item complexity affect performance, and (2) how does language exposure and vocabulary affect performance on language-like items?

The third aim is to explore whether bilingual children with a diagnosis of Developmental Language Disorder (DLD) can be distinguished from children with typical language development, based on the performance on tasks tapping into language skills (vocabulary) as well as phonological working memory (NWR), and reports about early language development, language proficiency, communicative behaviour, and current language exposure and use.

Three research questions are posed in the present study:

- **RQ1**: How do vocabulary skills develop with age in both languages of 4–7-year-old Arabic-Swedish-speaking bilinguals, and which external factors influence that development?
- **RQ2**: How do 4–7-year-old Arabic-Swedish-speaking bilinguals perform on NWR tasks, and how is their performance affected by item length and complexity, language-likeness, and language exposure and vocabulary?
- **RQ3**: How do bilingual children with a DLD diagnosis perform on vocabulary and NWR tasks, what are the reported backgrounds, language abilities and communicative behaviours of bilingual children with a DLD diagnosis, and how can these be used to identify DLD in bilinguals?

3 Methods

In this chapter, the methods of the present study are reported. Section 3.1 describes the participants, section 3.2 the materials, and section 3.3 the data collection procedure. Descriptions of coding and analysis of the Cross-linguistic Lexical Tasks and the non-word repetition tasks will appear in each results chapter (Chapter 4 Vocabulary, Chapter 5 Phonological working memory, and Chapter 6 The clinical study).

3.1 Participants

In this section, the participants of the cross-sectional study (section 3.1.1), and the clinical study (section 3.1.2) are described. The participants were recruited by reaching out to children and their parents mainly in three larger cities in central Sweden. Prior to their participation, the parents of all children signed a form ensuring their informed consent, and filled in a questionnaire. The questionnaire provided information on the social and linguistic background of the children and their parents. All participants spoke Arabic and Swedish. The questionnaire was rather extensive, including 36 questions targeting language development, language exposure, language use in the family, parental education, occupation and language skills, concerns regarding atypical language development, as well as patterns of home activities such as book reading and storytelling.

For the participants with a DLD diagnosis, parents were given a questionnaire with similar questions as in the cross-sectional study. Some additional questions queried for how long the child had been in contact with an SLP, and who took the initiative for SLP assessment. Additionally, parents as well as school-teachers and SLPs were interviewed. Interviews with parents were conducted by Pascale Wehbe (SLP and master student in speech-language pathology) and Rima Haddad (PhD student of Linguistics), who also conducted the data collection in Arabic. Interviews with SLPs were conducted by the author, as well as Pascale Wehbe. Furthermore, the author conducted interviews with the school teachers of the children in the clinical study.

Each child received a code for the sake of anonymity. For all children participating in the cross-sectional study, the code consisted of five letters: BiAra (short for Bilingual Arabic), followed by a number for the child's age group, followed by two digits for unique identification. For the children

participating in the clinical study, all codes started with the same seven letters, BiAra-LI (where 'LI' is short for 'Language Impairment'), followed by two digits for unique identification.[15] Following these principles, BiAra4-19 is a 4-year-old participant of the cross-sectional study, and BiAraLI-02 is a participant in the clinical study.

3.1.1 The participants of the cross-sectional study

3.1.1.1 General information

The data for the cross-sectional study was collected between September 2017 and March 2019. Ninety-nine Arabic-Swedish-speaking bilinguals participated. The aim was to include 100 children, with an even distribution between age groups. Some of the participants in the cross-sectional study were reached through contacting preschools and schools in the greater Stockholm and Uppsala regions. The vast majority of children were recruited through contacts with associations and congregations who provided activities for Arabic-speaking children in the suitable age range. Some others were recruited through personal contacts of Arabic-speaking research assistants. The children participating in the study attended 53 different preschools and schools in this region.

In total, 116 children were recruited for the cross-sectional study. Of these, 17 children were excluded for various reasons. Two children were excluded after the first session, because it became evident that they had not yet reached their fourth birthday. One of these children was instead tested a year later (BiAra4-17). Six children were excluded after the first session, since their language skills were very rudimentary in that language, and they showed minimal signs of comprehension or participation. One child was excluded after the first session (in Arabic) after it was revealed that she went to a French-medium school and therefore did not speak Swedish. Eight children were excluded after finishing both sessions. Five of these children were excluded after discussions with the team, since they could not complete all tasks in one or both languages. Two children spoke Sudanese varieties of Arabic, which were too distant from the experimenters' Arabic varieties and the prepared language versions of the data collection materials. Finally, one child was excluded from the cross-sectional study as it turned out that she 1.5 years later (now having a DLD diagnosis) was recruited for the clinical study (previously BiAra5-20, now BiAraLI-08).

Table 3.1 provides an overview of the children in the cross-sectional study. The proportion of girls vs. boys was slightly different for each age group.

[15] In the present thesis, the term 'Developmental Language Disorder' (DLD) is used. However, at the start of the project, 'Language Impairment' (LI) was used to refer to the condition, which is reflected in the use of 'LI' in the codes of the participants in the clinical study.

However, the gender distribution for the group as a whole was even. All children were between 4;0 and 7;11 years old.

Table 3.1. Participants in the cross-sectional study: number, gender, mean age (years; months) and age range (years; months) per age group.

	4-year-olds	5-year-olds	6-year-olds	7-year-olds	Total
N	22	24	29	24	99
Girls/boys	12/10	9/15	12/17	16/8	49/50
Mean age	4;5	5;6	6;6	7;7	6;1
Age range	4;0–4;11	5;0–5;11	6;0–6;11	7;1–7;11	4;0–7;11

All 4-year-olds and 5-year-olds attended preschool. In the 6-year-old group, three children attended preschool, while 26 attended *förskoleklass*.[16] In the 7-year-old-group, 19 children were in the first grade of primary school. For five 7-year-olds, it was unknown which grade they were in. Since the majority of the children in the 7-year-old group attended first grade in school, most of them had started receiving formal reading and writing instruction.

Forty-four children were born in an Arabic-speaking country, and fifty-four were born in Sweden. One child, a 6-year-old, was born in an English-speaking country. The participants in the study spoke different varieties of Arabic, reflecting the Arabic-speaking population in Sweden. As can be seen in Table 3.2, the majority of the children spoke Syrian or Palestinian varieties, and a smaller sub-group spoke Iraqi varieties. Nine children spoke Lebanese varieties, while a handful spoke Egyptian varieties.[17] Many of these children were reported to be exposed to more than one variety of Arabic. The variety stated for each child in the table is the variety that the parents reported as the main variety the child spoke or was exposed to by next of kin. Some children were reported to speak a third language apart from Arabic and Swedish in the home. Most of the time, this third language was English; most of these children were reported to speak only a little English, while a couple had more extensive use of English. One 6-year-old was reported to also speak Modern Standard Arabic in addition to the vernacular spoken in the home. One 7-year-old spoke Kurdish (Sorani) in addition to Arabic. Children who were trilingual, or who had passive knowledge of an additional language were not excluded from the study, since multilingualism is a reality for many Arabic-Swedish-speaking families.

[16] *Förskoleklass* is a preparatory year between (non-mandatory) preschool and primary school.
[17] As previously mentioned, two children who spoke a Sudanese variety were excluded from the study, since their variety differed so much from that of the experimenter that they had trouble understanding each other during testing.

Table 3.2. Overview of the percentage of Arabic varieties spoken by the children in each age group (numbers in parentheses).

	4-year-olds	5-year-olds	6-year-olds	7-year-olds	Total
Egyptian	1.0% (1)	2.0% (2)	0.0% (0)	1.0% (1)	4.0% (4)
Iraqi	2.0% (2)	2.0% (2)	8.1% (8)	5.1% (5)	17.2% (17)
Lebanese	3.0% (3)	2.0% (2)	2.0% (2)	2.0% (2)	9.1% (9)
Palestinian	10.1% (10)	7.1% (7)	4.0% (4)	6.1% (6)	27.3% (27)
Syrian	6.1% (6)	11.1% (11)	15.3% (15)	10.1% (10)	42.4% (42)

3.1.1.2 Language development according to the parents

One of the inclusion criteria for the cross-sectional study was that no child should have atypical language development. In order to ensure that all children met this criterion, the parental questionnaire included questions targeting language development and parental concerns about language development. According to the parental questionnaire, none of the 99 children had been diagnosed with a language disorder.[18]

The parents of five children reported that they had consulted a speech and language pathologist (SLP) at least once about their child. General reasons for contact were minor problems with pronunciation such as lisping or difficulties with particular phonemes. The parents of one child (BiAra4-17) reported that the nurse at the Child Health Care Centre had concerns and referred the child to an SLP, but the SLP's assessment did not show that the child had a language disorder. The parents of BiAra6-05 stated that they had sought consultation from an SLP because the child did not speak much. They also reported that the child was a late talker, saying his first words in the home language at 35 months. This child was included in the study, since he did not have a diagnosis of DLD. Furthermore, he could complete all language tasks in both languages, and scored well above the mean (0.5 to 1 SD above) on both Arabic and Swedish vocabulary (CLTs).

Two 4-year-olds (BiAra4-17 and BiAra4-24) were reported to have had temporary hearing problems, and one of them (BiAra4-17, also mentioned above regarding consulting an SLP) had had many ear infections. Since these children's results were comparable to those of the other children in the same age group, none of them were excluded from the study.

[18] As was mentioned earlier, one child was excluded from the cross-sectional study after it became known that she had received a DLD diagnosis after the time of testing.

The parents of four children (BiAra5-25, BiAra6-05, BiAra6-26 and BiAra7-13) reported that their children had a late language development in Arabic. All four children apart from BiAra6-26 were reported to have a normal language development in Swedish, and they all had Swedish vocabulary scores well above the mean for their respective age groups. Therefore, they were all included in the study. The parents of six children reported that their child had a late language development in Swedish. However, all these children were reported to have a normal or early language development in Arabic, and they all scored within the normal range (and often well above) on the vocabulary test in at least one language.

One child, BiAra6-26, was reported to have a late language development in *both* languages. This child had a brother with a DLD diagnosis (BiAra-LI-01). Since he did not himself have a DLD diagnosis, he was still included in the study. However, this child will be discussed in later chapters, since he performed poorly on many language tasks.

The parents of 18 children thought that their child's language development was in some way different to that of their peers. More than half of the parents reported that it differed in a positive way, for instance that the child was faster in their language development or had better language skills compared to their peers. Some other comments concerned the child's bilingualism, such as switching language dominance, starting to speak one of the languages much later than the other, or them having a different language development compared to monolingual children but not compared to other bilingual children. The parents of BiAra6-27 reported that they thought that their child's language development was different because it took him almost 1.5 years to start to speak Swedish after arriving in the country. However, he had vocabulary scores that were above the mean in both languages.

Seven children were reported to have a relative who currently or previously had difficulties with speech, language or reading. Three children (BiAra5-12, BiAra6-02 and BiAra6-20) had siblings with developmental dyslexia/reading and writing difficulties. One child, BiAra6-19, had a parent with late language development. BiAra6-26 (previously discussed) had a sibling with diagnosed DLD. For one child (BiAra4-07), there was no information about which relative was affected or the nature of the speech, language or reading difficulties. Finally, with respect to BiAra6-14 it was reported that one of the child's parents spoke only Arabic. In the context of heredity for speech, language or reading problems, this comment is difficult to interpret, but it could possibly mean that the parent has difficulties learning Swedish because of inherent language learning problems. Heredity for speech, language or reading difficulties was not in itself a reason for exclusion, and all these children but two had vocabulary scores in the expected range. BiAra5-12 and BiAra6-20 had very low vocabulary scores (z-scores below –1.25) in one of their languages, and will be discussed in later chapters.

Parents were also asked about the appearance of the first word and the first multi-word utterance. For all children but four the first word appeared in Arabic before Swedish, and seven children had their first multi-word utterance in Swedish. The most common age for the first word to appear (in either language) was 12 months (N=27). The most common age for the first multi-word utterance to occur (in either language) was 24 months (N=31).[19] The parents of seven children did not report the age of the first word and the parents of nine children did not report the age of the first multi-word utterance.

3.1.1.3 Age of onset and rated language proficiency in Arabic and Swedish

The parents were asked, via the questionnaire, to state at which age the child started to receive regular input in Arabic and Swedish, respectively. They were also asked to rate the child's language proficiency in Arabic and Swedish comprehension and production. Table 3.3 provides a summary of age of onset for Arabic and Swedish respectively. All children except for one had regular Arabic input since birth. BiAra5-25 started hearing Arabic between age 1;0 and 2;0, and had regular input in Swedish since birth. For one child, BiAra7-05, age of onset for both Arabic and Swedish was unknown. While there was close to no variation in age of onset for Arabic, age of onset for Swedish varied considerably. A minority, 6%, had regular input in Swedish since birth, and for 53.5%, regular exposure to Swedish had started after age 3;0. Twenty children (20.2%) had had less than two years (24 months) of exposure to Swedish at the time of testing. These children were not excluded from the study, since they were immersed in the Swedish language via preschool or school, and they could complete all tasks in Swedish. Children with migrant and recent refugee backgrounds are common in present-day Sweden; therefore, it was decided that they should not be excluded from the study based on a specific age of onset or number of years in Sweden. As long as the children could complete the tasks in both languages, they were included in the study.

[19] The answers on these questions were highly skewed: a clear majority answered 12 and 24 months respectively, and the data did not resemble a normal distribution. Therefore, means and SDs are not reported.

Table 3.3. Percentage of children with an age of onset (AoO) of exposure for each year to Arabic and Swedish respectively (numbers in parentheses

	AoO in Arabic	AoO in Swedish
Birth to age 1;0	98.0% (97)	6.1% (6)
Age 1;0 to 2;0	1.0% (1)	24.2% (24)
Age 2;0 to 3;0	0.0% (0)	16.2% (16)
Age 3;0 to 4;0	0.0% (0)	26.3% (26)
Age 4;0 to 5;0	0.0% (0)	13.1% (13)
Age 5;0 to 6;0	0.0% (0)	11.1% (11)
Age 6;0 to 7;0	0.0% (0)	1.0% (1)
Age 7;0 to 8;0	0.0% (0)	1.0% (1)
Missing information	1.0% (1)	1.0% (1)
Total	100% (99)	100% (99)

Note. The reason for the span of age of onset to Swedish (e.g. 4;0–5;0) is that the parents were asked in the questionnaire to tick a box during which year of life the child started to receive regular exposure to each language.

Parents were asked to estimate their child's language proficiency with respect to comprehension and production of Arabic and Swedish on a five-point scale. The categories were *'very good', 'good', 'so-so', 'poor'* and *'very poor'*. As can be seen in Table 3.4, the vast majority of children were reported to have very good or good comprehension and production in Arabic as well as Swedish. Patterns were similar for comprehension and production in Swedish, but differed slightly between comprehension and production in Arabic, where parents generally estimated their children's comprehension to be a bit better than the production.

Table 3.4. Percentage of children for each category of reported proficiencies for comprehension and production of Arabic and Swedish (numbers in parentheses).

	Arabic		**Swedish**	
	Comp	**Prod**	**Comp**	**Prod**
Very good	69.7% (69)	54.5% (54)	45.5% (45)	40.4% (40)
Good	23.2% (23)	29.3% (29)	42.4% (42)	41.4% (41)
So-so	4.0% (4)	11.1% (11)	10.1% (10)	13.1% (13)
Poor	0.0% (0)	2.0% (2)	1.0% (1)	0.0% (0)
Very poor	1.0% (1)	0.0% (0)	0.0% (0)	2.0% (2)
Missing information	2.0% (2)	3.0% (3)	1.0% (1)	3.0% (3)
Total	100% (99)	100% (99)	100% (99)	100% (99)

3.1.1.4 Social and linguistic background of the parents

Socioeconomic status is a factor that is known to affect children's language development. Because of this, the questionnaire contained questions probing the level of education and the occupation of both parents. In the current study, parental education was used as a proxy for SES. In the questionnaire, questions were asked separately for each parent about their highest level of education. The answers were coded according to the United Nations' ISCED 2011 classification (UNESCO Institute for Statistics, 2012), where education levels are ranked in nine levels, ranging from 0 (early childhood education) to 8 (doctoral degree). Education levels varied considerably within the group, with 0 being the lowest and 8 the highest. The mean level of education was 4.46 (N=90, SD=1.02) for Parent 1 and 4.10 (N=91, SD=1.95) for Parent 2. Information regarding SES was missing for 17 parents in total; in six cases, information was missing for both parents.

The parents were also asked about their birthplace, in which country they grew up, and for how long they had been in Sweden. As can be seen in Table 3.5, the vast majority of parents had been born and raised in an Arabic-speaking country. A handful had grown up in Sweden, and only one was born in Sweden. Information was missing for either place of birth or growing up for nine parents. For two participants, no information was available for either parent 1 or parent 2.

Table 3.5. Overview of percentage of parents according to country of birth and where they grew up (numbers in parentheses).

	Parent 1	**Parent 2**
Born and raised in an Arabic-speaking country	90.9% (90)	82.8% (82)
Born and raised in Sweden	0.0% (0)	1.0% (1)
Born in an Arabic-speaking country, raised in Sweden	2.0% (2)	6.1% (6)
Born in an Arabic-speaking country, raised in other country	1.0% (1)	0.0% (0)
Born in other country, raised in an Arabic-speaking country	0.0% (0)	3.0% (3)
Missing information	6.1% (6)	7.1% (7)
Total	100% (99)	100% (99)

The length of stay in Sweden (for all parents but the one who was born and raised in Sweden) varied substantially, with 10 months being the shortest and 31 years being the longest. The mean length of stay for Parent 1 was 9.34 years (N=93 SD=7.19), and 10.10 years for Parent 2 (N=91, SD=8.13). There was

missing information with respect to length of stay for 13 parents in total; in three cases, information was missing for both parents.

As can be seen in Table 3.6, the majority of the parents had Arabic as their first language (L1). One parent had Turkish and Kurdish (Sorani) as their first language. Both parents of one child mentioned that their first language was not Arabic, but did not specify which language it was instead.[20] Information was missing for 12 parents in total; in three cases, information was missing for both parents.

Table 3.6. Distribution of parents by first language (numbers in parentheses).

	Parent 1	**Parent 2**
Arabic	94.9% (94)	89.9% (89)
Other language	2.0% (2)	1.0% (1)
Missing information	3.0% (3)	9.1% (9)
Total	100% (99)	100% (99)

3.1.1.5 Language use in the family

There were multiple questions in the questionnaire targeting language use in the family. These were: which language(s) the parents spoke to the child, which language(s) the children spoke to the parents, which language(s) the parents spoke to each other, and which language(s) the child spoke with their siblings.

Table 3.7 shows the parents' reported language use with their child. In five cases, information regarding language spoken to the child was missing for one parent. In these cases, information was used for the one parent for which information was available, since it can be assumed that they were the primary caregiver and their language use reflected that in the home overall. Most children grew up in households where both parents reported that they spoke only or mostly Arabic. In eight households, one parent spoke predominately Arabic, and the other parent spoke equal amounts of Arabic and Swedish. In another household, one parent spoke mostly Arabic and the other parent spoke mostly Swedish. Two households reported that one parent spoke other language combinations to their child (BiAra4-04: Arabic and German; BiAra7-11: Kurdish (Sorani) and Swedish). In both cases, the other parent spoke only or mostly Arabic to the child. For one child, information about which language the parents spoke to the child was missing. No households reported that both parents spoke predominately Swedish to their children.

[20] Based on the information that they were from Syria, it is likely that this other language was Syriac/Aramaic.

Table 3.7. Distribution of families by reported parental language use with the child (numbers in parentheses).

Both parents mainly Arabic	79.8% (79)
1 parent mainly Arabic, 1 parent 50/50	8.1% (8)
1 parent mainly Arabic, 1 parent mainly Swedish	1.0% (1)
Both 50/50	8.1% (8)
1 parent mainly Swedish, 1 parent 50/50	0.0% (0)
Both parents mainly Swedish	0.0% (0)
Other	2.0% (2)
Missing information	1.0% (1)
Total	100% (99)

Table 3.8 contains information about which language(s) the parents spoke to each other. The vast majority reported that they spoke (almost exclusively) Arabic, while the parents of four children reported that they spoke both Arabic and Swedish to each other. The parents of two children did not report any information here, in one case because one parent was deceased, and in the other case because the parents were separated and reported that they did not speak to each other. In three cases, it was unknown which language(s) the parents spoke to each other. No households reported that the parents spoke predominately Swedish with each other.

Table 3.8. Parents' language use with each other (numbers in parentheses).

Only Arabic	Arabic and Swedish	Only Swedish	Other	Missing info	Total
90.9% (90)	4.0% (4)	0.0% (0)	2.0% (2)	3.0% (3)	100% (99)

Table 3.9 shows the distribution of the children's language use with their parents. Similarly as for the parents, the most common pattern for the children was to speak predominately Arabic to both parents, or predominately Arabic to one parent and equal amounts of Arabic and Swedish to the other parent. Thirteen children spoke equal amounts of Arabic and Swedish to both parents. Eight children were reported to speak predominately Swedish to both parents (which was a different pattern compared to the parents, where no parent reported that they spoke predominately Swedish to their child). One child spoke Kurdish (Sorani) and Swedish to one parent, and Arabic to the other parent. One child was reported to have a deceased father; therefore, the communicative pattern reported here is based only on the mothers' answer.

Table 3.9. Children's language use with their parents (numbers in parentheses).

To both parents mainly Arabic	67.7% (67)
To 1 parent mainly Arabic, to 1 parent 50/50	8.1% (8)
To 1 parent mainly Arabic, to 1 parent mainly Swedish	1.0% (1)
To both parents 50/50	13.1% (13)
To 1 parent mainly Swedish, to 1 parent 50/50	0.0% (0)
To both parents mainly Swedish	8.1% (8)
Other	1.0% (1)
Missing information	1.0% (1)
Total	100% (99)

Even though most children spoke predominately Arabic to their parents, another pattern was evident for language use between the siblings. As can be seen in Table 3.10, about a third of the children (36.4%) spoke predominately Arabic to their siblings, and another third spoke equal amounts of Arabic and Swedish respectively. About one fifth of the children (18.2%) were reported to speak mostly Swedish to their siblings. Four children did not have any siblings, and for five children information was missing for language use with siblings.

Table 3.10. Children's language use with their sibling(s) (numbers in parentheses).

No sibling	Mostly Arabic	Equal amounts	Mostly Swedish	Missing information	Total
4.0% (4)	36.4% (36)	36.4% (36)	18.2% (18)	5.1% (5)	100% (99)

3.1.1.6 Estimated exposure to Arabic and Swedish

Quality and quantity of input have emerged in the literature as being highly influential factors in bilingual language development. In order to investigate the amount of relative exposure the children had to Arabic and Swedish during the day, the parents were asked to estimate the proportion of daily input in each language. The question posed was 'How often does your child hear both languages in their everyday life? Make a mark on the scale'. The scale had seven categories, reflecting different levels of relative exposure to each language. The levels were '1: Swedish 5%, Arabic 95%', '2: Swedish 20%, Arabic 80%', '3: Swedish 40%, Arabic 60%', '4: Swedish 50%, Arabic 50%', '5: Swedish 60%, Arabic 40%', '6: Swedish 80%, Arabic 20%', '7: Swedish 95%, Arabic 5%'. There was also one additional eighth category, marked 'other'. In the 'other' category it was possible for parents to write other exposure patterns, for instance if the child received exposure to an additional

language besides Arabic and Swedish. Table 3.11 shows the distribution of exposure to Arabic and Swedish by age group.

According to the parents' estimates, about a third of the children (36.4%) received similar amounts of exposure (50/50) in each language. Another 18.2% children received 60% exposure to Arabic and 40% exposure to Swedish, and 22.2% children received 40% exposure to Arabic and 60% exposure to Swedish. This means that most children had relatively balanced exposure to each language during the day. Thirteen children (13.1%), mainly in the two youngest age groups, had 80% or more daily exposure to Arabic, and only three children were reported to have 80% or more daily exposure to Swedish. The parents of two children reported that their child had other patterns of exposure, BiAra7-11, who also received input in Kurdish (Sorani), and BiAra7-17, who received input in English apart from Arabic and Swedish.

Table 3.11. Daily language exposure to Swedish and Arabic (numbers in parentheses).

Age group	Swe 5% Ara 95%	Swe 20% Ara 80%	Swe 40% Ara 60%	Swe 50% Ara 50%	Swe 60% Ara 40%	Swe 80% Ara 20%	Swe 95% Ara 5%	Other	Missing information
4	1.0% (1)	4.0% (4)	5.1% (5)	6.1% (6)	5.1% (5)	1.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)
5	0.0% (0)	5.1% (5)	5.1% (5)	7.1% (7)	5.1% (5)	2.0% (2)	0.0% (0)	0.0% (0)	0.0% (0)
6	0.0% (0)	1.0% (1)	5.1% (5)	13.1% (13)	8.1% (8)	2.0% (2)	0.0% (0)	0.0% (0)	0.0% (0)
7	0.0% (0)	2.0% (2)	3.0% (3)	10.1% (10)	4.0% (4)	1.0% (1)	1.0% (1)	2.0% (2)	1.0% (1)
Total	1.0% (1)	12.1% (12)	18.2% (18)	36.4% (36)	22.2% (22)	6.1% (6)	1.0% (1)	2.0% (2)	1.0% (1)

3.1.1.7 Mother tongue instruction

Mother Tongue Instruction (MTI) is available to children in Sweden who speak a language other than Swedish in the home. For Arabic-speaking children, there are also some private initiatives that provide MTI via congregations (such as orthodox churches and mosques), or associations. Whether children attend MTI or not is likely to affect their language skills in the home language. The parental questionnaire included questions regarding whether the child attended MTI, and if so: how it was organised, for how many hours a week, as well as whether the instruction took place with other children or not.

As shown in Table 3.12, a total of 64 children (65%) were reported to attend MTI. The amount of hours per week ranged from 0.67 to eight, averaging at 2.18 hours (N=55, SD=1.84).[21] The majority of the children who attended MTI did so together with other children (N=45).[22] Among the four- and 5-year-olds there were fewer children who were reported to attend MTI. However, among those in the two younger age groups who did attend MTI, it was more common that the instruction was organised via a private initiative than in the two older age groups. One likely reason for this is that the municipality is not required to organise MTI instruction in preschool as it is in school. Also, some municipalities who previously offered MTI instruction to children attending preschool have revoked this (Lindström, 2016), and some parents have turned to private initiatives instead. In the two older age groups, only seven of the children were reported not to attend MTI, and the majority of those who did attend followed the instruction provided by the municipality.

[21] The parents of nine children did not answer the question regarding how many hours their child attended MTI. Eight out of these children were 6- and 7-year-olds, attending only MTI provided by the municipality. It is therefore likely that the amount of instruction per week was around one hour.

[22] Out of the 64 children whose parents reported that they attended MTI, the parents of nine children did not provide information regarding whether it was together with other children or not. However, most of them attended MTI only via the municipality; therefore, it is likely that the instruction took place in a group of other children.

Table 3.12. MTI instruction by type and age group (numbers in parentheses).

Age group	Yes, via municipality	Yes, private initiative	Yes, both municipality and private	No	Missing information
4	2.0% (2)	7.1% (7)	1.0% (1)	10.1% (10)	2.0% (2)
5	1.0% (1)	8.1% (8)	0.0% (0)	14.1% (14)	1.0% (1)
6	19.2% (19)	3.0% (3)	1.0% (1)	6.1% (6)	0.0% (0)
7	16.2% (16)	3.0% (3)	3.0% (3)	1.0% (1)	1.0% (1)
Total	38.4% (38)	21.2% (22)	5.1% (5)	31.3% (31)	4.0% (4)

3.1.1.8 Participants of the cross-sectional study: summary

This section contains a brief summary of the information provided by the parents of the participants in the cross-sectional study with respect to general characteristics, language development and proficiency, language use in the family, exposure patterns and mother tongue instruction.

Almost half of the children (44%) were born in an Arabic-speaking country. Some children (seven) were reported to also know some English apart from Arabic and Swedish, and one child also knew some Modern Standard Arabic (MSA), in addition to the vernacular spoken in the home. One child spoke Kurdish (Sorani) in the home apart from Arabic ad Swedish. Most children spoke Levantine varieties of Arabic (Syrian: 42, Palestinian: 27, and Lebanese: 9), seventeen spoke Iraqi varieties, and four children spoke Egyptian varieties. While there was close to no variation in age of onset for Arabic (being at birth for all children apart from one child), age of onset to Swedish varied considerably, ranging from birth to age seven. The vast majority of the parents were born and had grown up in an Arabic-speaking country; years of residence ranged from 10 months to 31 years. Only one parent was born in Sweden. Eight parents had been born in an Arabic-speaking country but grown up in Sweden. The vast majority of the parents had Arabic as their L1, and in most families both parents spoke predominately Arabic to their children (79) and to each other (90). No parent spoke predominately Swedish to their child, but eight children spoke predominately Swedish to their parents. About two-thirds of the children (67) spoke predominately Arabic to both parents, but only about one third (36) children spoke predominately Arabic to their siblings. It was as common for children to use both Arabic and Swedish in about equal amounts, and nearly 20% used mostly Swedish when communicating with their siblings. About a third (36) of the children were reported to receive equal amounts of daily input in both

languages, while around 40% heard slightly more Arabic (N=18, Arabic 60%) and more Swedish (N=22, Swedish 60%) respectively. There were almost twice as many children who were reported to have 80% or more exposure to Arabic (13) than who had 80% or more Swedish (7%). Finally, about two thirds (64) of the whole group were reported to attend MTI. Almost all children in the two older age groups attended MTI. Children in the two younger age groups who attended MTI most often did so via some private initiative, while the older children most often attended only the instruction provided by the municipality (i.e. in school).

3.1.2 The participants of the clinical study

3.1.2.1 General information

Recruitment of children with a DLD diagnosis was conducted via phone calls, emails and social media by the author of the present study and SLP master student Pascale Wehbe. Around 30 SLPs working in both public health care and private SLP clinics as well as specialised preschool and school units for children with DLD in the Stockholm, Uppsala and Västerås regions were contacted. During initial contact, the SLP was given a description of the research project and they were asked whether they knew about any children who fit the inclusion criteria. The inclusion criteria for participating in the study were the following:

- Age 4;0–7;11
- Being regularly exposed to Arabic and Swedish
- The Arabic variety the children were exposed to should match one of the varieties included in the cross-sectional study (i.e. Levantine, Iraqi or Egyptian)
- Having a DLD diagnosis as confirmed by an SLP. The aim was to include children who had comprehension difficulties as well as production difficulties (in Swedish 'generell språkstörning', *general language disorder*). However, children who were classified as having exclusively comprehension difficulties (in Swedish 'impressiv språkstörning', *impressive language disorder*) or production difficulties (in Swedish 'expressiv språkstörning', *expressive language disorder*) could also participate. Furthermore, children who currently had an unspecified diagnosis (in Swedish 'ospecificerad språkstörning', *unspecified language disorder*), but where the SLP suspected DLD could also participate. However, children with exclusively phonological deficits (in

Swedish 'fonologisk språkstörning', *phonological disorder*) were not eligible for participation.[23]

In accordance with Bishop et al. (2017, p. 1071), children who had a known biomedical condition associated with language difficulties (for instance Down syndrome), a diagnosis within the autism spectrum, or intellectual disability (ID) were not eligible to participate. However, having a diagnosis of attention deficit hyperactivity disorder (ADHD) was not an exclusion criteria, considering the high rates of co-occurrence with DLD (Bishop et al., 2017, p. 1072). Furthermore, an unconfirmed suspicion (by the parents, SLP or teacher) of for instance autism or ID was not a reason for exclusion.

The SLPs were given information materials to give to the parents of potential participants, including an information letter and a consent form, both available in Arabic and Swedish. The SLP presented the parents with the information materials, including a link to a video containing the same information as in the letter, intended as a way of familiarising the parents with the experimenters and telling them about the study in an easy-to-understand way. In total, seventeen children and their parents were asked by their SLP to participate in the study. Of these children, four turned down participation and one was excluded because they spoke a different Arabic dialect (Sudanese). One child was also turned down because they lived too far away for data collection to be feasible. Eleven children participated in the end, seven boys and four girls. The recruitment and data collection for the current study overlapped with Pascale Wehbe's master thesis that had a narrower age range (5;0–6;11) (Wehbe, 2020). As a result, the children in the clinical study had a narrower age range (5;0–7;3) than the children in the cross-sectional study (4;0–7;11), but the mean age was similar in both groups (mean age cross-sectional group: 6;1, mean age clinical group: 6;2). Seven children were recruited through the same SLP clinic, two children via an SLP working in the school health care service in a municipality, one child was recruited through a second SLP clinic, and another child was recruited through a 'språkförskola', *language preschool*, a specialised preschool unit for children with severe DLD. All children lived in the *Mälardalen* region, including the cities Stockholm, Uppsala and Västerås as well as surrounding municipalities. The data collection was conducted between January and September 2019.

The parents were given a questionnaire that was nearly identical to the one used in the cross-sectional study. Thus, it contained questions targeting language development, language exposure, language use in the family, parental education, occupation and language skills, as well as patterns of home activities such as book reading and storytelling. Questions querying the nature

[23] In the Anglo-Saxon terminology, this condition is called speech sound disorder, and is generally not perceived as a *language* disorder (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 consortium, 2017, p. 1073).

and length of SLP contact as well as possible language difficulties in the family were also included. In addition to filling in the questionnaire, parents were interviewed by either Pascale Wehbe or Rima Haddad. The questions asked during the interview concerned the same topics as in the questionnaire, but provided more in-depth information with respect to for instance how the parents viewed their child's language development over time, their attitudes and beliefs regarding language development and bilingualism, and whether they were concerned. The author of the present study interviewed the teachers when seeing the child in (pre)school for the Swedish testing session. The teacher interview included questions about the child's language skills, their communicative and social behaviour, whether they could follow instructions, and how they behaved during linguistic awareness and book reading activities. In addition, interviews with the SLPs were conducted by the author of the present study and Pascale Wehbe. The SLPs were asked how the child had been assessed (in which language(s), and which materials were used), age at referral, possible therapy and the child's development over time, the parents' attitudes towards therapy, current diagnosis, and what they consider as the most striking or problematic thing about the child's language. The questions for the interviews with parents, SLPs and teachers were developed by the BiLI-TAS team, and first used during the clinical study conducted on Turkish-speaking children in the BiTur-project, as described in Öztekin (2019). The original interview templates were slightly modified in order to suit the current study. Table 3.13 provides information about the children's age at testing, characteristics of the (pre)school (type and grade), diagnosis as well as age at initial SLP contact and possible language therapy.

Table 3.13. Information about the children's age at testing, (pre)school type, diagnosis, and SLP referral and therapy.

Child	Age	(Pre-) School	Diagnosis	SLP referral and therapy
BiAra LI-01	6;8	Förskole-klass	Unspec. LD, the SLP suspects gen. LD	Age 4;3, regular therapy
BiAra LI-02	6;1	Pre-school	gen. LD	Age 3, mainly indirect therapy
BiAra LI-03	5;7	Pre-school	gen. LD	Age 2;9, sporadic therapy
BiAra LI-04	6;0	Pre-school	Expr. LD, previously gen. LD	Age 4;6, regular therapy

BiAra LI-05	7;3	School, 1st grade	gen. LD	Age 6;2, no therapy
BiAra LI-06	5;4	Pre-school	gen. LD	Age 4;4, mainly indirect therapy
BiAra LI-07	6;1	Pre-school	gen. LD	Age 4;1, regular therapy until recently
BiAra LI-08	7;1	Förskole-klass	gen. LD	Age 6;10, no therapy at the clinic but support via SLP in school
BiAra LI-09	6;7	Förskole-klass	gen. LD	Age 3, received therapy at the clinic earlier, now has support via SLP in school
BiAra LI-10	5;0	Språk-förskola	gen. LD	Age 2;3, received therapy at the clinic earlier, now has support via SLP in preschool
BiAra LI-11	6;4	Förskole-klass	gen. LD	1st referral at age 3, 2nd at 5;1, received therapy at the clinic earlier

Note. 'Förskoleklass' is a preparatory year between preschool and first grade. 'Språkförskola' is a specialised preschool unit for children with severe DLD. Unspec. LD=unspecified language disorder, gen. LD=general language disorder, expr. LD=expressive language disorder. *Indirect therapy* means that the SLP advises the parents on how to interact with their child in order to strengthen their language skills. The meaning of *regular therapy* may vary, depending on for instance differences in access to speech and language therapy in different regions of the national health service.

3.1.2.2 Language background and input

The majority of the children in the clinical study were born in Sweden, but four children were born in an Arabic-speaking country (BiAraLI-01, BiAraLI-04, BiAraLI-08 and BiAraLI-11). Table 3.14 shows the Arabic variety and the age of onset for Arabic and Swedish for each participant in the clinical study. The most common Arabic variety in the clinical group was Iraqi, spoken by six children (55%). This differed from the cross-sectional group, where 17.2% spoke an Iraqi variety. Syrian was the second most common variety, spoken by four children (36%). One child spoke a Palestinian variety. All children were exposed to Arabic from birth, but age of onset for Swedish varied from the second year of life to the sixth year of life. Six children (55%) started hearing Swedish before age three (BiAraLI-02, BiAraLI-03, BiAraLI-06,

BiAraLI-07, BiAraLI-09, BiAraLI-10), and five children (45%) after the age of three (BiAraLI-01, BiAraLI-04, BiAraLI-05, BiAraLI-08, BiAraLI-11).

Table 3.14. Arabic variety and age of onset for Arabic and Swedish.

			Age of onset	
Child	**Age**	**Arabic variety**	**Arabic**	**Swedish**
BiAraLI-01	6;8	Iraqi	at birth	5;0–6;0
BiAraLI-02	6;1	Iraqi	at birth	1;0–2;0
BiAraLI-03	5;7	Syrian	at birth	1;0–2;0
BiAraLI-04	6;0	Syrian	at birth	4;0–5;0
BiAraLI-05	7;3	Iraqi	at birth	4;0–5;0
BiAraLI-06	5;4	Syrian	at birth	1;0–2;0
BiAraLI-07	6;1	Iraqi	at birth	1;0–2;0
BiAraLI-08	7;1	Syrian	at birth	3;0–4;0
BiAraLI-09	6;7	Iraqi	at birth	1;0–2;0
BiAraLI-10	5;0	Iraqi	at birth	1;0–2;0
BiAraLI-11	6;4	Palestinian	at birth	4;0–5;0

Note. The reason for the span of age of onset to Swedish (e.g. 4;0–5;0) is that the parents were asked in the questionnaire to tick a box during which year of life the child started to receive regular exposure to each language.

Table 3.15 shows patterns of language input and use for each individual child in the clinical study. As can be seen in the table, reported daily language exposure for the two languages were even. Six children were reported to hear equal amounts of Arabic and Swedish throughout the day. Two children (BiAraLI-01, BiAraLI-05) had slightly more Swedish (60%) than Arabic (40%), and two children (BiAraLI-07, BiAraLI-11) had slightly more Arabic (60%) than Swedish (40%). Only one child (BiAraLI-02) had substantially uneven exposure rates, with predominately Arabic (80%) and very little Swedish (20%).

All parents reported that Arabic was their first language, and it was also the main language of communication between the parents. Although this information is missing for three children (BiAraLI-01, BiAraLI-02 and BiAraLI-03), it is likely that the parents spoke predominately Arabic with each other, since they all had Arabic as their first language and they spoke mostly Arabic with their children. Arabic was the language in which parents communicated with their children most often. Only the parents of one child (BiAraLI-10) reported that they used equal amounts of Arabic and Swedish when communicating with their child. When communicating with their parents most children used predominately Arabic, although three children (BiAraLI-03, BiAraLI-06 and BiAraLI-10) were reported to use equal

amounts of Arabic and Swedish. When communicating with their siblings, most children used both Arabic and Swedish (55%). Three children (BiAraLI-02, BiAraLI-08 and BiAraLI-11) were reported to use mostly Arabic and only one child (BiAraLI-04) used mostly Swedish with their siblings. Thus, the same pattern emerged as in the cross-sectional sample, in that parents were more inclined to use the home language with each other and with their children. Although most children used predominately Arabic with their parents, use of Swedish was prominent between the children and their siblings.

Table 3.15. Patterns of input and language use.

Child	Daily language exposure	Parent to parent	Parents to child	Child to parents	Child with sibling(s)
BiAra LI-01	Swe 60%, Ara40%	---	pred. Ara	pred. Ara	both Ara and Swe
BiAra LI-02	Swe 20%, Ara 80%	---	pred. Ara	pred. Ara	mostly Ara
BiAra LI-03	Swe 50%, Ara 50%	---	pred. Ara	50% Ara, 50% Swe	both Ara and Swe
BiAra LI-04	Swe 50%, Ara 50%	Ara	pred. Ara	pred. Ara	mostly Swe
BiAra LI-05	Swe 60%, Ara 40%	single parent	pred. Ara	pred. Ara	both Ara and Swe
BiAra LI-06	Swe 50%, Ara 50%	Ara	pred. Ara	50% Ara, 50% Swe	both Ara and Swe
BiAra LI-07	Swe 40%, Ara 60%	Ara	pred. Ara	pred. Ara	both Ara and Swe
BiAra LI-08	Swe 50%, Ara 50%	Ara	pred. Ara	pred. Ara	mostly Ara
BiAra LI-09	Swe 50%, Ara 50%	Ara	pred. Ara	pred. Ara	no siblings
BiAra LI-10	Swe 50%, Ara 50%	Ara	50% Ara, 50% Swe	50% Ara, 50% Swe	both Ara and Swe
BiAra LI-11	Swe 40%, Ara 60%	Ara	pred. Ara	pred. Ara	mostly Ara

Note. '---' indicates missing information. Pred. = predominately.

Furthermore, the questionnaire also contained questions about possible language and literacy problems in the family. According to parental reports, six children (55%) in the clinical study had a relative with language and/or literacy difficulties: BiAraLI-01, BiAraLI-03, BiAraLI-04, BiAraLI-05, BiAraLI-06 and BiAraLI-08. This proportion was very different from the cross-sectional study, where only 7% were reported to have relatives with language or literacy difficulties. No child was reported to have had hearing problems or ear infections.

3.1.2.3 Parents' evaluation of children's language abilities

The parents were asked to rate their children's general language development as well as their current language proficiency in Arabic and Swedish. As shown in Table 3.16, all children were reported to have a late language development in Arabic, apart from BiAraLI-04 and BiAraLI-08 who were both reported to have a normal language development. The ratings for Swedish were more varied. Although the majority of the children (N=6, 55%) were reported to have a late language development also in Swedish, the parents of three children reported a normal development, and one child was even reported to have an early development in Swedish. Although general language development was rated to be late in at least one language for the majority of children (82%), current language proficiency was generally not described as poor. For comprehension in Arabic, proficiency was reported to be good or very good for nine children (82%), very poor for one child, and so-so for one child. For Swedish, comprehension was rated as very good or good for seven children (64%), and so-so for four children. For production in Arabic, proficiency was reported to be good for nine children (82%), and so-so for two children. For Swedish, production was rated as very good or good for six children (55%), and so-so for five children. Current language proficiency was generally rated as better in Arabic than Swedish.

Table 3.16. Parents' evaluations of their children's overall language development and current proficiency in Arabic and Swedish.

	Language development		Language proficiency			
			Comprehension		Production	
Child	**Arabic**	**Swedish**	**Arabic**	**Swedish**	**Arabic**	**Swedish**
BiAra LI-01	late	late	very poor	so-so	so-so	so-so
BiAra LI-02	late	normal	very good	good	good	good
BiAra LI-03	late	late	good	good	good	good
BiAra LI-04	normal	early	good	good	good	good
BiAra LI-05	late	---	good	good	good	good
BiAra LI-06	late	late	very good	very good	good	good
BiAra LI-07	late	late	so-so	so-so	good	so-so
BiAra LI-08	normal	normal	good	so-so	good	so-so
BiAra LI-09	late	late	very good	good	good	good
BiAra LI-10	late	late	good	good	so-so	so-so
BiAra LI-11	late	normal	very good	so-so	good	so-so

Note. '---' indicates missing information.

Somewhat surprisingly, the reported proficiency levels of the clinical group were similar to those of the cross-sectional group in both comprehension and production in Arabic and Swedish. However, there were notable discrepancies between the answers that parents provided in the questionnaire and information that emerged during the parental interviews. During the interviews, the children's language skills were generally described as being poorer compared to what was reported about the same topic in the questionnaire.

During the parental interview, parents were asked to describe the nature of their child's language difficulties, and what they thought were possible reasons for these difficulties. The answers to these questions are presented in Table 3.17.

Table 3.17. Parents' descriptions concerning their child's language difficulties and possible reasons for these.

Child	Language difficulties and behaviour
BiAra LI-01	He does not interact with other children. He listens but does not speak much, and rarely poses any questions. He has problems with comprehension as well as making himself understood.
BiAra LI-02	She is late in her language development compared to other Arabic-Swedish-speaking bilinguals. She has difficulties understanding what is said to her. She used to have large difficulties expressing herself, having to point and gesture in order to make herself understood, but it has improved.
BiAra LI-03	He has weak language comprehension. Sometimes when the parents do not understand him, he can get angry. His Arabic is not as good as other Arabic-Swedish-speaking bilinguals his age. He was late in his language development from the start, but the parents think that his language skills were further negatively impacted by the fact that he went to a preschool where they did not develop the children's language skills.
BiAra LI-04	He is very talkative and likes to interact with other children as well as adults. His language skills are good in both languages; the only thing that he has difficulties with is the pronunciation of some speech sounds.
BiAra LI-05	The school advised to see an SLP for language assessment. He does not have any problems in Arabic, but he quite recently started to learn Swedish and he does not get enough Swedish input.
BiAra LI-06	He mixes the languages, and uses Swedish words when speaking Arabic. He understands what the parents say, and the parents understand him. He had difficulties expressing himself sometimes when he was younger, but this has improved a lot. His language development was slow in the beginning, but now he learns fast. Perhaps there is heredity involved, as his little brother has similar difficulties.
BiAra LI-07	He talked very little in the beginning. He was even late in developing Arabic. He has a weak vocabulary, which leads to difficulties with finding words. He also has difficulties with pronunciation. The parents do not know if there is a

	specific reason for his difficulties. The parents consulted a psychologist, but their opinion was that the child did not have any additional difficulties.
BiAra LI-08	She does not speak much, neither in Arabic nor in Swedish. She has difficulties with both language comprehension and production. She only tries to find Arabic-speaking friends at school. Some relatives say that the child does not answer when she is spoken to in Arabic.
BiAra LI-09	He was late in his language development, but now everything is ok in Arabic. He is very shy and does not like to speak much. He does not have a large vocabulary in Swedish.
BiAra LI-10	She is very social and likes to interact with others. She is late in her language development compared to other Arabic-Swedish-speaking children. Her language comprehension is poor compared to her peers. Sometimes, the parents have difficulties understanding her. Her difficulties could be due to heredity, as her older sister had similar difficulties.
BiAra LI-11	She is very shy. She is aware of her difficulties and avoids saying things she knows will pose a problem for her. She was late in her early language development, but she has improved a lot. She has good comprehension. When the parents do not understand her, she uses gestures or objects to show what she means. Now the biggest concern is her pronunciation difficulties. Some say that it could be hereditary, but the parents do not know of any relatives with similar problems.

3.1.2.4 SLPs assessment and evaluation of children's language abilities

The interviews with the SLPs included questions about how the child had been assessed and which materials were used. The SLPs were also asked to describe the children with respect to their language profiles and communicative behaviour. This information was provided by the SLP who had conducted the initial assessment, or obtained from the child's medical record at the SLP clinic. For three children, information about the assessment and diagnosis was available through written statements from the SLP. Most children had been assessed in both languages in some way, either by a bilingual Arabic-Swedish-speaking SLP (N=7), or by making use of an interpreter (N=2). Two children had been assessed only in Swedish. The SLPs mention a range of materials that were used during assessment, which are listed in the following.

- Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983). A picture naming test originally constructed for assessing expressive

vocabulary in adults with aphasia. Swedish translation by Tallberg (2005); Swedish reference data from 152 children age 6–15 (age groups 6, 9, 12 and 15) (Brusewitz & Tallberg, 2010).

- Bussagan [Bus Story Test]. Original version by Renfrew (1969), translated into Swedish and standardised by Svensson and Tuominen-Eriksson (2000). Assessment of retelling a story told with picture aid, intended for children age 3:9–6:8. Swedish reference data from 110 children.
- Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4; Semel, Wiig, & Secord, 2013). A test battery used to assess both comprehension and production of different language domains (vocabulary, morphosyntax, comprehension of spoken paragraphs), as well as working memory, sentence repetition and word finding abilities. 'Pan-Scandinavian' norms (i.e. a joint norming sample for Danish-, Norwegian- and Swedish-speaking children) available from 600 children age 5;0-12;11.
- The FAS word fluency task. Swedish standardisation and reference data from 130 Swedish speaking children age 6–15 (age groups 6, 9, 12 and 15) available in an unpublished master's thesis in speech-language pathology (Carlsson, 2009).
- Fonem- och benämningstest vid flerspråkighet [Phoneme and naming test for bilinguals]. Clinic internal material, not published.
- Nya Fonemtestet [the Phoneme test] (second edition) (Hellqvist, 2013). Assessment of expressive phonology and disorders in preschool children (from age 3). No reference data.
- Grammatiktest för barn (GRAMBA) [Grammar test for children] (Hansson & Nettelbladt, 2010). Test of expressive morphosyntax in Swedish-speaking children age 3–6 with norms from 661 children.
- Grammatisk undersökning av svenska som andraspråk (GrUS) [Grammatical assessment of Swedish as a second language] (Salameh, 2015). A material for assessing expressive morphosyntax in L2 learners of Swedish primarily in the preschool and early school ages. Criterion based assessment of elicited morphosyntactic structures.
- Multilingual Assessment Instrument of Narratives (MAIN; Gagarina et al., 2012). Assessing comprehension and production of narratives for children age 3–10.
- Non-word repetition (NWR), specific task not stated.
- Nya NELLI – Neurolingvistisk undersökningsmodell för språkstörda barn [Neurolinguistic assessment model for language impaired children] (Holmberg & Sahlén, 2000). A test battery with many subtasks, for example NWR, sentence repetition, picture aided telling, retelling and more. The criteria for evaluating the results are described separately for each subtask, focusing primarily on qualitative assessment. Intended for children age 4 and up.

- Ordracet [The word race] (Eklund, 1996). Assessment of rapid automatic naming, 80 black-and-white pictures. Available in two formats: electronic and booklet. Reference data available for 5;6–6;5-year-olds for the booklet version, and for 6;0–10;11-year-olds in two SLP master's theses (Krüger-Wahlqvist, 2012; Samson & Rasmussen, 2003).
- Peabody Picture Vocabulary Test, third edition (PPVT-III; Dunn & Dunn, 1997) and fourth edition (PPVT-IV; Dunn & Dunn, 2007). Assessment of receptive vocabulary via picture choice. The third edition has black and white drawings and the fourth edition contains colour drawings. Standardised and normed for American English. No official Swedish translation. American norms for age 2;9–adults.
- The New Reynell Developmental Language Scales (NRDLS; Edwards, Letts, & Sinka, 2011). Assessment of receptive and expressive vocabulary and morphosyntax. It has been translated and adapted for Swedish by Lundeborg Hammarström, Kjellmer and Hansson (2016), with norms from 530 children age 2;0–6;11.
- Picture sequences: informal assessment of sequencing and telling abilities, specific task or material not stated.
- Språkligt impressivt test för barn (SIT) [Impressive language test for children] (Hellqvist, 2011). Test for assessing language comprehension in children age 3–7. Reference data from 32 children.
- Språklig medvetenhet hos förskolebarn [Linguistic awareness in preschool children] (Lagergren & Larsson, 1992). Assessment of linguistic awareness, focusing on phonological aspects. No reference data.
- Språklig medvetenhet hos förskolebarn och skolbarn [Linguistic awareness in preschool and school age children] (Magnusson & Nauclér, 1993). Assessment of linguistic awareness, focusing on phonological aspects. Reference data exists for 6-year-olds, first graders (corresponding to age seven) and fourth graders (corresponding to age ten) with and without language impairment.
- Test for Reception of Grammar, second edition (TROG-2; Bishop, 2003). Assessment of comprehension of morphosyntactic structures by picture choice. Translated into Swedish and normed (N=650) for children age 4;0–12;11 (plus additional estimated norms for children age 13;0–16;5) (Bishop, 2009b).

Many of the tests and materials that are used in clinical practice by Swedish SLPs are originally in English. Some of them have been translated into Swedish, and have an official Swedish standardised version and Swedish norms (CELF-4, NRDLS, and TROG-2). Other tests are being used in translated versions that are not official (e.g. BNT). Some of the tests have reference data for Swedish-speaking children, often collected as part of an SLP master's thesis. However, the data collection procedure is rarely as thorough as is expected from a proper norming procedure (i.e. with respect to

representativeness and sample size), and thus they should not be considered as 'proper' norms. Among the standardised tests mentioned in this section, only GRAMBA, Nya NELLI, Ordracet and SIT were originally developed for Swedish.

In Table 3.18, the assessment procedure and materials are listed for each child in the clinical study. To the best of our knowledge, there are no standardised versions of any of these assessment materials that are available for Arabic, regardless of variety. Arabic-speaking SLPs may have their own 'standardised' translations, but whenever tests are administered via an interpreter, the test items are likely to be translated ad hoc during assessment. Table 3.19 contains a description of the child's language difficulties and communicative behaviour according to the SLP.

Table 3.18. Description of the procedure and materials used by the SLP when conducting language assessment.

Child	Assessment procedure and materials
BiAra LI-01	*Assessed in both languages by the SLP* Arabic: SIT Swedish: Ordracet, GRAMBA
BiAra LI-02	*Assessed in both languages by the SLP* Arabic: SIT Swedish: NRDLS, TROG-2, Nya fonemtestet, CELF-4 (parts), picture sequences (telling)
BiAra LI-03	*Assessed in both languages by the SLP* Arabic: SIT Swedish: NRDLS, Nya fonemtestet, CELF-4 (parts), GRAMBA
BiAra LI-04	*Assessed in both languages by the SLP* Arabic: SIT Swedish: SIT, TROG-2, PPVT-IV, GRAMBA
BiAra LI-05	*Assessed in both languages by the SLP* Materials (language not stated): CELF-4 (whole), BNT, MAIN (informal telling, not conducted according to standard procedure), PPVT-4, TROG-2, Bussagan, GrUS, FAS, NWR, fonem- och benämningstest vid flerspråkighet
BiAra LI-06	*Assessed in both languages by the SLP* Arabic: SIT Swedish: NRDLS, GRAMBA, Nya fonemtestet

BiAra LI-07	*Assessed in Swedish only* Materials: CELF-4 (whole), GRAMBA, Nya fonemtestet, Nya Nelli: ’Branden’ (telling), NWR
BiAra LI-08	*Assessed in both languages with interpreter* Arabic: PPVT-III, BNT, TROG-2 Swedish: PPVT-III, BNT, TROG-2, picture sequences (telling), språklig medvetenhet (Nauclér and Magnusson), NWR
BiAra LI-09	*Assessed in Swedish only* Materials: SIT, PPVT, Ordracet, GRAMBA, Bussagan, språklig medvetenhet (Lagergren and Larsson)
BiAra LI-10	*Assessed in both languages with interpreter* Assessment procedure and materials are not explicitly described. Informal assessment of basic concepts (colours; animals) and spontaneous speech at first assessment (age 2;3).
BiAra LI-11	*Assessed in both languages by the SLP* Materials (language not stated): CELF-4 (parts), SIT, GrUS, Nya fonemtestet, sequence pictures (telling)

Note. BNT=Boston Naming Test, CELF-4=Clinical Evaluation of Language Fundamentals, GRAMBA=Grammatiktest för barn, GrUS=Grammatisk undersökning av svenska som andraspråk, MAIN=Multilingual Assessment Instrument of Narratives, NWR=Non-word repetition, PPVT=Peabody Picture Vocabulary Test, NRDLS=The New Reynell Developmental Language Scales, SIT=Språkligt impressivt test för barn, TROG-2=Test for Reception of Grammar.

Table 3.19. Language difficulties and communicative behaviour according to the SLP.

Child	Language difficulties and behaviour
BiAra LI-01	He has immense language difficulties, both in comprehension and production. The most striking difficulty is his weak vocabulary and considerable difficulties expressing himself in an understandable way. He works well during therapy and does not seem to have concentration difficulties, but does not seem to make much progress.
BiAra LI-02	Weak comprehension skills seem to be the most prominent difficulty. She has increased her vocabulary and expressive skills in therapy, but still has difficulties with expressive morphosyntax and the vocabulary is not age appropriate. She works well during therapy and is motivated, but her concentration can drop sometimes when it gets too difficult.

BiAra LI-03	He has weak comprehension skills and difficulties with expressive morphosyntax. He seldom took communicative initiatives at the beginning of therapy, but this has improved over time.
BiAra LI-04	Weak vocabulary is the most prominent language difficulty, along with deficits in expressive morphosyntax and pronunciation difficulties. Comprehension difficulties were present at the start of therapy, but his comprehension ability is now judged to be age appropriate. He has developed a lot over time during therapy.
BiAra LI-05	The child was seen once by the SLP (for assessment). He had a hard time concentrating and had difficulties understanding instructions. Deficient comprehension is the most prominent language difficulty, but he also has weak expressive skills and a small vocabulary.
BiAra LI-06	He is very talkative and speaks in long utterances, but it can be difficult to understand him. He has pragmatic difficulties; he has a hard time getting his message across. His weak comprehension is the most prominent language difficulty.
BiAra LI-07	A weak vocabulary and deficient language comprehension seems to be the most prominent difficulties. He works well during therapy most of the time, but sometimes has a hard time focusing. He has developed his expressive ability in therapy.
BiAra LI-08	She has a weak vocabulary and difficulties with morphosyntax in both production and comprehension. Her expressive abilities are particularly affected. The SLP recommends that she should be assessed by a psychologist, but does not say why (e.g. suspecting ID or a neuropsychiatric disorder such as autism or ADHD).
BiAra LI-09	The assessment shows language difficulties across the board: deficits in comprehension, vocabulary, expressive morphosyntax and linguistic awareness.
BiAra LI-10	She has considerable language difficulties. She talks a lot, but is very difficult to understand. Her utterances are morphosyntactically and semantically incomplete. She mixes Arabic, Swedish and English sometimes without realising that others do not understand her. She likes interacting with peers, but the language difficulties often lead to conflicts.
BiAra LI-11	She is very quiet and shy. She is aware of her difficulties and avoids speaking because of this. She speaks in very short utterances, with notable pronunciation difficulties. She has deficient expressive morphosyntax and comprehension.

3.1.2.5 Teacher's evaluation of children's language abilities

During the (pre)school visit, the teachers were interviewed about the children's language skills and communicative behaviour. For the children who went to school, the main teacher was interviewed, and for the children who went to preschool, one of the preschool staff who knew them well were interviewed. For two children (BiAraLI-08 and BiAraLI-09), the main teacher was unavailable at the time. Thus, information was gathered from the special education teacher and the school SLP instead. For six children (BiAraLI-01, BiAraLI-04, BiAraLI-05, BiAraLI-06, BiAraLI-08 and BiAraLI-11), either the teacher or someone else from the staff spoke Arabic in addition to Swedish. This meant that they had an idea about the children's language abilities not only in Swedish but also in Arabic. The teachers were asked about the children's language abilities, social behaviour and interaction with other children and how they work to develop and facilitate the children's language skills. They were also asked which advice they would give to the parents of these particular children. The teacher's descriptions of the children's language abilities and communicative behaviour is presented in Table 3.20.

Many teachers mentioned that they try to work with the whole group to facilitate and develop the children's language skills, for instance by using augmentative and alternative communication methods (e.g. pictures and hand signs) and working with language exercises in smaller groups. However, a majority of the teachers experienced that they were short on resources, either because of lack of time or because they felt that they do not have the adequate knowledge needed to provide support for these children. When asked about advice to parents, many teachers said that it is important that they speak Arabic at home, since the children will develop their Swedish in (pre)school. Many teachers also stressed the importance of book reading and making sure that parents talk a lot with their children.

Table 3.20. Language difficulties and communicative behaviour according to the teacher.

Child	Language difficulties and behaviour
BiAra LI-01	He is eager to speak in front of the class, but is very difficult to understand. He has a weak vocabulary and speaks in short utterances with 2–3 words. He also has large difficulties with comprehension and needs individual instructions. There are frequent misunderstandings between him and his peers. He often loses attention when he does not understand.
BiAra LI-02	She used to produce nonsense words that were difficult to understand. Sometimes she would also imitate others. Her vocabulary has improved, but she has erroneous morphosyntax and word finding difficulties. She needs short and clear instructions in order to understand. She is easily distracted.

	Earlier she was often involved in conflicts, but it has happened less frequently during the past year.
BiAra LI-03	He is unfocused and has difficulties with listening. His speech is unclear and he does not speak in whole sentences. It is difficult to know how large his vocabulary is, since he has word finding difficulties. His comprehension seems better than production. He wants to decide everything when playing with his peers, which leads to conflicts.
BiAra LI-04	His language skills are on par with his peers. His vocabulary is not very strong, but most children attending the same preschool have a weak vocabulary in Swedish. The only thing that he has difficulties with is the pronunciation of some speech sounds. He plays well with the other children and is seldom involved in conflicts.
BiAra LI-05	He has difficulties with comprehension in all languages (Arabic, Swedish, and English). He has weak expressive skills and uses gestures and his own kind of 'sign language' to show what he means. He rarely speaks. He only has one friend in class. The mother is an English teacher and speaks English at home.
BiAra LI-06	His language skills have improved a lot during the past year. His comprehension and production skills are ok, but he has a weak vocabulary. He generally plays well with the other children, but sometimes he wants to decide everything. He is very meticulous about routines and rules, for instance with cleaning and washing his hands when he knows you are supposed to.
BiAra LI-07	He has a weak vocabulary and sometimes his sentences are 'funny', but most of the time he can make himself understood. He usually asks whenever he does not understand. He plays well with the other children, and is not involved in conflicts.
BiAra LI-08*	She has profound language difficulties. She has very limited expressive abilities in Swedish, and most often produces isolated words. Comprehension difficulties can be observed both during classroom activities and in interaction with peers and staff. Her language skills seem to be stronger in Arabic. She prefers to speak with children in Arabic, and does not seek contact with children who speak Swedish only. She has difficulties with social interaction and often behaves in ways that are perceived as impolite or mean by peers and school staff; this often leads to conflicts.
BiAra LI-09*	He is very shy, and prefers not to speak in front of the class. He is often quiet even when in a small group. However, he works well together with the special education teacher and has shown

	that he has good verbal reasoning skills and can make detailed statements.
BiAra LI-10	It is very difficult to understand her if you do not know her very well; she has her own way of communicating which involves mixing Arabic, Swedish, English and body language. She has weak comprehension and vocabulary skills. There are often misunderstandings which lead to conflicts with the other children. The preschool teachers often need to act as 'interpreters' between her and the other children. She has difficulties keeping focused, and it is tricky to know what came first: the short attention span or the language difficulties.
BiAra LI-11	The teacher does not know her very well, since she started school only two weeks prior to data collection. She is very quiet and unwilling to speak; she needs support in order to speak in front of the class. She seldom takes the initiative to play with the other children; when she does, it is with children she knows well. She speaks in short utterances, often containing only one or two words, but sometimes longer in Arabic. They use picture support to aid her comprehension.

Note: '*' = the schoolteacher was unavailable for an interview at the time of testing, so this information comes from the special education teacher and the school SLP.

3.1.2.6 The participants in the clinical study: summary

In sum, all children in the clinical study had been diagnosed with Developmental Language Disorder (in Swedish: 'språkstörning'). A majority (9/11) had 'general language disorder', encompassing both receptive and expressive skills. One child had 'expressive language disorder', and another child had an unspecified diagnosis due to the fact that he was referred for a psychological assessment, but the result of this assessment was not known to the SLP. Although some children had been assessed by a psychologist or were described as having a short attention span, none of them had any additional neuropsychiatric diagnosis such as autism or ADHD, or intellectual disability. All children had been exposed to Arabic since birth, but age of onset for Swedish varied from age 1 to age 5, which matched the children in the cross-sectional group. However, there was a higher proportion of the children in the clinical group who spoke an Iraqi variety of Arabic than in the cross-sectional group. No children were reported to have had any hearing problems, but many had family or relatives with language difficulties, and the proportion was higher than in the cross-sectional group. Most parents stated that the general language development was late in Arabic, but current language proficiency was often reported to be good or very good (on par with the children in the cross-sectional sample). However, when interviewed more in detail about their children's language skills, the majority of the parents described that their child had communication problems. Having a late onset of the first word or

the first multi-word utterance was more common in the clinical group compared to the cross-sectional sample. In most cases, the parents, the SLP and the teacher had similar views about the child's language skills and communicative behaviour, but there were some exceptions. The parents and the teacher of BiAraLI-04 thought that he had problems only with pronunciation, despite the fact that the SLP stated that he also had problems with expressive morphosyntax and a weak vocabulary in both languages. The mother of BiAraLI-05 did not think that her child had any language difficulties, but she was advised by the school staff to see an SLP for language assessment. She was of the opinion that her child has weak Swedish skills because he hasn't had sufficient input in Swedish. However, the SLP and the main teacher, both being speakers of Arabic, say that the child has weak skills in both languages. Finally, the parents and the school staff of BiAraLI-09 did not share the opinion of the SLP who described language difficulties in both comprehension and production. The parents and the school staff both said that the boy is very shy and avoids speaking in groups (e.g. in front of the class), but that he seems to have adequate language skills when talking one on one.

3.2 Materials

This section contains a description of the materials used in the current study: the Cross-Linguistic Lexical Tasks (CLTs; Haman et al., 2015) and four non-word repetition (NWR) tasks: a language-specific Swedish task (Radeborg et al., 2006), a Swedish and an Arabic version of the quasi-universal non-word repetition task (Chiat, 2015), and the Non-word Repetition Task-Lebanese (Abou Melhem et al., 2011).

3.2.1 Cross-linguistic Lexical Tasks (CLTs)

The Cross-linguistic lexical task (CLT) is a picture-based vocabulary assessment material (Haman et al., 2015). It was developed by a group of researchers (Working Group 3, focusing on phonological and lexical processing) within the COST Action IS0408.[24] Each CLT has four subtasks: comprehension of nouns, comprehension of verbs, production of nouns, and production of verbs. Each part consists of 30 items plus two practice items, making 120 test items, and eight practice items altogether.

The CLT was constructed to be used primarily with 3–5-year-old (bilingual) children. So far, there are versions for 29 languages.[25] All language

[24] COST is short for *European Cooperation in Science and Technology*, an organisation that has an EU-funded programme allowing researchers to initialise research networks for different topics. The COST Action IS0408 'Language Impairment in a multilingual society: Linguistic patterns and the road to assessment' ran from between 2009 to 3013.
[25] More language versions appear continuously, see http://psychologia.pl/clts/#projects.

versions are developed from a common list of concepts with corresponding pictures, i.e. no CLT version in one language is a translation from another language. Rather, each version is developed to suit the cultural and linguistic context for each language, and also takes into account the level of item difficulty (which differs between languages). Item difficulty is operationalised as an index of estimated age of acquisition and phonological and morphological complexity. In the present study, the Swedish version and an adaptation of the Lebanese Arabic version were used. The Swedish version was developed by Gisela Håkansson, Natasha Ringblom and Josefin Lindgren (Ringblom, Håkansson, & Lindgren, 2014). The Lebanese Arabic version was developed by Christel Khoury Aouad Saliby, Edith Kouba Hreich and Camille Messarra (Khoury Aouad Saliby, Kouba Hreich, & Messarra, 2017). The Arabic-speaking population in Sweden is diverse, consisting of speakers speaking many different varieties. Because of this, the Lebanese version was further developed by Rima Haddad, PhD student of Linguistics, in order for it to be valid to use in the Swedish context (Haddad, 2017). Dictionaries (English-Arabic) and speakers were consulted for varieties from nine different locations in the Levant and Iraq, as well as Modern Standard Arabic (MSA). The speakers (n=9, ages ranging from 30–45) were sent a list of the target items, in either Swedish or English. For comprehension items, the informants were asked *how would you express this concept?*, and for production items they were asked *what do you call this noun/verb?*. By using this procedure, a handful (1–3) of vocabulary labels could be identified for each variety as likely to be used by a young speaker of that variety. All in all, five different adaptations were constructed for Syrian, Palestinian, Lebanese and Iraqi (Mosul variety and Baghdad variety) Arabic.

The CLTs are presented in a booklet format.[26] The comprehension part consists of 60 test items (30 verbs and 30 nouns), plus four practice items (two for each sub-test). It is a picture-identification task, where the child is shown four colour pictures and is asked for example *where is the ant?* (nouns) or *who is waving?* (verbs). The child is instructed to indicate the correct picture by pointing (children who are familiar with numbers may opt to say the number of the picture instead). Examples are shown in Figure 3.1. The production part also consists of 60 test items (30 verbs and 30 nouns) and four practice items. This is a picture-naming task, where a child is shown one picture at a time, and asked *what is this?* (nouns), or *what is (s)he doing/what happens here?* (verbs). Examples are shown in Figure 3.2.

[26] However, some language versions exist in digital format.

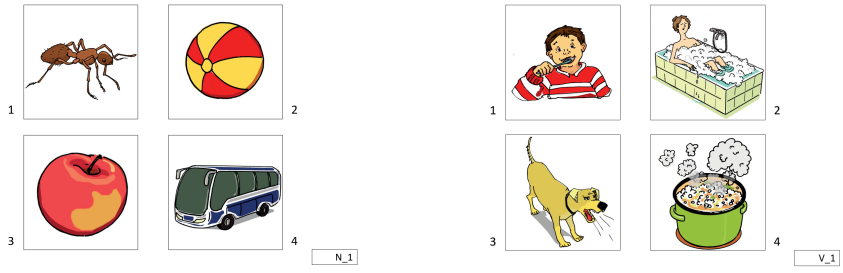

Figure 3.1. Examples from the Swedish CLT comprehension, nouns (left) and verbs (right). ©University of Warsaw.

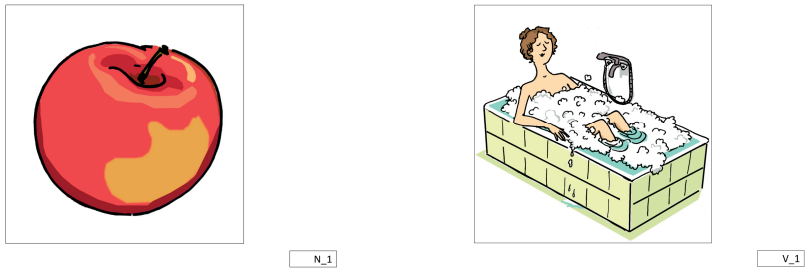

Figure 3.2. Examples from the Swedish CLT production, nouns (left) and verbs (right). ©University of Warsaw.

The CLTs were administered according to the standard procedure explained by Haman et al. (2015). Both language versions were presented to the child in booklets, with laminated pictures in A4 (comprehension) or A5 (production) format. The answers were noted on paper forms. All sessions were recorded with video as well as audio, for later inspection. For comprehension, the child was shown the pictures and was then asked to indicate the correct picture. If the child did not answer, or asked for clarification, the experimenter repeated the question once. If the child still did not answer, the experimenter noted 'no answer' in the form and proceeded with the next item. When conducting the comprehension subtasks in Arabic, the procedure was modified slightly from standard practice, in order not to disadvantage children who happened not to be familiar with the word in the variety of Arabic the experimenter used for a lexical concept. The children spoke different varieties of Arabic, originating from different geographic locations in the Levant region, Iraq and Egypt. Arabic varieties spoken in these areas vary considerably, not only across countries but also between different regions in countries. Therefore, it was decided that the children should be given multiple options of synonyms for the comprehension items if they seemed not to understand the word the experimenter used. Five different versions were created for CLT comprehension in Arabic: Two for Iraqi (Mosul variety and Baghdad variety),

one for Lebanese, one for Palestinian and one for Syrian. For each test item in each dialectal version, there was a list of words based on the information obtained from the dictionaries and informants (as described above). The experimenter was instructed to use the dialect version that each child was reported to speak, and prompted the child using the first word in the list. If the child did not seem to understand immediately, the experimenter asked the question again, but now using a synonym on the list. In some cases, there was also a third option that children were asked if they still did not seem to understand the word that the experimenter used. All in all, children could have up to three different word options in the Arabic comprehension part.

For CLT production, the child was asked to name the object or action shown in the picture, and the answer was written down on the form. If the child did not answer, or asked for clarification, the experimenter repeated the question once before proceeding to the next item. The experimenter only provided neutral feedback such as 'aha' or 'mmm' regardless whether the child's answer was correct or not, and did not answer any questions from the child about his or her performance. At the end of each subtask the child was given praise and a sticker.

3.2.2 Non-word repetition tasks

Non-word repetition entails repeating of a series of phonological nonsense forms. In the present study, four NWR tasks were used, all developed for children in preschool and early school age.

First, there was a Swedish language-specific task (LS-Swe), which was originally developed by Barthelom and Åkesson (1995) in their MA thesis in speech-language pathology, and was later published (Radeborg et al., 2006). The LS-Swe encompasses 24 test items of 2–5 syllables that follow Swedish phonotactics, and contains phonemes that are typical of Swedish; nineteen consonant phonemes (/p, b, t, d, k, g, m, n, ŋ, ɾ, f, v, s, ɕ, ɧ, ʂ, h, j, l/) and fifteen vowel phonemes (/i, ɪ, y, ʏ, e, ɛ, œ, ɑ, a, o, ɔ, u, ʊ, ʉ, ɵ/). The items have syllables with varying phonological complexity: there are open and closed syllables, with and without consonant clusters in onset and coda. Additionally, they are pronounced with stress patterns that are typical of Swedish, i.e. with varying main stress and vowel duration in different syllables in the non-word. The LS-Swe items were recorded by a female speaker of Swedish (speaking a central Swedish dialect).

Second, two different language versions (a Swedish and an Arabic) of an NWR task that was developed within the COST Action IS0408, the quasi-universal NWR task (Chiat, 2015), were used.[27] The task was designed to be compatible with the lexical phonology of many languages. As such, it contains

[27] The quasi-universal non-word repetition task is now referred to as the crosslinguistic nonword repetition task (CL-NWR) by Chiat & Polišenská (2016).

items of 2–5 syllables, with no consonant clusters and no codas (only open syllables). The full range of phonemes includes eleven consonants (/p, b, t, d, k, g, s, z, m, n, l/) and three vowels (/a, i, u/). For the purpose of this study, a Swedish version was created by the author and an Arabic version was created by Rima Haddad (PhD student of Linguistics). From a list of 84 candidate items, 16 items are chosen for each language version, excluding items that contain phonemes that do not exist in that language, or real words or inflections in that language.[28] Thus, the Swedish version contains no items with a /z/ (as Swedish does not contain /z/), and the Arabic version contains no items with a /p/ or a /g/ (as /g/ does not exist in some spoken varieties of Arabic). The QU-Swe items were recorded by a female speaker of Swedish. The QU-Ara items were recorded by a female speaker of Syrian Arabic, and the QU-Swe items were recorded by the same female speaker who recorded the LS-Swe items. In both languages, all items were pronounced with quasi-neutral prosody (Chiat, 2015, p. 138), where all syllables were equally stressed (i.e. they carried equal length and pitch) apart from final-syllable lengthening and pitch drop marking the end of an utterance.

Finally, there was another task that had been developed within the COST Action IS0408 network, the Non-word repetition task-Lebanese (NWRT-Leb; Abou Melhem et al., 2011), which was modelled on the NWR-FRENCH task (constructed in 2011; a later version of the French task is published in dos Santos and Ferré (2018)).[29] This task was constructed to investigate how phonological complexity impacts NWR performance in typically developing children and children with DLD. The task contains 30 items of 1–3 syllables with and without consonant clusters and codas. There were three different types of syllables, all present in Lebanese Arabic, French and English: CV, CCV or CVC. There are seven phonemes, four consonants (/b, l, k, f/) and three vowels (/a, i, u/), phonemes that all exist in Lebanese Arabic, French and English. Each item contained three to seven phonemes. The NWRT-Leb items were recorded at the Department of Speech and Language Therapy, St Joseph University, by a female speaker of Lebanese Arabic (Abou Melhem, 2017).

For all NWR tasks, audio files were created where each item was played one after the other, with a three second pause in between each non-word. All tasks were presented with increasing level of difficulty, i.e. starting with the items that were the shortest (had the lowest number of syllables), and gradually increased with one additional syllable. The audio recordings were incorporated into audio-visual power-point presentations to be administered

[28] In some cases it was not possible to find an item that did not breach one of the restrictions, in which case an item including syllable combinations that could make up real word were chosen. However, since all items were pronounced with a quasi-neutral prosody, these syllable combinations were not pronounced with typical stress, but with equal weight on all syllables.
[29] Both the Lebanese and the French tasks are now part of the 'LITMUS' test battery with tools designed for assessing language abilities in bilingual children (see Armon-Lotem, de Jong & Meir, 2015).

to the children in the form of an imitation task. The LS-Swe and the two QU tasks feature a parrot, and the NWRT-Leb features an alien that the child is instructed to imitate. A list of all items in the respective NWR tasks can be found in the Appendix (Tables A3.1–A3.4).

The child was instructed to imitate what the parrot or alien said, exactly as they said it. Before starting the presentation, the experimenter asked the child to repeat two practice items (not scored) in order to make sure that the child understood the instructions. The task was presented to the child on a smartphone, and the audio was played via noise-cancelling headphones. The task was played in its entirety without pauses. However, if the child was silent in the beginning of the task, the experimenter stopped, repeated the instruction and started over from the beginning. The experimenter did not answer any questions from the child about his or her performance. At the end of each NWR task the child was given praise and a sticker. All sessions were audio-recorded, and the NWR tasks were transcribed and scored after the session. For details on the transcription and scoring of the NWR tasks, see Chapter 5 (section 5.3).

3.3 Procedure

This section contains a description of the procedure that was used during data collection for the cross-sectional study (section 3.3.1) and the clinical study (section 3.3.2). The schemes used for counterbalancing the order of testing with respect to language and tasks will be described in section 3.3.3.

3.3.1 The cross-sectional study

The same general procedure was followed with each child. Data was collected on two separate occasions, one for each language. Half of the children were supposed to be tested in Arabic first, and the other half were tested in Swedish first. The aim was to have 5–7 days between the first and the second session, but due to events such as illness, school holidays or experimenter scheduling, this was not always possible. The time between the two sessions varied between 2–46 days, with the most common time span being 7 days.

All Swedish data was collected by the author, save for two children who were tested by Karin Koltay (SLP and trained research assistant). The Arabic data was collected by Rima Haddad and three trained research assistants, Zeinab Shareef, Amal Choumar and Pascale Wehbe. Each child was seen individually in a quiet room at (pre)school, in the home, or at a cultural or religious centre. One child was tested on the premises of Uppsala University. In order to ensure a monolingual setting, the experimenter talked to the child only in the language of testing, and pretended that she did not know the other language.

All assessment materials that were not currently being used were kept out of sight of the child. Before starting the session, the experimenter talked to the child and asked some simple questions in order to familiarise the child with the situation and to make sure that he or she could follow instructions and answer questions. After that, the experimenter explained to the child what they were going to do during the session, i.e. repeating funny words, looking at picture booklets, telling stories and answer some questions.[30] At the beginning of each session, the child was shown a visual timeline, where each task was represented with an adhesive icon (i.e. a parrot/alien for NWR, envelopes for the MAIN, and picture booklets for the CLTs). After each task was finished, the child could remove the icon. In this way, the child was prepared for what was going to happen next and always knew how many tasks were left to complete.

The testing proceeded according to the predefined order for each child, see section 3.3.3 for a description of the counterbalancing systems. The experimenter gave the same instructions to all children, closely following a prewritten script for each task. Each data collection session took approximately 30–45 minutes.

All sessions were both video and audio recorded. The video camera was placed next to the child and the experimenter, if possible slightly behind child in order to capture pointing, gestures and facial expressions.[31] The audio recorder was placed on the table in close proximity to the child's mouth.

The same procedure was followed in the Arabic and Swedish session. First, all children did one NWR task (LS-Swe, NWRT-Leb, QU-Swe or QU-Ara), then told one story from the MAIN (Cat or Dog). This was followed by the CLT, which was counterbalanced between the subtasks (nouns and verbs, comprehension and production). Next, a second MAIN story was told (Baby Birds or Baby Goats), and after that a second NWR task was completed (LS-Swe, NWRT-Leb, QU-Swe or QU-Ara). The task order is outlined in Figure 3.3. After each task, the child was given a sticker, and at the end of each session they were praised and thanked for their participation. After the second session, they were given a diploma.

[30] The test battery also included a narrative task, the Multilingual Assessment Instrument for narratives (MAIN; Gagarina et al., 2012). The MAIN was designed to assess narrative abilities in children aged 3–9, particularly regarding comprehension and production of narrative macrostructure. The material contains two pairs of depicted narratives that were constructed to be parallel in their story grammar structure, thus enabling a comparison between narratives told in two different languages by a bilingual child. The material can also be used to investigate microstructural aspects of narratives, such as morphosyntactic complexity or lexical diversity. The results from the MAIN are not included in this dissertation.

[31] This setup was relevant for the transcription of the narrative data.

Figure 3.3. Task order. NWR=Non-word repetition, MAIN=Multilingual Assessment Instrument for Narratives, CLT=Cross-linguistic Lexical Task.

3.3.2 The clinical study

The children were tested once in Arabic (either in the home or at the (pre)school) by Pascale Wehbe or Rima Haddad, and once in Swedish by the author at the (pre)school. Half of the children were tested in Arabic first, and the other half in Swedish first. The children were tested with the same materials as in the cross-sectional study (NWR, the MAIN and the CLT). The procedure matched the one used in the cross-sectional study: all children were tested in a quiet room, with the same placement of the video camera and the audio recorder, and all tasks were presented in the same way as in the cross-sectional study.

3.3.3 Counterbalancing

Counterbalancing of the different tasks and which language was tested first was carried out within each age group in the cross-sectional study. As previously mentioned, half of the children should start with Arabic first and the other half with Swedish first. Due to circumstances such as differently sized age groups, drop-outs and excluded participants, 53 children were tested in Arabic first, and 46 children were tested in Swedish first. The counterbalancing system is displayed in Table 3.21 for Arabic as first language and Table 3.22 for Swedish as first language. The system was repeated for every 8th child.

Table 3.21. Counterbalancing of the tasks for children who started with the Arabic session.

	Arabic 1					Swedish 2				
Child	**NWR1**	**MAIN1**	**CLT**	**MAIN2**	**NWR2**	**NWR1**	**MAIN1**	**CLT**	**MAIN2**	**NWR2**
1	NWRT-Leb	Cat	1	BB	QU-Ara	LS-Swe	Dog	1	BG	QU-Swe
2	QU-Ara	Cat	2	BG	NWRT-Leb	QU-Swe	Dog	2	BB	LS-Swe
3	NWRT-Leb	Dog	3	BB	QU-Ara	LS-Swe	Cat	3	BG	QU-Swe
4	QU-Ara	Dog	4	BG	NWRT-Leb	QU-Swe	Cat	4	BB	LS-Swe
5	NWRT-Leb	Cat	1	BB	QU-Ara	LS-Swe	Dog	1	BG	QU-Swe
6	QU-Ara	Cat	2	BG	NWRT-Leb	QU-Swe	Dog	2	BB	LS-Swe
7	NWRT-Leb	Dog	3	BB	QU-Ara	LS-Swe	Cat	3	BG	QU-Swe
8	QU-Ara	Dog	4	BG	NWRT-Leb	QU-Swe	Cat	4	BB	LS-Swe

Note. BB=Baby Birds, BG=Baby Goats (MAIN stories).

Table 3.22. Counterbalancing of the tasks for children who started with the Swedish session.

	Swedish 1					Arabic 2				
Child	**NWR1**	**MAIN1**	**CLT**	**MAIN2**	**NWR2**	**NWR1**	**MAIN1**	**CLT**	**MAIN2**	**NWR2**
1	QU-Swe	Cat	1	BB	LS-Swe	QU-Ara	Dog	1	BG	NWRT-Leb
2	LS-Swe	Cat	2	BG	QU-Swe	NWRT-Leb	Dog	2	BB	QU-Ara
3	QU-Swe	Dog	3	BB	LS-Swe	QU-Ara	Cat	3	BG	NWRT-Leb
4	LS-Swe	Dog	4	BG	QU-Swe	NWRT-Leb	Cat	4	BB	QU-Ara
5	QU-Swe	Cat	1	BB	LS-Swe	QU-Ara	Dog	1	BG	NWRT-Leb
6	LS-Swe	Cat	2	BG	QU-Swe	NWRT-Leb	Dog	2	BB	QU-Ara
7	QU-Swe	Dog	3	BB	LS-Swe	QU-Ara	Cat	3	BG	NWRT-Leb
8	LS-Swe	Dog	4	BG	QU-Swe	NWRT-Leb	Cat	4	BB	QU-Ara

Note. BB=Baby Birds, BG=Baby Goats (MAIN stories).

As displayed in Table 3.23, the CLT subtasks were also counterbalanced between the four subtasks, following the guidelines by Haman et al. (2015).

Table 3.23. Counterbalancing system for the Cross-Linguistic Lexical Tasks (CLTs). Comp = comprehension, prod = production.

	Order 1	**Order 2**	**Order 3**	**Order 4**
Part 1	Noun comp	Verb comp	Noun prod	Verb prod
Part 2	Verb comp	Noun comp	Verb prod	Noun prod
Part 3	Noun prod	Verb prod	Noun comp	Verb comp
Part 4	Verb prod	Noun prod	Verb comp	Noun comp

To give an example of how the counterbalancing system was applied, child number 2 in Table 3.21 started with the QU-Ara NWR task, did the CLTs according to CLT order 2 and finished with the NWRT-Leb in Arabic in the first session. During the second session, the same child started with the QU-Swe task, did the CLTs according to CLT order 2 and finished with the LS-Swe task in Swedish.

For the clinical study, the same counterbalancing system was utilised as in the cross-sectional study, with the exception that all children did the CLTs according to order 1.[32] Thus, six children started with Arabic in the first session, and five children started with Swedish in the first session.

[32] This was done since it was assumed that some of the children in the clinical study would have considerable difficulties with speech production. Thus, starting with noun comprehension was a way of making the task less demanding.

4 The cross-sectional study: Vocabulary

Vocabulary is a central aspect of knowing a language, and serves as an effective medium for communicating the meaning of concrete objects, abstract concepts, actions, events and properties. Vocabulary is also an important resource for learning; vocabulary size during preschool is a strong predictor of reading comprehension beyond the early school years (Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017). At group level, scoring low on vocabulary tests is common for bilinguals (compared to monolingual peers) as well as children with DLD. Therefore, it is crucial to identify which factors influence vocabulary development in bilingual pre-schoolers. While it is not possible to measure exactly how many words a child has in their mental lexicon, one can make comparisons with other children the same age via vocabulary tests or checklists. In research investigating vocabulary skills in individuals, a distinction is typically made between two main aspects of vocabulary knowledge: vocabulary breadth (the size of the lexicon) and vocabulary depth (how much knowledge you have of a word).[33] The focus in this study will be on vocabulary breadth, as measured with the Cross-linguistic lexical tasks (CLTs), encompassing both comprehension and production.

The setup of the chapter is as follows. Earlier studies of bilingual vocabulary development are summarised in the background section (section 4.1). Scoring of the CLTs is explained (section 4.2), followed by analyses and results for vocabulary comprehension and production in Arabic and Swedish (Section 4.3). The chapter closes with a section where the findings from the current study are discussed in light of previous studies (section 4.4).

In this chapter, the vocabulary development of 99 bilinguals aged 4–7 is explored cross-sectionally in the minority language Arabic and in the majority language Swedish. The following general research questions are asked:

- How does vocabulary comprehension and production develop with age in Arabic and in Swedish?
- Are there any differences between vocabulary comprehension and production in Arabic and Swedish?

[33] However, see Vermeer (2001) for a critical discussion on the binary separation of these aspects of vocabulary knowledge. Vermeer argues that there is no conceptual distinction between the two, but rather that they are strongly related to each other.

- What is the relationship between age, language input, socio-economic status (SES) and vocabulary scores?

4.1 Background literature: Bilingual vocabulary development

Studies of vocabulary size typically explore either receptive (comprehension) or expressive (production) knowledge. Vocabulary comprehension is usually assessed with picture selection tasks, such as the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) and vocabulary production is often assessed with picture naming tasks such as the Boston Naming Test (BNT; Kaplan et al., 1983).[34] The Cross-Linguistic Lexical Tasks (CLTs) – used in the current study – include two receptive picture choice tasks as well as two expressive picture naming tasks targeting nouns and verbs respectively.

Children's vocabulary is expected to grow with age, due to cognitive development and an increase in the number of concepts encountered in the input. In bilinguals, certain factors may impact the developmental trajectories differently in each of their languages, the main two factors being SES and input. As we shall see in the following, the effect of SES is often mediated by input variables, as parental level of education or income tends to co-vary with input patterns in some contexts.

This background section provides a review of the literature on vocabulary development in children. The primary focus will be on bilingual children in the late preschool and early primary school years, but whenever relevant, studies on monolinguals and younger or older children will also be mentioned. The first section (4.1.1) provides a general overview of development with age. The second section (4.1.2) reviews the role of language input. The third section (4.1.3) discusses the effect of SES with regards to development in the minority as well as the majority language. In the fourth (4.1.4) and fifth (4.1.5) sections, previous studies on vocabulary in monolingual and bilingual Swedish-speaking and Arabic-speaking children are summarised.

4.1.1 Age

Of relevance for the current study, Haman et al. (2017) conducted a large scale cross-linguistic study of 639 monolingual children from 15 different countries (age 3;0–6;11), who were tested on the CLT in their respective language. The 17 languages were Afrikaans, British English, Catalan, Finnish, German, Hebrew, isiXhosa, Italian, Lithuanian, Luxembourgish, Norwegian, Polish,

[34] Additionally, parental checklists such as the MacArthur-Bates Communicative Development Inventories (CDIs) are often used to investigate vocabulary skills in very young children (age 1–3) (Fenson et al., 1993).

Serbian, Slovak, South African English, Swedish, and Turkish. There was a moderate to strong positive correlation between age in months and vocabulary scores in the whole sample, as well as for 11 of the individual languages (one of them being Swedish). In languages where no age development was present, small samples, narrow age ranges, or both, could explain the lack of an age effect. Overall, Haman et al.'s study showed that monolingual children in the preschool age can be expected to increase their performance on the CLT as they grow older. Moreover, word class effects were investigated for 16 of the 17 language versions. In 13 of the investigated languages, children scored better on nouns than on verbs (although this was not the case for Swedish).

In the Welsh bilingual societal context, Gathercole, Thomas, Roberts, Hughes and Hughes (2013) used a cross-sectional design with 427 children in four age groups (2–3-year-olds, 4–5-year-olds, 7–8-year-olds, and 13–15-year-olds) to explore the development of vocabulary comprehension in English and Welsh. In Wales, both English and Welsh enjoy high status, and neither language is associated with a particular social or ethnic group. They used the British Picture Vocabulary Scale (BPVS, 2nd ed.; Dunn, Dunn, Whetton, & Burley, 1997) in English, and a Welsh vocabulary comprehension test, the *Prawf Geirfa Cymraeg*, which was developed for the purpose of the study. The participants were divided into four groups, depending on the type of language use in the family: monolingual English, bilinguals hearing only English at home, bilinguals hearing Welsh and English at home, and bilinguals hearing only Welsh at home. Gathercole et al. found that children in all groups improved their scores in the majority language English with age. Also, bilinguals from all three exposure type groups improved their scores in the scores minority language Welsh with age.

However, many bilingual children, especially those with immigrant backgrounds, do not grow up in communities where both languages have equal status. This means that as children grow older, they become more oriented towards the majority language, in part as a result of schooling and generally becoming more involved with the majority community (Pearson, 2007). In fact, many studies find that while bilingual children increase their vocabulary scores in the majority language over time, vocabulary growth curves may not increase to the same extent in the minority language, or even stagnate (Bialystok, Luk, Peets, & Yang, 2010; Cobo-Lewis et al., 2002a, 2002b; Dijkstra, Kuiken, Jorna, & Klinkenberg, 2016; Gagarina et al., 2014; Ganuza & Hedman, 2019; V. C. M. Gathercole & Thomas, 2009; Leseman, 2000; Lindgren, 2018; Lindgren & Bohnacker, 2020; Öztekin, 2019).

For bilingual children, time may influence vocabulary development in their respective languages in different ways compared to monolinguals. Not only does it vary at which age children become bilingual, but patterns of language use in the family may also change during childhood, which in turn influences the child's input patterns. This means that even though bilingual children are expected to increase the number of *concepts* in their mental lexicon as they

grow older, the number of vocabulary *labels* does not necessarily increase to the same extent in both languages. Rather, there is often an imbalance in the number of vocabulary items known in each language. Furthermore, which language is the dominant one may shift over time as a result of changes in the conditions affecting language input. The role of language input will be discussed further in the next section.

4.1.2 Input

Studies of the role of input in language development are not only interested in the *quantity* (amount) of input, but also the *quality* (what kind) of input a child receives. In their seminal study, Hart and Risley (1995) followed 42 monolingual English-speaking children in the USA longitudinally, starting at around 12 months, and measured the input they received in the home repeatedly for about 2.5 years.[35] Hart and Risley found that there were substantial differences regarding the amount of child-directed speech in the home. These differences had a large cumulative effect on how many words the children heard over time. The children who received more input had larger expressive vocabularies, and they had steeper vocabulary growth curves compared to the children who received less input, as measured by the number of tokens in the children's speech. In families who talked more, the quality of the input was also generally higher (as measured by a number of variables connected to variation in the input and parental interaction style). Furthermore, these patterns were found to be related to SES; children who grew up in higher-SES homes generally had more language input of higher quality, and as a result they had larger vocabularies and steeper growth curves than children from lower-SES backgrounds.

Other studies concerning monolinguals have since confirmed the importance of quality and quantity of input in children's vocabulary development. Rowe (2012) followed 50 monolingual English-speaking children in the USA and videotaped free-play sessions in the home with their primary caregiver at three times, when the children were 18, 30 and 42 months. A vocabulary comprehension test (the PPVT) was administered at 30, 42 and 54 months, and analyses were conducted to see which aspects in the parental input that predicted children's vocabulary scores one year later. The analyses showed that during the second year of life, *quantity* was the most important predictor. During the third and fourth years of life however, measures of input *quality* in the parents' speech, such as vocabulary diversity and sophistication as well as the use of decontextualised language such as narrative or explanatory discourse, were the best predictors of the children's vocabulary comprehension one year later.

[35] The input variables included quantity measures such as the average number of words addressed to the child per hour, as well as quality measures – richness and variation.

Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina and Trustwell (2013) investigated referential transparency in the video recordings from Rowe (2012) by asking adult raters to make quality assessments of the parent-child interactions. Clips from the free-play sessions at 14 and 18 months were shown to the participants, who were asked to guess the target word from context alone (only nouns were used as targets, and the word was muted by a beep). Each vignette was assigned a referential transparency score based on the percentage of correct ratings made by the adult raters. The analyses showed that increased *informativity* (i.e. a higher referential transparency score) had a positive effect on children's vocabulary comprehension (PPVT scores) three years later.

In sum, for very young children sheer quantity of input (number of words) seems to be facilitative in increasing vocabulary size, but beyond the second year of life, the quality of the input grows more important. Although both Hart and Risley (1995) and Cartmill et al. (2013) claim that increased quantity of language input also leads to an increased number of high quality interactions, Rowe (2012) suggests that it is more important to concentrate on the quality of child-directed speech rather than on sheer quantity only.

In bilingual language development, investigating the effect of input means that the added factor of being exposed to two languages also needs to be taken into account. Bilinguals have more sources of variation in the input, i.e. how much time during the day and in which contexts children are exposed to each language (Paradis & Grüter, 2014). Measures of input quantity can be operationalised in different ways. A common way of calculating input quantity is to estimate percent daily exposure to each language via parental questionnaires. These estimates may be simple approximations of current daily exposure to each language. Another common measure is to ask for the age of acquisition for each language, which in turn can be used to calculate the length of exposure. Both measures are used in the present study. Other studies make use of more elaborate measures where parents are asked to report how many hours a week and in which contexts the child has been exposed to each language over a longer period (months or years). Such measures try to get at the cumulative effect of input over time (Thordardottir, 2011; Tuller, 2015). Yet another way of operationalising input quantity is by comparing groups of bilingual children with different patterns of language use in the family, i.e. children hearing only the minority language at home, vs. children hearing both languages, vs. children hearing only the majority language at home, such as the aforementioned study by Gathercole et al. (2013) on Welsh-English bilinguals.

Many studies have investigated the effect of input on the language proficiency in each language of bilingual children. Hoff, Core, Place, Rumiche, Señor and Parra (2012) used the English and Spanish versions of the MacArthur-Bates CDIs with 103 toddlers growing up in South Florida, the US. They were divided into four groups: monolingual English (N=56),

bilingual English dominant (N=18), balanced bilinguals (N=14), and bilingual Spanish dominant (N=15). Measures from the CDIs were collected at 1;10, 2;1 and 2;6 years, and estimates of language exposure were obtained through parental interviews. While the children who were dominant in English and the balanced bilinguals performed much closer to the monolingual English norm, the children who were dominant in Spanish performed substantially lower in English than all other groups. For the minority language Spanish, the authors found that the Spanish-dominant children received the highest scores, and that both balanced bilinguals and the English-dominant children scored much lower and were indistinguishable from each other in the statistical analysis. Thus, at group level, even children who received equal amounts of input in the minority and the majority language performed on par with the English-dominant children. However, it should be noted that at this age there is a great deal of variance in the vocabulary size of monolinguals. Thus, scoring within the monolingual norm may be 'easier' compared to older children.

In a similar study of slightly older Spanish-English bilinguals, Hoff, Rumiche, Burridge, Ribot and Welsh (2014) compared the development of expressive vocabulary over time in three groups of children: monolingual English-speaking (N=31), bilinguals with only one of their parents being a native speaker of Spanish (N=15), and bilinguals with both parents being native speakers of Spanish (N=11). All children were from mid-high-SES backgrounds, as the majority of the parents in all groups had college degrees. The children were followed in a longitudinal design. Measures of vocabulary production were obtained at 22, 25 and 30 months with the CDIs, and at 48 months with the Expressive One Word Picture Vocabulary Test (EOWPVT). Hoff et al. found that at group level, the bilingual children performed at or near the monolingual norm in the majority language.[36] However, the group of children who had two Spanish-speaking parents had significantly lower scores in English at 48 months compared to the monolinguals. Furthermore, the vocabulary scores of children who had one English-speaking parent were boosted if the parents spoke more English at home. The same pattern was not found for children who had two Spanish-speaking parents. On the contrary, speaking more English in the home did not increase these children's English expressive vocabulary but rather had a negative effect on their Spanish expressive vocabulary, thus running the risk of promoting subtractive bilingualism rather than facilitating the development of the majority language.

Cattani, Abbott-Smith, Farag, Krott, Arreckx, Dennis and Floccia (2014) compared 35 bilingual children (with 18 different L1s) to 36 monolingual children between 28–32 months growing up in the UK on a set of language measures including CDIs and vocabulary comprehension (the British Picture

[36] Again, it should be emphasised that there is much variation in the vocabulary size of monolinguals in early childhood, thus, scoring within the monolingual norm may be easier at this age than when children grow older.

Vocabulary Scale, BPVS) in English. They found that children who received at least 60% exposure to English were highly likely to score within monolingual norms on the CDI and BPVS measures.

The role of input for language attainment in bilinguals has also been studied in preschool and school age children, and some of these studies will be reviewed in the following. Thordardottir (2011) investigated the relationship between the relative amount of exposure to English and French and receptive and expressive vocabulary in 84 4-year-olds (range: 4;6–5;0) growing up in Montreal, a bilingual community in Canada where English and French enjoy equal status. Thirty-five children were monolinguals: 19 were exposed to French only, 16 to English only, and 49 children were bilingual with varying cumulative exposure during the life span (20 had mostly French, 16 had mostly English and 13 had roughly equivalent amounts of exposure to both languages). All bilingual children had received continuous exposure to both languages starting before age three. Children who received 40–60% exposure to one language over time scored similarly to monolinguals on *receptive* vocabulary in that language, but much more input (80%) was needed in order for children to acquire monolingual-like proficiency in *expressive* vocabulary.

Turning back to the aforementioned study by Gathercole et al. (2013) on Welsh-English bilinguals, Gathercole et al. found that differences between the groups (such as the children being exposed to more English scoring better in English than children from more Welsh-oriented homes, and vice versa) tended to diminish over time. That is, there were larger differences between the exposure type groups in the younger ages than in the older ages. This diminishing difference over time likely reflects a change in the exposure patterns for these children, for instance as an effect of schooling (increasing exposure to the language that is less frequently spoken in the home).

Prevoo, Malda, Mesman, Emmen, Yeniad, van Ijzendoorn and Linting (2014) investigated the effect of SES, maternal language use, and home reading input on minority and majority language vocabulary in a group of 111 Turkish-speaking children (age 5;5–6;10) in the Netherlands. The children were tested on vocabulary comprehension in Turkish (PPVT), and on vocabulary production in Dutch (EOWPVT). The mothers were asked to rate on a five-point scale how much the child was exposed to Turkish and Dutch respectively. The authors found that there was a positive correlation between age and productive vocabulary in the majority language Dutch, but no correlation between age and receptive vocabulary in the minority language Turkish. However, there were clear correlations between the relative amount of input in the minority and majority language and vocabulary scores in both Turkish and Dutch.

Bohnacker, Lindgren and Öztekin (2016) investigated the effect of age, SES, and a number of input variables on expressive vocabulary in the minority language in two groups of 4–6-year-old Swedish-speaking bilinguals with German (N=38) or Turkish (N=40) as their home language. Neither age nor

SES affected the scores in the home language in either group, but parental language use was strongly linked to vocabulary scores. Children had significantly higher vocabulary scores when the parents spoke only the home language to each other, or when the parents spoke mainly the home language to the child. Furthermore, for children whose parents did not speak mainly in the home language with their child (or if only one parent did so), receiving additional input from other interlocutors such as friends had a positive effect on expressive vocabulary scores in their home language.

Finally, something should be said about age of onset of bilingualism and the impact on vocabulary development. Much of the research about bilingualism during early childhood has centred around the question of whether there is a fundamental difference between bilingualism that starts at or close after birth (so-called simultaneous bilingualism) and bilingualism that starts after the basics of the first language has been established (sequential bilingualism). The exact definitions of early vs. late age of onset vary between researchers and studies, but many agree to draw the line around age three. That is, simultaneous language acquisition happens when a child is regularly exposed to two languages before age three, and sequential bilingualism happens after age three (Paradis, 2007). Several researchers have argued that there is a fundamental difference between simultaneous and sequential bilingual language acquisition, see for instance De Houwer (2009, pp. 4–7) and Meisel (2009).

Some studies have compared the vocabulary size in the majority language between simultaneous and sequential bilinguals. However, since researchers do not always operationalise simultaneous and sequential bilingualism the same way, it is difficult to compare the findings across studies. For instance, in the aforementioned study by Thordardottir (2011) on Canadian French-English 4-year-olds, the children who had an early age of onset to English (before 6 months) did not score better on vocabulary comprehension or production compared to the children who had a late age of onset to English (after 20 months). By contrast, Gross, Buac and Kaushanskaya (2014) found that for Spanish-English bilinguals aged 4–7, simultaneous bilinguals (N=39) outperformed sequential bilinguals (N=19) on both vocabulary comprehension and production. Gross, Buac and Kaushanskaya defined simultaneous bilinguals as those children who started to *produce* two-word phrases in English *before* age three, and sequential bilinguals as those children who started to *produce* two-word phrases in English *after* age three.

As Kupisch (2018) points out, age of onset to the majority language may co-vary with other factors influencing language acquisition, one such factor being *amount of exposure* to each language. Therefore, investigating the effect of age of onset while controlling for amount of exposure may uncover the relative effect of each factor. Unsworth (2016) compared the Dutch receptive vocabulary of English L1/Dutch L2 children (N=87) with an age of onset to Dutch before age four (N=44, mean age of onset for Dutch: 2;4, mean age at

testing: 7;3) and after age four (N=43, mean age of onset for Dutch: 5;5, mean age at testing: 9;1) respectively. There was no significant difference in receptive vocabulary scores between children with an age of onset before vs. after age four. Unsworth also found that current amount of exposure to Dutch was a significant covariate; a higher degree of current daily exposure to Dutch was associated with a higher score on the receptive vocabulary test. However, there was no significant effect of cumulative length of exposure to Dutch or age at testing.[37] Thordardottir (2019) investigated the effect of acquisition type (simultaneous bilinguals vs. sequential bilinguals vs. monolinguals), age of exposure (the age at which significant exposure to French started) and cumulative amount of exposure (reported amount of hours spent in each language throughout the child's life) on vocabulary comprehension and production in French in 132 children in grade 1 and 3.[38] Sequential bilingualism was defined as having an age of onset after 36 months (3 years). Analyses on the whole sample of children (including children who had regular exposure to languages other than English or French) indicated that acquisition type (simultaneous vs. sequential) was not a relevant predictor of vocabulary scores, when age of exposure and cumulative amount of exposure to French was taken into account. An additional analysis was conducted on a subsample of the children who were being exposed exclusively to English and French (N=49), with receptive vocabulary in French as the outcome measure. It showed that age of onset was not a significant predictor of French vocabulary scores in the early school years, while cumulative amount of exposure to French was. In fact, it was even a slightly stronger predictor than chronological age.

In sum, input quantity and quality affect children's vocabulary development, where more input of better quality facilitates overall vocabulary growth. For bilinguals, variations in patterns of minority vs. majority language use further influence the development of vocabulary in the respective language. If parents speak the minority language at home, it seems to boost minority language vocabulary skills. By contrast, if parents speak the majority language at home (when not being native speakers of that language), it does not seem to boost majority language skills but rather hampers the development of vocabulary skills in the minority language. The studies summarised here show that there is large variation in the vocabulary skills of bilingual children. However, some general conclusions can be drawn with respect to input and the connection to vocabulary development. As shown by Thordardottir (2011), a higher degree of relative input is needed for a bilingual child to score

[37] *Cumulative length of exposure* here means the proportion of exposure to each language based on language input from the parents and in (pre)school. The proportion was calculated separately for each one-year-period in the child's life, and then summed up, thus giving the total amount of exposure to Dutch (in years) over time (Unsworth, 2013a).

[38] The children in grade 1 had a mean age of 6;10 (years;months), and the children in grade 3 had a mean age of 8;10 (years;months).

within the monolingual norm for vocabulary production compared to comprehension. Furthermore, although it is often claimed that there is a fundamental difference between simultaneous and sequential bilingualism, the findings by Thordardottir (2011, 2019) and Unsworth (2016) suggest that such binary categorisations of simultaneous vs. sequential bilinguals (strictly based on age of onset of bilingualism) may not be relevant classifications for predicting later vocabulary size, at least not in the early school age. Rather, more fine-grained measures of cumulative input over time as well as current relative exposure to each language seem to better explain the effect of language experience on vocabulary skills. Finally, patterns of language input seem to be related to SES, which will be discussed in the next section.

4.1.3 Socioeconomic status (SES)

As previously mentioned, SES has emerged as a key factor in vocabulary development. SES can be defined and operationalised in different ways, commonly by level of education, occupation or income, or any combination of these measures. A common way of operationalising SES is to ask for years of education of both caregivers, or by the primary caregiver (often the mother). There are several potential factors related to SES that may affect children's vocabulary development. For instance, SES may affect the chances of education, as well as the quality of education. Furthermore, SES may affect the quality and quantity of linguistic input a child receives. As mentioned above, Hart and Risley (1995) showed that SES was related to vocabulary use, vocabulary growth in a group of monolingual toddlers growing up in the USA. Rowe (2012) found that SES (education level of the primary caregiver) was a significant predictor of higher-quality input during parent-child interaction, as well as better receptive vocabulary scores in a group of 50 monolingual children growing up in the USA. For bilinguals however, SES may affect the home language and the majority language differently. According to a review article by Pearson (2007) on social factors and childhood bilingualism in the USA, SES has a clear influence on majority language development, but its role in minority language maintenance is less consistent. In the following, some studies concerning vocabulary development and SES in bilingual children will be reviewed.

Cobo-Lewis, Pearson, Eilers, and Umbel (2002a, 2002b) examined majority and minority language skills in a group of 952 school-age children in kindergarten, 2nd grade and 5th grade growing up in the USA. They investigated the effect of SES on a number of language and literacy skills, among them vocabulary comprehension and production in the majority language English and the minority language Spanish. They found that children from high-SES families did significantly better in the majority language English in most tasks. For the minority language Spanish, the results were not as conclusive; children from households with low SES scored higher

compared to those with high SES on two tasks targeting vocabulary skills in the oral domain, but there was no difference in performance between the groups on most tasks.

Calvo and Bialystok (2014) investigated the effects of bilingualism and SES on vocabulary comprehension (PPVT) in the majority language English in a group of 175 6–7-year-olds growing up in Toronto, Canada. Information on the children's SES was obtained via parental questionnaires, where parents reported their education, occupation and income level. Bilingual children from lower-SES backgrounds had significantly lower scores than did the bilinguals from high-SES backgrounds.

Prevoo et al. (2014) investigated the vocabulary skills of 111 Turkish-Dutch bilinguals (age 5;5–6;10), and found that while there was no relation between SES and vocabulary scores in the minority language Turkish, children from high-SES backgrounds generally had higher scores in the majority language Dutch. Structural equation models were utilised in order to disentangle the causal effects of multiple predictors. They showed that SES predicted maternal language use, which in turn predicted the child's vocabulary skills in both Turkish and Dutch. The relative exposure to each language thus mediated the effect of SES.

Buac, Gross and Kaushanskaya (2014) investigated the effect of primary caregiver vocabulary knowledge on the vocabulary skills of 58 bilingual Spanish-English speaking 5–7-year-olds growing up in Wisconsin, the US. The children were tested on vocabulary comprehension and production in English and Spanish, as were their primary caregivers (for most children this was the mother but for some it was the father (N=10), or the grandmother (N=2)). While SES emerged as a significant predictor of the children's vocabulary skills in the majority language English, no relationship between SES and vocabulary skills was found for the minority language Spanish. Furthermore, English vocabulary skills were highly sensitive to primary caregivers' vocabulary skills in English. The authors suggested that the primary caregivers' vocabulary skills in the majority language might serve as the underlying factor affecting children's vocabulary scores in the majority language.

Meir and Armon-Lotem (2017) investigated the performance of vocabulary production in the majority language (Hebrew) in 88 Russian-Hebrew-speaking 5–6-year-old bilinguals in Israel, half with low SES, and half with mid-high-SES. Contrasting findings in many previous studies, Meir and Armon-Lotem did not find any differences in scores between the two groups.

In short, higher SES is generally associated with better vocabulary scores in the majority language (but see for instance Meir and Armon-Lotem (2017) for differing results). Prevoo et al. (2014) and Buac et al. (2014) suggest that this effect may in part be a result of the fact that in many contexts, SES tends to co-vary with parental language skills in the majority language. Caregivers with high SES often have a higher proficiency in the majority language, which

may serve as a positive influence of children's majority language skills. For the minority language, results are not as conclusive. For instance, for Spanish-English children in the USA, Buac et al. (2014) found a significant effect of SES in the majority language, but not in the minority language. Cobo-Lewis et al. (2002a, 2002b) found clear effects of SES in the majority language, and to some extent also in the minority language. Similarly, no effect of SES was found for the minority language skills of Turkish-Dutch children in the Netherlands (Prevoo et al., 2014), but high SES was associated with better vocabulary scores in the majority language. For Swedish bilinguals with Turkish as their home language Bohnacker et al. (2016) found no effect of SES.

4.1.4 Vocabulary in Swedish-speaking monolingual and bilingual children with a home language other than Arabic

In the Swedish context, there are few published studies on vocabulary skills in monolinguals or bilinguals. A number of BA and MA theses in speech-language pathology have investigated vocabulary in monolingual and bilingual Swedish-speaking children. Some of these will be mentioned in the next paragraph. The rest of the section contains an overview of published articles and doctoral theses on vocabulary in Swedish monolinguals and bilinguals. Concerning Arabic-Swedish-speaking bilinguals in particular, a handful of BA and MA theses and published studies exist that will be further addressed in the next section.

Westlin and Ytterdal (2007) investigated vocabulary comprehension and production in 28 children age 3–4, using a Swedish translation of the PPVT and the BNT as translated by Tallberg (2005), and found age effects for both tasks. Krüger-Wahlqvist (2012) investigated rapid naming in a group of 6–10-year-old Swedish monolinguals using the Swedish test Ordracet (Eklund, 1996), and found age effects for both accuracy and naming speed. Karner and Mattsin (2017) investigated vocabulary skills in a group of 172 monolingual children in grade 1–9 (age 7–15), using the PPVT in a Swedish translation (Ahlström & Ljungman, 2011). Raw scores increased with age. Furthermore, the Swedish translation did not match the difficulty level of the original task; the Swedish children had higher raw scores compared to the North American norming sample.

Brusewitz and Tallberg (2010) investigated vocabulary production in 152 monolinguals school-age children (ages 6–15) using the BNT. There were four groups: 6-year-olds, 9-year-olds, 12-year-olds and 15-year-olds. The scores generally increased with age, but there were significant gains in scores only between age groups 6 and 9, and 12 and 15.

Ganuza and Hedman (2019) investigated the impact of mother tongue instruction on children's vocabulary knowledge and reading skills in 120

Somali-Swedish bilinguals in grades 1–6 (between the ages of 6–12). They used translated versions of the PPVT in Swedish (Ahlström & Ljungman, 2011), and Somali (developed for the purpose of the study), as well as tasks targeting lexical depth (synonyms, antonyms and hypernyms) in both languages. The children in the younger school years (grade 1–3) scored equally well on the vocabulary tasks in both Swedish and Somali, whereas the children in the older school years (grade 4–6) scored consistently better in Swedish than Somali. A longitudinal follow-up was conducted with 46 children from the original sample one year after the first testing. The analyses showed that while the participants showed clear gains in both vocabulary comprehension (PPVT) and vocabulary depth (synonyms, antonyms and hypernyms) in Swedish, for Somali there was a significant gain only in vocabulary depth.

Lindgren (2018) investigated vocabulary production using CLTs in 72 monolingual, 46 German-Swedish and 48 Turkish-Swedish-speaking children age 4–6. All children were assessed on Swedish CLT vocabulary production, and the bilinguals were also assessed on CLT vocabulary production in their respective minority language (German or Turkish). This study comprised all children in the German sample in Bohnacker et al. (2016), as well as a subsample of the Turkish-speaking group in Öztekin (2019). While an age effect was seen for all groups in Swedish, there was no age effect in the home language for neither the German-speaking nor the Turkish-speaking group.[39]

Öztekin (2019) investigated the effect of age, SES and different measures of language input on vocabulary comprehension and production in 102 Turkish-Swedish-speaking bilinguals (age 4;0–8;1, including all children from Bohnacker et al. (2016)). The children were tested on the Turkish and Swedish versions of the CLTs. In this sample, which was larger and had a wider age range compared to Lindgren (2018), the children increased their scores in vocabulary comprehension and production in *both* languages with age, even though the age effect was stronger for the majority language Swedish compared to the home language Turkish. There were positive, weak correlations between age and vocabulary skills in Turkish (for comprehension $p < .05$, and for production approaching significance). Öztekin also investigated differences between age groups. For minority language comprehension, there was only a significant difference between the 4-year-olds and the 7-year-olds. For production, there were no significant differences between the age groups. For the majority language Swedish, there were moderate to strong positive correlations between age and vocabulary skills for both comprehension and production. The stronger age effects in Swedish were also reflected when comparing the performance between age groups. For comprehension, the 4-year-olds performed significantly lower than five-, six-

[39] Lindgren and Bohnacker (2020) report the same pattern also for vocabulary comprehension in the German-Swedish-speaking children.

and 7-year-olds. The 5-year-olds performed significantly lower than the 6-year-olds, but there was no difference between 6-year-olds and 7-year-olds. For production, the 4-year-olds scored lower than the six- and the 7-year-olds, and the 5-year-olds also scored significantly lower than the six- and the 7-year-olds. Although SES as measured by parental education was not related to vocabulary scores in either language, Öztekin did find effects of input. Children with an estimated daily exposure of 80% or more Swedish during the day performed significantly better on Swedish vocabulary production, and marginally so for comprehension. Furthermore, there was a significant effect of parental language use. Children whose parents spoke only or mostly Turkish to them had significantly higher scores in Turkish vocabulary comprehension and production than children with other input patterns.

4.1.5 Vocabulary in Arabic-speaking bilingual children

In this section, studies of vocabulary skills in Arabic-speaking bilinguals will be summarised. These include BA and MA theses (Alkass Yousef & Bergström, 2011; Mikoczy & Nyman, 2008; Petersen & Wail Yassin, 2012) as well as published studies on vocabulary skills in Arabic-Swedish-speaking bilinguals, and a study on Arabic-speaking bilinguals in Lebanon. Many of the studies investigating vocabulary skills in Arabic-Swedish-speaking bilinguals make use of the Kent and Rosanoff word association task.[40]

Mikoczy and Nyman (2008) investigated vocabulary size (using the PPVT) and organisation (using the Kent and Rosanoff word association list) in the majority language Swedish of 49 Arabic-Swedish-speaking bilinguals in grade 4 (age 10–11). They found that the Arabic-speaking children in their sample scored lower on the PPVT compared to findings from earlier studies concerning bilinguals in Sweden. Alkass Yousef and Bergström (2011) investigated the vocabulary skills of 16 children age 6;2 to 7;0. The children were tested on vocabulary comprehension and production in Swedish with the PPVT and the Swedish naming task Ordracet. The bilinguals were also tested on both tasks in their home language, with Arabic translations developed by the first author. Alkass Yousef and Bergström found that the bilinguals scored on par with Swedish monolinguals in Swedish, and they did better in Swedish than Arabic in both comprehension and production. Petersen and Wail Yassin (2012) investigated vocabulary size (a vocabulary test probing school

[40] As mentioned in Chapter 1, this task was originally developed for the assessment of psychiatric disorders in adults in the early 20th century (Kent & Rosanoff, 1910), but has since been used in studies in linguistics (Namei, 2004), and has recently caught attention in Sweden as an assessment tool of bilingual children's vocabulary skills (Holmström, 2015; Salameh, 2003, 2011). The task entails hearing a word, and making an association to the first word one comes to think of. Holmström (2015) and Salameh (2003, 2011) classify the associations as sound-based, syntagmatic, paradigmatic or 'other', where paradigmatic associations are considered to be most advanced.

vocabulary), organisation (the Kent and Rosanoff list), verbal fluency and verbal analogies in 36 Arabic-Swedish-speaking bilinguals (16 with a DLD diagnosis, 20 without a DLD diagnosis) between the ages of 6–9. They found that the children with DLD scored consistently lower than the TD children on all vocabulary measures.

As already mentioned in Chapter 1, Salameh (2011) investigated the effect of bilingual instruction on vocabulary size (Swedish and Arabic versions of the PPVT, developed for the study) and lexical organisation (the Kent and Rosanoff list) in Arabic-Swedish-speaking pupils in grade four (10–11-year-olds). The experiment group consisted of 16 bilingual children who had had bilingual Arabic-Swedish instruction in school since first grade, and they were compared to a control group of 33 age-matched children from the same ethnically diverse area in Sweden, but who had received monolingual schooling. All children were reported to have typical language development. Salameh found that the pupils who received monolingual instruction had higher PPVT scores in Swedish (but this difference was not significant). Furthermore, the children who received bilingual instruction scored significantly higher on the PPVT in Arabic than the children in the control group.

Another study that was also mentioned in Chapter 1 is Holmström (2015), who investigated the vocabulary skills of 88 monolingual and bilingual Arabic-Swedish-speaking children with and without DLD between the ages of 5;11–9;3 (including the participants from Petersen and Wail Yassin (2012)). Twenty-four children had a DLD diagnosis (15 bilingual, 9 monolingual) and 64 children had typical language development (29 bilingual, 35 monolingual). Vocabulary skills were assessed for comprehension (a task probing school-related words), production (a task developed for the study probing nouns) and lexical organisation (the Kent and Rosanoff task). Holmström found that the combined measurement of vocabulary comprehension and production showed the highest degree of separation between the TD and the DLD group. Moreover, the bilingual children (both with and without a DLD diagnosis) scored better than the monolingual children on the Kent and Rosanoff task. Holmström speculated that the bilingual group might include a higher proportion of children misclassified as having DLD compared to the monolingual group, and that the scores in the bilingual group were therefore higher. However, Holmström does not discuss the fact that the bilinguals were tested twice (once in each language) and the combined score from both languages were compared to the monolinguals. Thus, the bilinguals had two chances to receive a score on each item, which may have contributed to their higher scores.

Finally, Khoury Aouad Saliby, dos Santos, Kouba Hreich and Messarra (2017) used the Lebanese version of the CLT to compare the performance of 32 TD children (age range 5;7 to 6;9) to 10 children with a DLD diagnosis (age range 5;9 to 7;10). In Lebanon, virtually all children become bilingual at

a very young age, being exposed to French or English in addition to Lebanese Arabic. The authors used conceptual scoring in the expressive task.[41] The experimenter always spoke Arabic during the sessions. For comprehension, all prompts were subsequently in Arabic. For production, the prompts were also in Arabic, but any accurate answers in either English or French were scored as correct. The TD group scored consistently higher than the DLD group on production of nouns and verbs as well as comprehension of verbs (noun comprehension was close to ceiling for both groups). Moreover, all children scored better on nouns than verbs (in comprehension and production), mirroring the results of Haman et al. (2017), who found a significant effect of word class in 13 of the 16 languages that were investigated for word class effects in their study.

In sum, although there are previous studies that investigate the vocabulary skills of Arabic-Swedish-speaking bilinguals, sample sizes are typically small, and when both languages are assessed, translated tests are often used. Large-scale studies that investigate the relationship between age, SES, language input and vocabulary skills in both languages of Arabic-Swedish-speaking children with typical language development are lacking. The present study uses CLTs to investigate vocabulary skills in both languages of Arabic-Swedish-speaking 4–7-year-old children. Furthermore, the relationship between age, SES, language input and vocabulary skills is also explored.

4.2 Scoring of the CLTs

For each child, the responses on the Cross-linguistic Lexical Tasks were scored and coded for comprehension and production in Arabic and Swedish. The maximum score for CLT comprehension was 60 points in each language. For each correct answer, one point was awarded. The author of the present study, a native speaker of Swedish, did the scoring in Swedish. For Arabic, Rima Haddad, PhD student of linguistics and native speaker of Lebanese Arabic, conducted the scoring, in collaboration with Ute Bohnacker, professor of linguistics, and Annette Månsson, lecturer in Semitic languages. Prior to conducting the scoring for the current study, extensive discussions and consensus rounds were carried out within the BiLI-TAS group, in order to ensure that scoring was done in a homogeneous way. These principles were summarised in a comprehensive document containing scoring guidelines, which was used when scoring the CLTs for the current study (Guidelines for scoring CLT, version March 2018).[42]

[41] Conceptual scoring is a method of assessing vocabulary skills in bilinguals where answers in any of the two languages are scored as correct (Bedore, Peña, García, & Cortez, 2005).

[42] The BiLI-TAS research team at that point consisted of Ute Bohnacker (Professor of Linguistics and PI of the project), Sibylle Dillström (SLP; PhD in German; researcher), Josefin Lindgren (PhD student of Linguistics) and Buket Öztekin (SLP; PhD student of Linguistics).

For comprehension, scoring was straightforward, since only responses selecting the correct picture were scored as correct. When scoring the production parts of the Swedish CLTs, a handful of unclear cases emerged. These were discussed among the members of the BiLI-TAS team until consensus was reached. For Arabic, scoring the CLTs involved a more complicated procedure, since the children spoke different varieties of Arabic and thus there was more variation in the number of lexical labels that could be considered correct. The process of scoring the Arabic production parts was conducted by Rima Haddad, and included discussions with Ute Bohnacker and Anette Månsson, as well as consulting experts and speakers of Arabic and in some cases, parents of the children who took part in the study.

The following principles were followed when scoring production. Apart from the target response, some additional responses were also scored as correct:

- deviant or slightly off-target pronunciation due to dialectal variation, cross-linguistic influence or age-appropriate developmental phonological processes (e.g. *sringer* instead of *springer* 'running' in Swedish, and *yakhbot* instead of *yatbukh* 'to cook' in Arabic),
- adult-like synonyms of the target word (e.g. *meta* 'angling' instead of *fiska* 'fishing' in Swedish),
- alternative, adult-like interpretations of the picture not corresponding to the target word (e.g. *warda* 'rose' instead of *zahra* 'flower' in Arabic),
- errors of grammatical gender (e.g. *ett banan* instead of *en banan* 'a banana' in Swedish),
- marking of tense, finiteness, person and plural was disregarded, as long as the child used the target lemma (e.g. both *smält* and *smälter*, 'melted' and 'melting', were scored as correct in Swedish),
- non-target but more specific words corresponding to the picture (e.g. *champinjon* 'champignon' instead of *svamp* 'mushroom' in Swedish)

Responses scored as incorrect were:

- words belonging to a different word class (e.g. *snö* 'snow' instead of *snöa* 'snowing' in Swedish),[43]
- words corresponding to the picture which were less specific than the target word (e.g. *ya3mel sura* 'make a picture' instead of *sawwar* 'to photograph' in Arabic),
- responses in another language (e.g. *strumpa* 'sock' in Swedish instead of *kalseet* 'sock' in Arabic),
- responses that were too phonologically deviant from the target word to be recognisable (e.g. *brebåge* instead of *regnbåge* 'rainbow' in Swedish)

[43] Recall that the prompt clearly queried a verb: *Vad händer här?* (What happens here?)

All participating children completed all sub-tasks of the CLT in both languages. The total number of responses was 23 760 (99 children x 2 languages x 120 test items, 60 for comprehension and 60 for production).

4.3 Results

The results of the study are presented in two sections. First, vocabulary scores and age are analysed in section 4.3.1, followed by vocabulary scores in relation to language input and socioeconomic status (SES) in section 4.3.2.

4.3.1 Vocabulary scores and age

This section starts with a description of the statistical analyses (4.3.1.1). Next, CLT scores are presented (4.3.1.2), for Arabic (4.3.1.3) and Swedish (4.3.1.4) in relation to age. Finally, a comparison will be made between Arabic and Swedish CLT scores (4.3.1.5).

The following specific research questions are investigated concerning vocabulary and age:

- Is there a difference between vocabulary scores in the two languages?
- How do vocabulary scores develop with age in Arabic and Swedish?
- What is the relationship between production scores and comprehension scores in each language?

4.3.1.1 Statistical analyses

The two languages were analysed separately for comprehension and production. Both the comprehension and the production task had 60 items in total, with 30 nouns and 30 verbs in each task.

Arabic and Swedish CLT total scores, comprehension scores and production scores were compared with paired samples t-tests. Next, age effects for comprehension scores and production scores were investigated by using one-way ANOVAs for age groups, and a Pearson correlation for age in months for Arabic and Swedish respectively. The relationship between comprehension and production was analysed for each language with a Pearson correlation. For all statistical analyses, the level of significance was consistently set to $p < .05$ (two-tailed). The analyses were made in R (R Core Team, 2020).

4.3.1.2 Arabic and Swedish CLT scores

In this section, the total scores, comprehension scores and production scores are reported for the Arabic and Swedish CLT.

Table 4.1. All ages combined (4;0–7;11, N=99) vocabulary (CLT) total, comprehension and production scores for Arabic and Swedish. Total score max 120 points, comprehension score max 60 points, and production score max 60 points.

	Arabic	Swedish
Total scores		
Mean	80.17	76.65
SD	18.79	21.94
Range	30–111	34–113
Comprehension		
Mean	47.47	45.66
SD	7.47	10.78
Range	25–59	18–60
Production		
Mean	32.70	30.99
SD	12.35	11.66
Range	1–53	10–53

There was no statistically significant difference between the total scores on the Arabic and the Swedish CLT ($t(98) = –1.25$, $p = .216$, $d = .17$. The same result was found when analysing the scores separately for comprehension and production; there was no significant difference between Arabic and Swedish scores (comprehension: ($t(98) = –1.62$, $p = .11$, $d = .20$), production: ($t(98) = –0.95$, $p = .35$, $d = .14$)). Comprehension scores were higher than production scores in both languages (Arabic: ($t(98) = 18.32$, $p < .001$, $d = 1.45$); Swedish: ($t(98) = 30.39$, $p < .001$, $d = 1.31$).

As displayed in Table 4.1, the ranges for the total scores are similar for Arabic and Swedish. However, when inspecting the ranges for comprehension and production separately, the patterns are different within each language. The range for Swedish comprehension is wider than Arabic comprehension, but for production the pattern is reversed, with a wider range in Arabic than in Swedish.[44] The top score was obtained for comprehension in Swedish, but not in Arabic (maximum 60 points).

The standard deviation (SD) for total scores is narrower for Arabic than for Swedish. Looking at the SDs separately for comprehension and production, the SD is larger for comprehension in Swedish than in Arabic, suggesting that on group level, the children scored more similarly to each other in Arabic comprehension than in Swedish comprehension. SDs for production scores

[44] However, the wider range in Arabic production compared to Swedish is due to one individual who performs very low, as can be seen in Figure 4.2.

were higher than SDs for comprehension in both languages, reflecting a more varied performance in production than comprehension. Notably however, the difference in SDs between comprehension and production was substantially larger in Arabic (4.88) than in Swedish (0.88). In what follows, total scores will not be reported, but only comprehension scores and production scores separately.

4.3.1.3 Arabic CLT scores and age

In this section, Arabic CLT comprehension scores and production scores are reported in relation to age. Age development is investigated via age groups in ANOVAs, and also linearly, including scatterplots showing results for individual children, and via Pearson correlation with age in months.

Table 4.2. Descriptive statistics for Arabic vocabulary scores (CLT) for each age group: comprehension and production. Comprehension score max 60 points, and production score max 60 points.

	4-year-olds N=23	**5-year-olds N=23**	**6-year-olds N=29**	**7-year-olds N=24**
Comprehension				
Mean	41.70	46.78	48.52	52.42
SD	7.50	6.76	7.51	3.62
Range	25–52	27–56	31–58	45–59
Production				
Mean	25.00	33.52	34.48	37.13
SD	12.16	11.46	13.24	9.22
Range	1–42	11–48	10–53	16–52

As presented in Table 4.2, mean scores for Arabic comprehension increased with age. A one-way ANOVA confirmed this pattern, showing a significant difference in scores between the age groups ($F(3,95) = 10.818$, $p < .001$, $\eta_p^2 = .255$). Post hoc analyses (with Bonferroni correction) revealed a significant difference between the 4-year-olds and the two older age groups (4-year-olds vs. 6-year-olds: $p < .01$; 4-year-olds vs. 7-year-olds: $p < .001$), and between the 5-year-olds and the 7-year-olds ($p < .05$). No other groups differed significantly from each other. The lowest score was found in the youngest age group, and the highest score was found in the oldest age group. Variation was the smallest among the 7-year-olds, as is reflected in Table 4.2 both in narrower range and smaller SD compared to the younger age groups.

As for Arabic production, mean scores also increased with age. A one-way ANOVA revealed a significant difference in scores between the age groups ($F(3,95) = 4.033$, $p > .01$, $\eta_p^2 = .113$). However, a post hoc analysis (with Bonferroni correction) revealed that there was only a significant difference between the 4-year-olds and the 7-year-olds ($p < .01$). No other groups

differed significantly from each other. The lowest score was found in the youngest age group, and the highest score was found among the 6-year-olds. Variation was similar across groups.

Figure 4.1 shows the Arabic comprehension scores plotted against age in months. As previously mentioned, there was a significant difference in scores between the 4-year-olds and the two oldest age groups, and between the 5-year-olds and the 7-year-olds. A Pearson correlation revealed that there was a moderate positive correlation between Arabic comprehension score and age in months (N = 99, $r = .497$, $p < .001$). The majority of the children (77.8%) scored between 50–90% correct. Four children, three 4-year-olds and one 5-year-old (4.0%), scored below 50%. Twenty-two children (22.2%), mainly 7-year-olds and 6-year-olds, performed at 90% or better. No 4-year-olds performed 90% or better on Arabic comprehension.

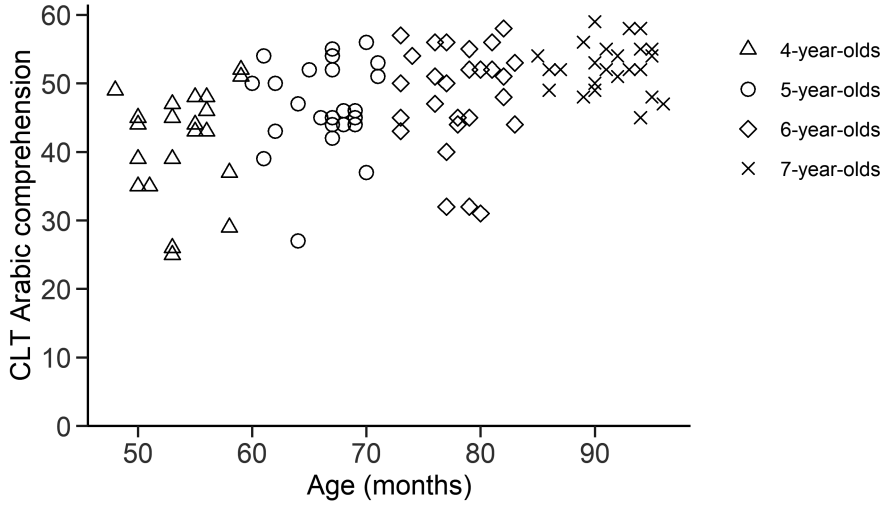

Figure 4.1. Scatterplot of CLT Arabic comprehension scores and age in months.

Now we turn to the lowest performing children for Arabic comprehension in each age group. For the 4-year-olds these children were BiAra4-02, BiAra4-15 and BiAra4-16. For BiAra4-02, one possible reason for the low scores may be test behaviour. The child had a hard time cooperating during all sessions, and scored low also in Arabic production. BiAra4-14 is a child whose parents report the majority (60%) of the input being in Swedish, and the child performs close to the mean in Swedish. BiAra4-16 has slightly better scores in Swedish compared to Arabic. Although the parents do not report exposure to or use of English, the child used many English words in the production parts of both the Arabic and the Swedish CLT. It is thus likely that there is exposure to English that may explain the low comprehension scores in Arabic. Turning to the 5-year-olds, there was one child, BiAra5-06, who can be considered as

an outlier compared to the group, scoring almost three SDs below the mean. The parents report that the child is exposed predominately to Swedish (80%), and the child scores close to the mean in Swedish comprehension. Among the 6-year-olds, three children perform much lower than the others, close to 50% (30 points). Of these three children, two of them (BiAra6-10 and BiAra6-11) have above mean performance in Swedish comprehension, and are also reported to have the majority of the daily exposure in Swedish (60% and 80% respectively). BiAra6-26 has low scores not only in Arabic comprehension, but also in Arabic production as well as in Swedish in both modalities. This child had trouble cooperating during both sessions, and had a brother who had a DLD diagnosis (BiAraLI-01). Among the 7-year-olds the scores were more homogeneous and no individuals were clearly performing much lower than the group as a whole.

Figure 4.2 shows the Arabic production scores plotted against age in months. As previously mentioned, there was a significant difference in scores between the youngest and the oldest age group. A Pearson correlation showed that there was a moderate positive correlation between Arabic comprehension score and age (N = 99, r = .326, p = .001). The majority of the children, 61.6%, performed between 50% and 90% on Arabic production. At the same time, a substantial group (38.4%) performed 50% or lower. There were children from all age groups performing at 50% or lower. No children performed at 90% or better.

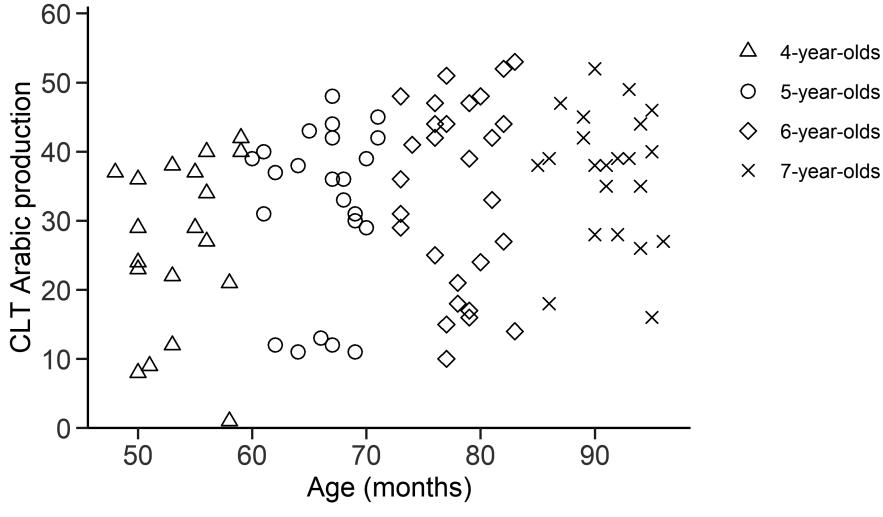

Figure 4.2. Scatterplot of CLT Arabic production scores and age in months.

Let us now look at some low performers in each age group. Among the 4-year-olds, a subgroup of six children performed very low, below 20 points (33%). The lowest performing child, BiAra4-02 (scoring only one point), was not fully cooperating during testing, and has already been discussed above (with respect to the Arabic comprehension scores). Some of the low-scoring 4-year-olds (BiAra4-09, BiAra4-14, BiAra4-15 and BiAra4-21) are reported to have the majority of their exposure in Swedish (60–80%), and score close to average or far above average in Swedish comprehension. BiAra4-16 was previously discussed, and received low scores also in Arabic production. In the 5-year-old group, there is a subgroup of five children (BiAra5-01, BiAra5-06, BiAra5-12, BiAra5-16 and BiAra5-25) performing very low compared to the group as a whole. The parents of these 5-year-olds report that the majority of the children's daily exposure is in Swedish (i.e. 60% or more), and all children score close to or above the mean on Swedish production. For the 6-year-olds there is no clear subgroup performing much lower than the rest, but rather two separate groups: one clustering low and another high. All children in the low-scoring group (BiAra5-01, BiAra5-06, BiAra5-12, BiAra5-16 and BiAra5-25) had less daily exposure to Arabic than Swedish. Among the 7-year-olds, two children (BiAra7-12 and BiAra7-13), stand out as very low performing compared to the rest of the group. Both however score more than one SD above the mean for Swedish comprehension. The parents of BiAra7-12 report that the overwhelming majority of the daily language exposure is in Swedish (95%), while the daily exposure pattern for BiAra7-13 is balanced (50-50). In the 7-year-old group, there is also a low-mid-scoring segment of four children. Three of these score more than one SD above the mean for Swedish production (BiAra7-02, BiAra7-03 and BiAra7-14), and have equal exposure patterns (BiAra7-02), or slightly more Swedish (BiAra7-03 and BiAra7-14). Finally, BiAra7-17 is also in this low-mid-scoring segment. His scores for Swedish production are also low, at –1.7 SD below the mean (and scoring below 1 SD in both comprehension and production in both languages). When investigating the background information of this child, exposure patterns may explain the low scores. The parents report that the child, apart from Arabic and Swedish, is also exposed to and uses a lot of English. The child used many English words in the production parts of both the Arabic and the Swedish CLT.

Next, the relationship between Arabic comprehension scores and Arabic production scores will be investigated. A Pearson correlation showed that there was a strong positive correlation between comprehension scores and production scores in Arabic ($N = 99$, $r = .779$, $p < .001$). This pattern is evident in Figure 4.3, where comprehension and production scores are plotted against each other for individual children.

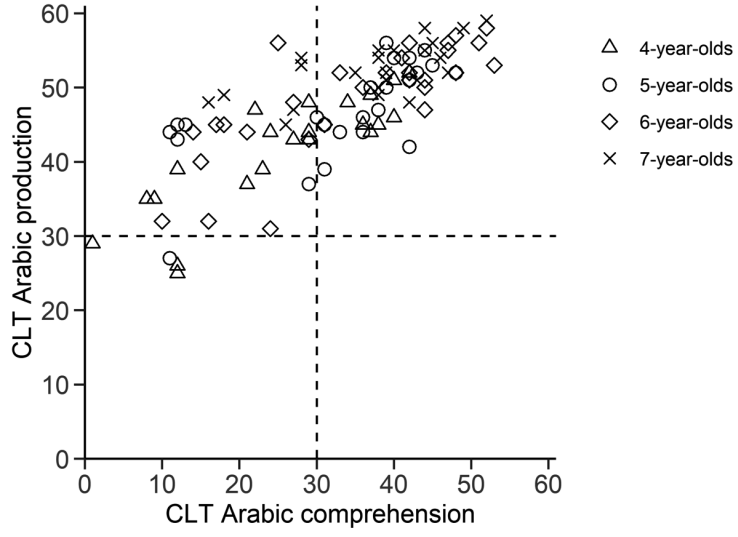

Figure 4.3. Scatterplot of CLT Arabic comprehension scores (x-axis) and CLT Arabic production scores (y-axis). Dashed lines mark 30 points (50%).

As shown in Figure 4.3, many children are found in the upper right quadrant, meaning that they scored better than 50% on both Arabic CLT comprehension and production. In fact, 14 children are in the upper half of the top right quadrant, scoring at or above 75% in both comprehension and production. Four children are represented in the lower left quadrant (scoring below 50% on both comprehension and production), three 4-year-olds (BiAra4-02, BiAra4-15, and BiAra4-16) and one 5-year-old (BiAra5-06). All children have been discussed previously. As for BiAra4-02, this child was not cooperating during testing, and it is likely that the results do not reflect the child's true abilities. BiAra4-15 has more exposure to Swedish than Arabic, and also scores close to the mean in Swedish comprehension and production. BiAra4-16 scored within 1 SD below the mean for both Swedish comprehension and production. BiAra5-06 has the majority of the daily exposure in Swedish (80%), and also scores close to the mean for the age group in both Swedish comprehension and production.

4.3.1.4 Swedish CLT scores and age

Table 4.3. Descriptive statistics for Swedish vocabulary scores (CLT) for each age group: comprehension and production. Comprehension score max 60 points, and production score max 60 points.

	4-year-olds N=23	5-year-olds N=23	6-year-olds N=29	7-year-olds N=24
Comprehension				
Mean	36.57	45.13	46.93	53.33
SD	8.48	9.22	10.18	8.52
Range	18–52	29–59	27–60	27–60
Production				
Mean	22.57	29.78	31.62	39.46
SD	7.15	9.98	11.50	11.25
Range	10–41	15–48	11–48	12–53

As presented in Table 4.3, mean scores for Swedish comprehension increased with age. A one-way ANOVA confirmed this pattern, showing a significant difference in scores between the age groups ($F(3,95) = 14.026$, $p < 0.001$, $\eta_p^2 = .307$). Post hoc analyses (with Bonferroni correction) revealed a significant difference between the 4-year-olds and the older age groups, as well as between the 5-year-olds and the 7-year-olds (4-year-olds vs. 5-year-olds: $p < .01$, 4-year-olds vs. 6-year-olds: $p < .001$, 4-year-olds vs. 7-year-olds: $p < .001$, 5-year-olds vs. 7-year-olds: $p < .05$). The lowest score was found in the youngest age group and the highest score was found in the two oldest age groups, with participants among both the 6-year-olds and the 7-year-olds reaching the maximum score. As shown in Table 4.3, the ranges (highest and lowest score) were similar among the three oldest age groups, but both the lowest and the highest score was lower among the 4-year-olds. However, standard deviations were similar across all age groups.

As for Swedish production, mean scores also increased with age. A one-way ANOVA revealed a significant difference in scores between the age groups ($F(3,95) = 10.962$, $p < .001$, $\eta_p^2 = .257$). Post hoc analyses (with Bonferroni correction) showed that there were significant differences between the 4-year-olds and the two oldest age groups, as well as between the 7-year-olds and the three younger age groups (4-year-olds vs. 6-year-olds: $p < .05$, 4-year-olds vs. 7-year-olds: $p < .001$, 5-year-olds vs. 7-year-olds: $p < .01$, 6-year-olds vs. 7-year-olds: $p < .05$).

The lowest score was found in the youngest age group, and the highest score in the oldest age group. There were children among both the 6-year-olds and the 7-year-olds who scored lower than the lowest performing child in the 5-year-old-group. The SD was slightly smaller in the youngest age group compared to the two oldest ones.

Figure 4.4 shows the Swedish comprehension scores plotted against age in months. As previously mentioned, there was a significant difference in scores between all age groups, except between the 5-year-olds and the 6-year-olds. A Pearson correlation showed that there was a moderate positive correlation between Swedish comprehension score and age (r=.514, N = 99, p<.001).

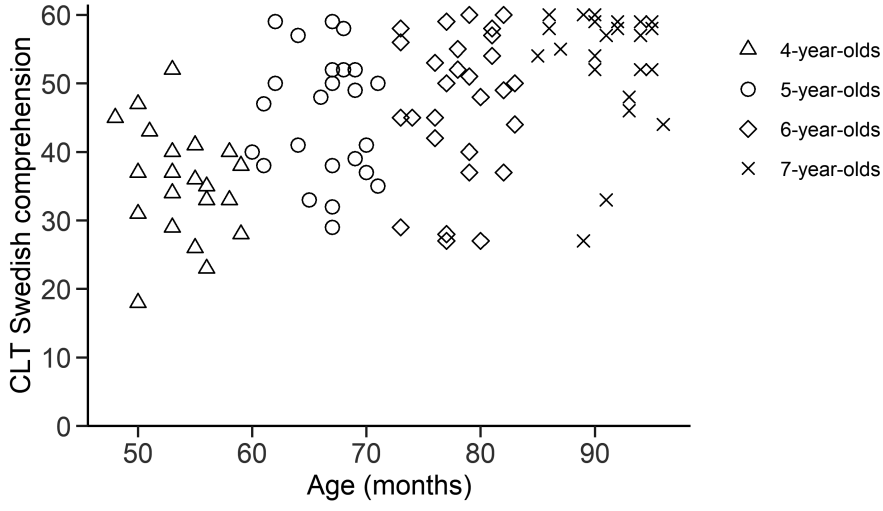

Figure 4.4. Scatterplot of CLT Swedish comprehension scores and age in months.

The lowest performing 4-year-old (BiAra4-12) was the only one scoring less than 20 points (33%) on Swedish comprehension. The child is reported to have the majority (80%) of the daily input in Arabic, and performs at the mean for Arabic comprehension. In the 5-year-old-group the results were rather homogeneous, with no group of children performing much higher or lower than their peers. Only one child, BiAra5-05, scored below 30 points (50%), and was reported to have little daily exposure to Swedish (20%). Four 5-year-olds (BiAra5-09, BiAra5-11, BiAra5-12 and BiAra5-25) scored above 54 points (90%). All of them were reported to have predominately Swedish in their daily language input. Among the 6-year-olds there is a sub-group of four children (BiAra6-20, BiAra6-25, BiAra6-26 and BiAra6-29) who perform lower than their peers. Remember that BiAra6-29 scored low in Arabic and Swedish in both modalities, and had a sibling with DLD. The rest of the low-scoring 6-year-olds all have better scores in Arabic comprehension compared to Swedish. Two 7-year-olds (BiAra7-05 and BiAra7-16) stand out as performing much lower than their peers. There is no background information about exposure patterns or language development available from the parents of BiAra7-05. However, the child mentioned during the session that he had relatively recently arrived in Sweden. BiAra7-16 had also recently arrived in

Sweden and only been exposed to Swedish for about six months, which explains the low scores in Swedish comprehension.

Figure 4.5 shows the Swedish production scores plotted against age in months. As previously mentioned, there was a significant difference in mean scores between all age groups, except for the 5-year-olds, who differed significantly only from the 7-year-olds. A Pearson correlation revealed that there was a moderate positive correlation between Swedish comprehension scores and age (N = 99, $r = .463$, $p < .001$).

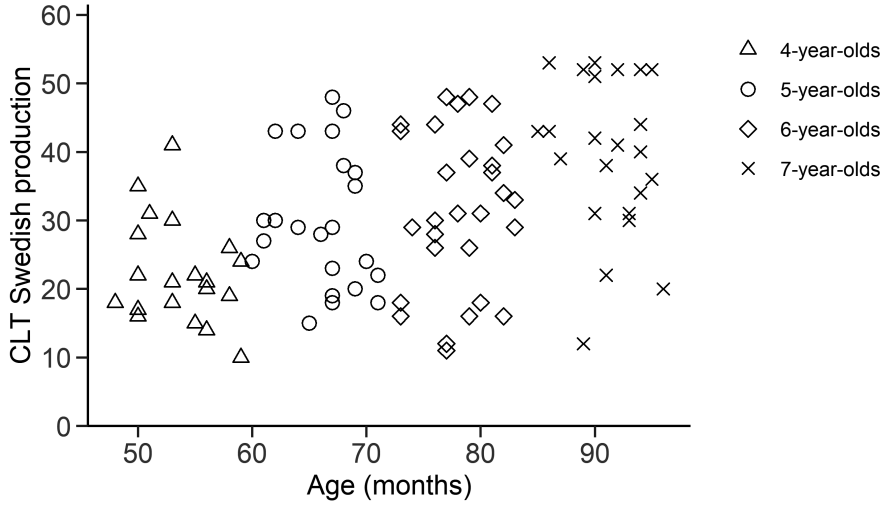

Figure 4.5. Scatterplot of CLT Swedish production scores and age in months.

The lowest scoring 4-year-old (BiAra4-08) received only ten points (17%) on Swedish production. The parents report that the majority of the exposure was in Arabic (80%) and that onset to Swedish was at age four. In the 6-year-old group there was a rather large subgroup of seven children that performed lower than the others. Some of these performed within one SD above the mean on Arabic production (BiAra6-06, BiAra6-15, BiAra6-20), or higher (BiAra6-25). BiAra6-25 was reported to have a late age of onset of Swedish (between age 4 and 5), and is clearly dominant in Arabic. BiAra6-26 has been mentioned before, scoring low on all CLT tasks in both languages, and having a sibling with a DLD diagnosis. However, he was also reported to have a relativley late age of onset to Swedish, between age 4 and 5. Regarding the 7-year-olds, three children (BiAra7-05, BiAra7-16 and BiAra7-17) perform much lower than the rest. The background characteristics of these children have been previously described: BiAra7-05 had a self-reported late age of onset of Swedish (parental reports were not available). BiAra7-16 also had a late age of onset of Swedish, while BiAra7-17 was exposed to English in

addition to Arabic and Swedish. In fact, many words in the Swedish production task were in English, or pronounced very English-like.

Next, the relationship between Swedish CLT comprehension scores and Swedish production scores will be explored. A Pearson correlation showed that there was a very strong positive correlation between comprehension scores and production scores in Swedish (N = 99, r = .911, p < .001). This pattern is evident in Figure 4.6, where comprehension and production scores are plotted against each other for individual children. Children in the upper right quadrant scored 50% or better in both comprehension and production. Children from all age groups are found here, but many of them were 7-year-olds.

As can be seen in Figure 4.6, there are a handful of children performing below 50% in both Swedish CLT comprehension and production (lower left quadrant). These are mainly 4-year-olds, but at least one child from each age group is represented here.

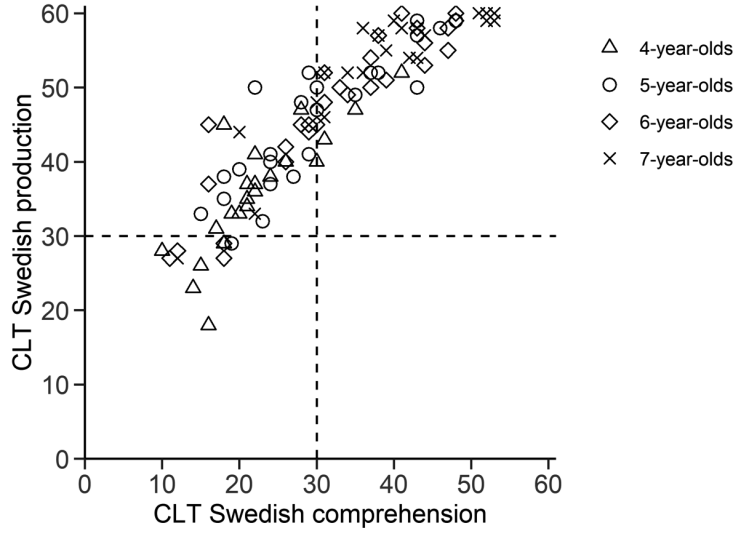

Figure 4.6. Scatterplot of CLT Swedish comprehension scores (x-axis) and CLT Swedish production scores (y-axis). Dashed lines mark 30 points (50%).

Of the 4-year-olds in the lower left quadrant, most of them have the majority (60–80%) of their daily input in Arabic (BiAra4-08, BiAra4-12 and BiAra4-16), while BiAra4-03 was reported to have a very recent age of onset to Swedish, starting in the fifth year of life. The only 5-year-old in this group, BiAra5-05, was reported to have a recent age of onset of Swedish (in the sixth year of life), as well as a majority of the daily exposure in Arabic (80%). All 6-year-olds in the lower left quadrant have been mentioned earlier (BiAra6-20, BiAra6-25, BiAra6-26 and BiAra6-29), all having a recent age of onset of

Swedish (starting at the fifth or sixth year of life). Remember also that BiAra6-26 had a sibling with diagnosed DLD. The only 7-year-old in this group was BiAra7-05, for which background information is not available, apart from a self-reported late age of onset of Swedish.

4.3.1.5 Comparison of age development of Arabic and Swedish comprehension and production

In this section, a comparison will be made of the group results in comprehension and production in the two languages over time. Figure 4.7 shows the mean scores of vocabulary comprehension and production in both Arabic and Swedish for each age group.[45]

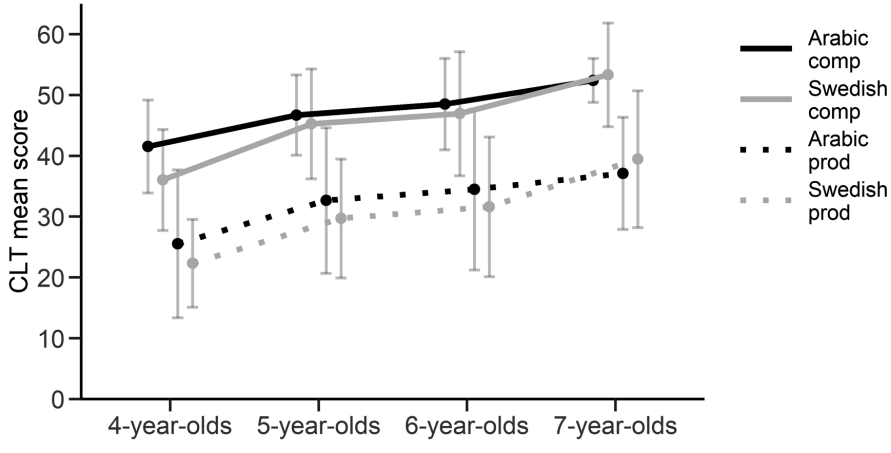

Figure 4.7. Line chart of CLT mean scores by age groups. Error bars indicate +/- 1 SD.

As can be seen in Figure 4.7, comprehension scores are higher than production scores in both languages and all age groups. The distance between production and comprehension also stays constant for both languages in all age groups. The most notable change in scores is between the lowest and the highest age group for both comprehension and production in both languages. Means increase between age four and five, stay almost constant between age groups five and six, and then increase again at age seven. Notably, mean scores for both comprehension and production are higher for Arabic in age groups four, five and six, but at age seven, the mean scores in Swedish are the same or slightly higher than the Arabic mean scores. Overall, this means that development over age is steeper for both comprehension and production in

[45] However, it is important to keep in mind that this is a cross-sectional study. This means that differences between the age groups cannot be attributed to age development alone, but it is likely that differences in the distribution of the groups may also contribute to the observed differences between the age groups.

Swedish than in Arabic. Finally, there is much individual variation, as shown by the large error bars (indicating +/- SD). This pattern is particularly evident for the production scores in both languages.

4.3.1.6 Vocabulary and age: summary

To summarise, there was no difference between Arabic and Swedish total CLT scores for all age groups combined, nor was there a difference between the languages when analysing comprehension and production separately. Comprehension scores were higher than production scores in both languages, which follows the expected pattern. When analysing mean scores for the age groups separately, similar patterns emerged for both comprehension and production in Arabic and Swedish. There was an increase in the mean scores between age four and five. The mean scores stayed rather constant between 5-year-olds and 6-year-olds, but increased again at age seven.

Several children (three 7-year-olds and two 6-year-olds) reached the maximum score (60 points) in Swedish comprehension, but no child did so for Arabic comprehension (the top score for Arabic, 59 points, was reached by one 7-year-old). At the same time, the lowest score for Swedish comprehension was seven points lower than for Arabic comprehension, which reflects a more diversified performance for Swedish than for Arabic. This is also reflected in the fact that the SD is larger for Swedish comprehension than for Arabic comprehension. For production, the picture was a bit different. While the top score was the same for both languages (53 points), the lowest score for Arabic was nine points lower than for Swedish. However, for production the SD was similar for both languages, thus reflecting more similar patterns between the languages for production than for comprehension.

Most children performed above 50% on Arabic comprehension, with only four children (mainly 4-year-olds) scoring below. Twenty-two children (among them nearly half of the 7-year-olds) performed above 90%. For Swedish comprehension, eleven children performed below 50%, notably mainly four- and 6-year-olds. However, nearly one third (29) of the total group (mainly seven- and 6-year-olds), performed at 90% or above. Turning to production scores, more than one third (38) of the whole group scored below 50% on Arabic production, and more than half (52) scored below 50% on Swedish production. No child scored above 90% on either Arabic or Swedish production.

Finally, the background information of the children scoring high or low compared to their peers (discussed above) seems to suggest that input factors such as age of onset and amount of daily exposure have an impact on vocabulary scores. This pattern has been attested in many studies (as reviewed in the introduction). Previous studies have also shown that there may be an association between SES and vocabulary skills. As mentioned earlier, one objective of the present study is to investigate the relationship between input factors, SES and vocabulary scores. This will be done in the next section.

Furthermore, it is worth mentioning that children who scored low may be candidates for undetected DLD. Some of them will be discussed in the clinical chapter (Chapter 6) focusing on the children with a DLD diagnosis.[46]

4.3.2 Language input, SES and vocabulary

In this section, background factors relating to language input and SES and their relationship with Arabic and Swedish CLT scores will be explored. First, a description of how the variables were operationalised is provided in section 4.2.3.1. This is followed by a description of the statistical analyses in section 4.2.3.2. The comprehension and production scores are analysed in connection to input and SES in section 4.2.3.3 for Arabic and in section 4.2.3.4 for Swedish. Finally, the results are summarised in section 4.2.3.5.

The following research questions are investigated concerning vocabulary, input and SES:

- What is the relationship between length of exposure to Swedish and Swedish vocabulary comprehension and production?
- What is the relationship between amount of daily exposure and vocabulary comprehension and production in Arabic and Swedish?
- What is the relationship between SES and vocabulary comprehension and production in Arabic and Swedish?

4.3.2.1 Operationalisation of variables

All background information was obtained from the parental questionnaire, which is described in more detail in section 3.1 in the Methods chapter.

Parents estimated their child's daily exposure to each language on a seven-point scale, reflecting different amounts of relative exposure to each language. The possible answers were '1: Swedish 5%, Arabic 95%', '2: Swedish 20%, Arabic 80%', '3: Swedish 40%, Arabic 60%', '4: Swedish 50%, Arabic 50%', '5: Swedish 60%, Arabic 40%', '6: Swedish 80%, Arabic 20%', '7: Swedish 95%, Arabic 5%'. There was also one additional eighth category, marked 'other'. In the 'other' category it was possible for parents note a different distribution, for instance if the child received exposure to an additional language besides Arabic and Swedish. The answers were recoded into two separate variables, reporting a percentage of exposure to each language. For instance, for a child whose parents ticked '3' in the questionnaire, the variable

[46] Although it is frequently argued that DLD must manifest in both languages of bilingual children in order to qualify for a diagnosis, scoring low or high compared to age peers may not necessarily say much unless input factors such as exposure patterns are also taken into account (Goral & Conner, 2013). Therefore, only some children will be discussed, namely those whose results on the vocabulary and/or NWR tasks are unexpected in light of their background information, in particular their exposure patterns.

'Reported daily exposure to Swedish' was coded as 40% Swedish, and the variable 'Reported daily exposure to Arabic' was coded as 60% Arabic.[47]

As for the length of exposure measure, parents were asked to report during which year of life the child started to receive regular input to Arabic and to Swedish. The question was posed for Arabic and Swedish separately, and coded in two variables called age of onset to Arabic and age of onset to Swedish, respectively. There was close to no variation in the age of onset of Arabic, since all children but one started to receive exposure to Arabic from birth. Therefore, this information was not analysed further. There was, however, considerable variation between the children in the age of onset of Swedish. This variable was transformed into a new variable called 'length of exposure to Swedish'. The levels of the original age of onset of Swedish variable were recoded. This was done by subtracting the corresponding months from the child's current age in months. For example, if a child who was 53 months at the time of testing and had a reported age of onset to Swedish between age 2;0–3;0, 24 months were subtracted from the current age in months, meaning that the child's length of exposure to Swedish was 29 months.

In this study, parental education was used as a proxy for socioeconomic status (SES). In the questionnaire, parents were asked in an open-ended question to report the highest level of education for each parent. The parents' answers were coded according to the International Standard Classification of Education (ISCED) scale, which has nine levels (UNESCO Institute for Statistics, 2012). For each child, the answer was coded for each parent separately, and a combined score was then calculated in order to receive the average score for each child. For some children, information on education level was missing for one (2) or both (6) parents. Furthermore, three children lived in single parent households, and consequently information about SES was available only for one parent. This was dealt with as follows. The mean value of SES was calculated based on the available information from all parents (parent 1 and parent 2). For the whole group, the mean education level was 4.27. For children where information was missing for one parent, the mean score for the whole group was used for parent 2. Thus, the SES score for these children consisted of the mean of the reported score for parent 1 and the mean score for the whole group. For children where information about the education level was missing for both parents, the mean SES score for the sample as a whole was used.[48] For the three children who grew up in single parent households, the value for Parent 1 was used. For more information, see section 3.1.1.4 (in the Methods chapter).

[47] For children whose parents had ticked the 'other' category, 'Reported daily exposure to Swedish' and 'Reported daily exposure to Arabic' were coded with the reported percentage for each language respectively.

[48] Mean score substitution as described by Widaman (2006) was applied.

4.3.2.2 Statistical analyses

CLT scores were analysed in four separate multiple linear regression models for each language, with comprehension and production scores in Arabic as the dependent variables in the first and the second analyses, and comprehension and production scores in Swedish as the dependent variables in the third and fourth analyses. The independent variables were: daily exposure, length of exposure (this variable was only included in the analyses on Swedish vocabulary since age of onset for Arabic was at birth for all but one child), and socioeconomic status (parental education). Since there was a significant correlation between age and CLT scores in both languages, chronological age (in months) was also included in the regression analyses as a control variable. In each linear regression model, the following is reported for each variable: coefficient (Coef.), standard error (SE), standardised coefficient (Std. coef.), p-value (p), Variance Inflation Factor (VIF) and adjusted R^2. The F-statistic (F) and the corresponding p-value are included in a note. For all statistical analyses, the level of significance was consistently set to $p < .05$ (two-tailed).

For one child (BiAra7-05), no information was available regarding daily exposure levels, age of onset to Swedish and Arabic, or parental education levels. Therefore, this child is not included in the analyses, and the number of children included in these analyses is 98.

4.3.2.3 Arabic CLT scores, input and SES

In this section, the combined effects of age, input (daily exposure to Arabic) and SES on the Arabic CLT scores will be investigated separately for comprehension and production. Since age of onset for Arabic was within the first year of life for all participants (except for one child, BiAra5-25, who started being exposed to Arabic in the second year of life), length of exposure to Arabic was not included in the analysis as an input measure.

In Figure 4.8A, Arabic CLT comprehension scores are plotted as a function of daily exposure. We can gather from this plot that most children are reported to have fairly even exposure to Arabic and Swedish respectively (between 40% and 60%). Furthermore, we see a tendency that the more children are exposed to Arabic, the better they perform on Arabic CLT comprehension. This pattern is also confirmed when analysing the relationship between daily exposure to Arabic and scores on the Arabic CLT comprehension with a Pearson correlation, which revealed a strong to moderate positive relationship ($N = 98$, $r = .31$, $p < .01$). In Figure 4.8B, Arabic CLT comprehension scores are plotted as a function of SES (parental education). Inspecting the distribution, there does not seem to be a relationship between SES and comprehension scores on the Arabic CLT. This pattern is confirmed when analysing the relationship between SES and comprehension scores on the Arabic CLT with a Pearson correlation ($N = 98$, $r = .03$, $p = .77$).

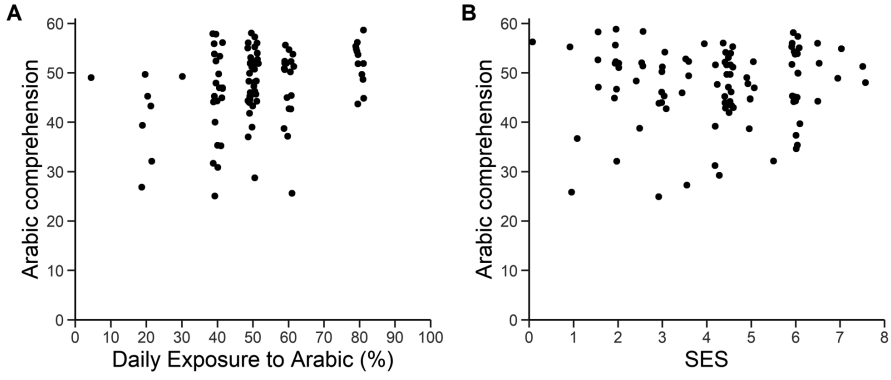

Figure 4.8. Scatterplots of CLT Arabic comprehension scores and (A) daily exposure to Arabic, and (B) SES.

In order to examine the combined effects of age, daily exposure to Arabic and SES on Arabic comprehension, a multiple linear regression was conducted, with comprehension scores on the Arabic CLT as a dependent variable and age, SES, and daily exposure to Arabic as independent variables. Table 4.4 gives an overview of the regression model. As displayed in Table 4.4, age and daily exposure to Arabic were significant predictors of Arabic comprehension scores, while SES was not. VIF values were below 5 for all independent variables, indicating low levels of multicollinearity.

Table 4.4. Scores on the Arabic CLT comprehension as a function of age (months), daily exposure to Arabic and SES.

	Coef.	**SE**	**Std. coef.**	***p***	**VIF**
Intercept	13.83	4.39		.002**	
Age	.31	.04	.57	<.001***	1.03
Daily exp Ara	.20	.04	.40	<.001***	1.03
SES	.26	.36	.06	.47	1.00
R^2 (adjusted)	.39				

Note. ** $p < .01$, *** $p < .001$. $F(3,94) = 21.91$, $p < .001$.

In Figure 4.9A, Arabic CLT production scores are plotted as a function of daily exposure to Arabic. Looking at the distribution, a similar but seemingly stronger pattern is evident for daily exposure for scores on CLT Arabic production compared to comprehension. Children with a higher level of reported daily input to Arabic scored better on Arabic production. This pattern is confirmed via a Pearson correlation that shows a moderate positive relationship (N = 98, $r = 0.48$, $p < .001$). In Figure 4.9B, Arabic CLT production scores are plotted as a function of SES (parental education). Inspecting the distribution, it seems like there might be a small tendency of a negative relationship between parental level of education and comprehension

scores on the Arabic CLT, such that children with higher SES score lower on the Arabic production. However, when investigating the relationship between daily exposure to Arabic and production scores on the Arabic CLT with a Pearson correlation, it becomes evident that this tendency is very small and not statistically significant (N = 98, r = –.11, p = .29).

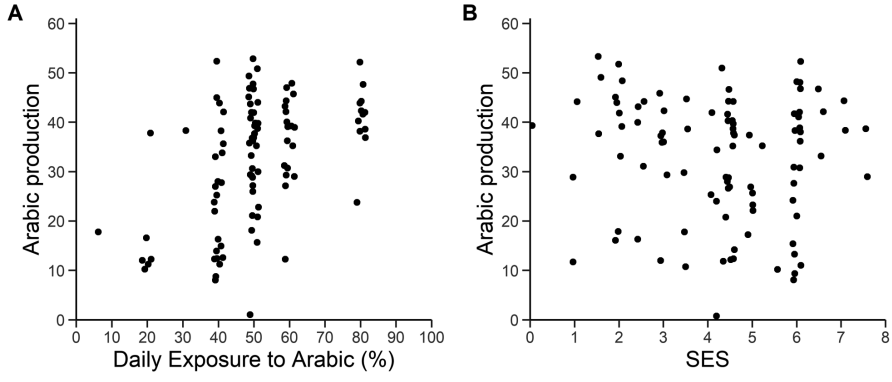

Figure 4.9. Scatterplots of CLT Arabic production scores and (A) daily exposure to Arabic, and (B) SES.

In order to investigate the combined effects of age, daily exposure to Arabic and SES on Arabic production, a multiple linear regression was conducted, with production scores on the Arabic CLT as a dependent variable and age, SES, and daily exposure to Arabic as independent variables. Table 4.5 gives an overview of the regression model. As we can see in Table 4.5, age and daily exposure to Arabic were significant predictors of Arabic production scores, while SES was not. VIF values were below 5 for all independent variables, indicating low levels of multicollinearity.

Table 4.5. Scores on the Arabic CLT production as a function of age (months), daily exposure to Arabic and SES.

	Coef.	**SE**	**Std. coef.**	***p***	**VIF**
Intercept	–12.78	7.30		.08	
Age	.36	.07	.40	<.001***	1.03
Daily exp Ara	.44	.07	.54	<.001***	1.03
SES	–.69	.60	–.09	.25	1.00
R^2 (adjusted)	.38				

Note. *** $p < .001$. $F(3,94) = 20.64$, $p < .001$.

4.3.2.4 Swedish CLT scores, input and SES

In this section, the combined effects of age, input (length of exposure to Swedish and daily exposure to Swedish) and SES will be investigated in relation to the Swedish CLT scores for comprehension and production.

Figure 4.10A shows a scatterplot of CLT Swedish comprehension scores as a function of length of exposure to Swedish. There appears to be a positive relationship between length of exposure to Swedish and CLT Swedish comprehension scores. This pattern is confirmed when conducting a Pearson correlation, which shows a moderate to strong positive correlation between Swedish CLT comprehension scores and length of exposure to Swedish ($N = 98$, $r = .63$, $p < .001$). Figure 4.10B shows scores for Swedish CLT comprehension as a function of daily exposure to Swedish. There seems to be a positive relationship between daily exposure to Swedish and scores for CLT Swedish comprehension. This pattern is confirmed when conducting a Pearson correlation, which shows a weak to moderate positive correlation between Swedish CLT comprehension scores and daily exposure to Swedish ($N = 98$, $r = .33$, $p < .001$). Figure 4.10C shows scores on the Swedish CLT comprehension as a function of SES (parental education level). As shown in the figure, there appears to be a slight tendency for a positive relationship between SES and Swedish CLT comprehension scores. When analysing this with a Pearson correlation however, it is evident that the relationship is weak and statistically nonsignificant ($N = 98$, $r = .14$, $p = .18$).

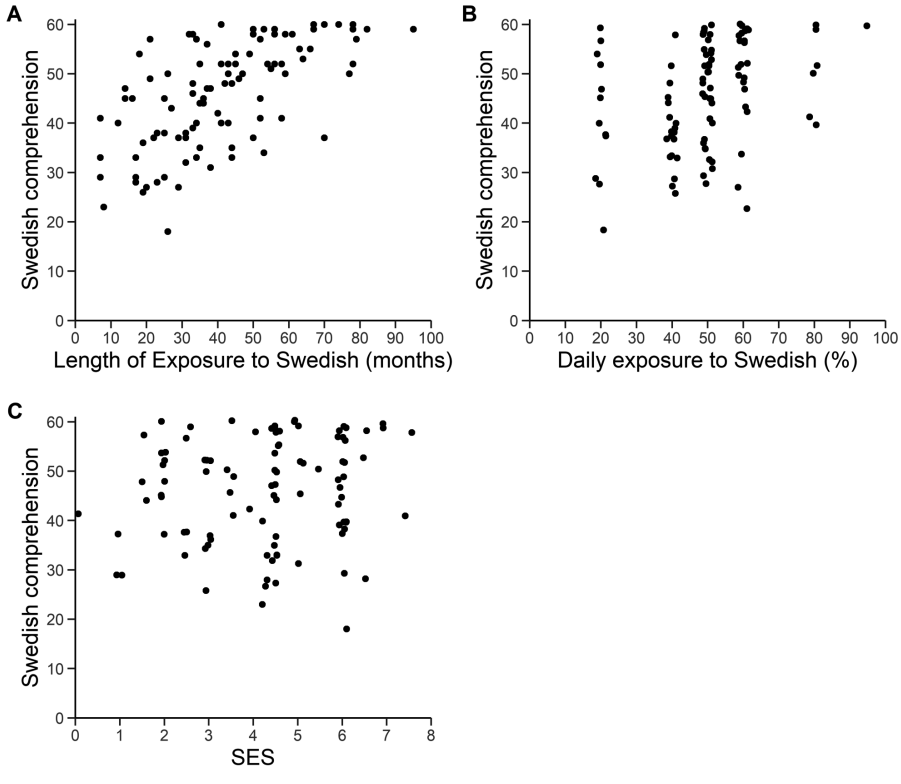

Figure 4.10. Scatterplots of CLT Swedish comprehension scores and (A) length of exposure to Swedish, (B) daily exposure to Swedish, and (C) SES.

In order to investigate the combined effects of age, input in Swedish and SES on Swedish vocabulary comprehension, a multiple linear regression was conducted, with comprehension scores on the Swedish CLT as a dependent variable, and age, SES, length of exposure to Swedish and daily exposure to Swedish as independent variables. As shown in Table 4.6, length of exposure to Swedish was a significant predictor of comprehension scores in Swedish, but daily exposure to Swedish and SES was not. The control variable age was also a significant predictor. Even though there was a positive correlation between daily exposure and comprehension scores, it was not significant when age, length of exposure and SES was also included in the model.[49] VIF values were below 5 for all independent variables, indicating low levels of multicollinearity.

[49] It should be mentioned that daily exposure approaches significance, as seen in Table 4.6, although the predictive power is low.

Table 4.6. Scores on the Swedish CLT comprehension as a function of age (months), length of exposure to Swedish, daily exposure to Swedish and SES.

	Coef.	**SE**	**Std. coef.**	***p***	**VIF**
Intercept	8.52	5.22		.11	
Age	.27	.06	.36	<.001***	1.26
LoE Swe	.22	.05	.41	<.001***	1.43
Daily exp Swe	.10	.05	.15	.06	1.15
SES	.77	.47	.12	.10	1.03
R^2 (adjusted)	.49				

Note. *** $p < .001$. $F(3,94) = 24.67$, $p < .001$.

Next, we turn to Swedish production. In Figure 4.11A, production scores on the Swedish CLT are plotted as a function of length of exposure to Swedish. Just as for the comprehension scores, there seems to be a positive correlation between production scores and length of exposure. This pattern is confirmed when analysing the relationship with a Pearson correlation, which shows a moderate to strong positive correlation between length of exposure to Swedish and scores on the Swedish CLT production (N = 98, $r = .62$, $p < .001$).

Figure 4.11B shows the Swedish production scores as a function of reported daily exposure to Swedish. Inspecting the figure, it seems like there is a positive relationship between daily exposure to Swedish and production scores. This pattern is confirmed when conducting a Pearson correlation, that shows a moderate correlation between exposure and production scores (N = 98, $r = .42$, $p < .001$). Figure 4.11C shows Swedish CLT production scores as a function of parental education (SES). The pattern does not suggest any correlation between the two variables, which is confirmed when analysing the relationship with a Pearson correlation (N = 98, $r = .10$, $p = .31$).

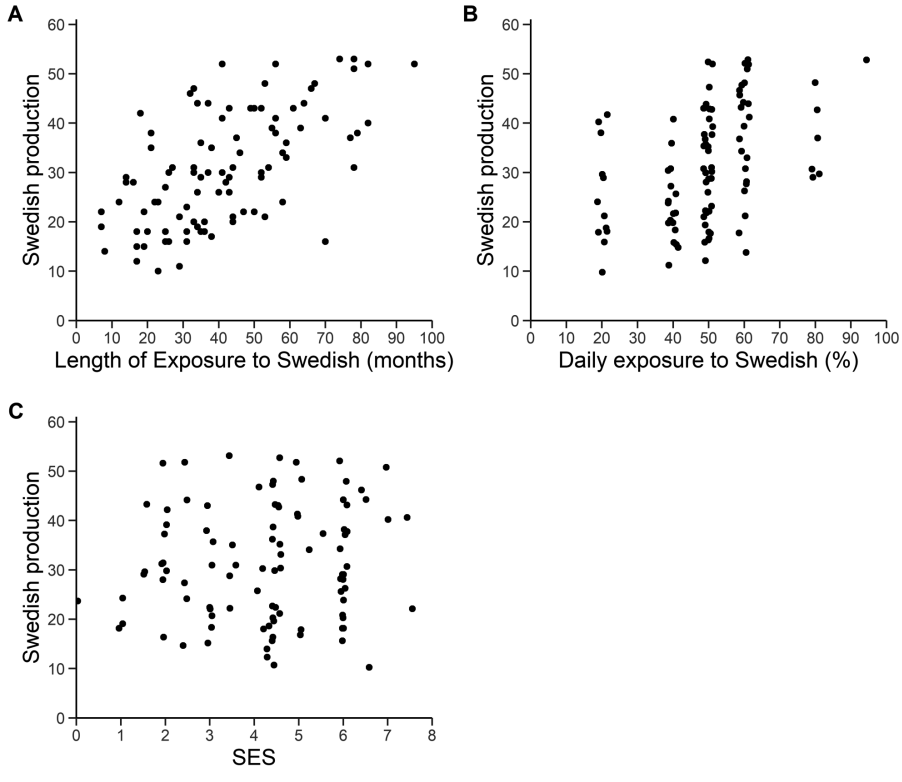

Figure 4.11. Scatterplots of CLT Swedish production scores and (A) length of exposure to Swedish (months), (B) daily exposure to Swedish (%), and (C) SES.

In order to investigate the combined effects of age, input in Swedish and SES on Swedish vocabulary production, a multiple linear regression was conducted, with production scores on the Swedish CLT as a dependent variable and age, SES, length of exposure to Swedish and daily exposure to Swedish as independent variables. As shown in Table 4.7, length of exposure to Swedish and daily exposure to Swedish were significant predictors of comprehension scores in Swedish, but SES was not. The control variable age was also a significant predictor. VIF values were below 5 for all independent variables, indicating low levels of multicollinearity.

Table 4.7. Scores on the Swedish CLT production as a function of age (months), length of exposure to Swedish, daily exposure to Swedish and SES.

	Coef.	SE	Std. coef.	*p*	VIF
Intercept	–7.80	5.76		.18	
Age (months)	.24	.07	.29	<.001***	1.26
LoE Swe	.23	.05	.39	<.001***	1.43
Daily exp Swe	.18	.06	.24	<.01**	1.15
SES	.61	.51	.09	.24	1.03
R^2 (adjusted)	.48				

Note. ** $p < .01$, *** $p < .001$. $F(3,94) = 22.97$, $p < .001$.

4.3.2.5 Vocabulary, language input and socioeconomic status: summary

In what follows, a summary will be provided with respect to the relationship between comprehension and production of Arabic and Swedish, and background factors such as age of onset, daily exposure, length of exposure and SES operationalised as parental level of education.

First, the results for comprehension of Arabic and Swedish will be summarised. For Arabic comprehension, there was a weak correlation with daily exposure, but no correlation with SES. The regression analysis showed that age and daily exposure were significant predictors of comprehension scores in Arabic, while SES was not. Age was a stronger predictor than daily exposure. For Swedish comprehension, there was also a weak to moderate correlation with daily exposure, but no correlation with SES. Furthermore, there was a moderate to strong correlation with length of exposure to Swedish. The regression analysis showed that age and length of exposure were significant predictors of Swedish comprehension scores, with similar predictive strength. Furthermore, daily exposure to Swedish and SES were not significant predictors of Swedish vocabulary comprehension.

Next, the results for production of Arabic and Swedish will be summarised. For Arabic production, there was a moderate positive correlation with daily exposure, but no correlation with SES. The regression model revealed that age and daily exposure were significant predictors of Arabic production scores, while SES was not. Age and daily exposure had similar predictive strength. For Swedish production, there was a moderate to strong correlation with length of exposure, and a moderate correlation with daily exposure, but no correlation with SES. The regression model showed that age, length of exposure and daily exposure were all significant predictors of Swedish production scores, but SES was not. Furthermore, length of exposure was a stronger predictor compared to age and daily exposure.

In sum, SES was not a significant predictor for CLT comprehension or production scores in either Arabic or Swedish. Daily exposure was a stronger

predictor of production scores than comprehension scores in Arabic. Similarly, for Swedish it was a significant predictor for production scores but not for comprehension scores. Length of exposure turned out to be a strong predictor of CLT scores in Swedish, for comprehension with similar strength as age; for production it was a stronger predictor than age.

4.4 Discussion

In this chapter, vocabulary comprehension and production was investigated in both languages of Arabic-Swedish-speaking bilinguals, using the Cross-linguistic Lexical Tasks (CLTs) in both languages. The vocabulary scores were analysed in relation to age in months, input (percent daily exposure and for Swedish length of exposure) and SES as operationalised by the parents' level of education. The following research questions were asked:

- How does vocabulary comprehension and production develop with age in Arabic and in Swedish?
- Are there any differences between vocabulary comprehension and production in Arabic and Swedish?
- What is the relationship between age, language input, socio-economic status (SES) and vocabulary scores?

In this section, findings from the current study are discussed in relation to previous studies with respect to age effects (section 4.4.1), input effects (section 4.4.2) and SES effects (section 4.4.3).

4.4.1 Age effects on vocabulary

In this study, age effects were investigated in two ways for comprehension and production of Arabic and Swedish: linearly via correlation and as part of the multivariate regression analyses, and for age groups with ANOVAs.

There was considerable individual variation in performance on comprehension and production of Arabic and Swedish alike. Nevertheless, a positive association was observed between age in months and scores on all vocabulary tasks (comprehension and production in both Arabic and Swedish). Performance increased with age for both the minority language Arabic, and the majority language Swedish. These results are in line with the findings of Haman et al. (2017), showing a positive correlation between chronological age and vocabulary scores in 639 *monolingual* children (age 3;0–6;11), who were tested on the CLT in their respective language version.

The findings regarding development with age were in some ways different in the current study compared to previous studies investigating vocabulary development in the majority and minority language of Swedish bilinguals.

Bohnacker et al. (2016) investigated age effects in vocabulary production in the minority language (German/Turkish) of 38 German-Swedish-speaking bilinguals and 40 Turkish-Swedish-speaking bilinguals age 4–6. There was no significant increase with age for vocabulary production in the bilingual children's minority languages. Lindgren (2018) investigated age effects on vocabulary production in Swedish monolinguals (N = 72), German-Swedish-speaking bilinguals (N=46) and Turkish-Swedish-speaking bilinguals (N=48) age 4–6. Although there was an increase in scores with age in the majority language Swedish for all groups, there was no increase with age for vocabulary production in the bilingual children's minority languages. In the current study, there was a significant development with age in both languages that was stronger for comprehension than production, and stronger in Swedish than Arabic. One possible explanation for the difference in results between the findings in Bohnacker et al. (2016) and Lindgren (2018) and the current study is that the participants in those studies had a narrower age range. Furthermore, only vocabulary production was analysed, which had a weaker relationship with age compared to vocabulary comprehension in the current study. Öztekin (2019, Chapter 4) investigated vocabulary comprehension and production in both the minority and the majority language of 102 Turkish-Swedish-speaking bilinguals age 4–7. Öztekin found a significant but small age effect in the minority language for comprehension (however, the majority of the children (61%) scored above 90% correct, suggesting a ceiling effect), and marginally so for production. For the majority language Swedish, there were strong correlations between chronological age and performance on both comprehension and production. The correlation coefficients were larger in Öztekin's study compared to the current study, suggesting a stronger effect of chronological age in the majority language (Swedish) in the Turkish-speaking group compared to the Arabic-speaking group in this study.[50]

In an international perspective, the results of the current study mirror many previous findings that show a clear development with age in the majority language (in this case Swedish; Bialystok et al., 2010; Cobo-Lewis et al., 2002a; Dijkstra et al., 2016; Prevoo et al., 2014). Although somewhat smaller than for the majority language, there was also a clear development with age for both comprehension and production in the minority language (Arabic), a pattern that is not always seen for minority language development. As previously mentioned, vocabulary growth in the minority language may be smaller than in the majority language (Ganuza & Hedman, 2019; V. C. M. Gathercole & Thomas, 2009; Hoff et al., 2014; Leseman, 2000; Öztekin, 2019, Chapter 4). This is often attributed to the fact that as children get older,

[50] The stronger age effect in Swedish in the Swedish-Turkish-speaking children was confirmed for ages 4–6 (Bohnacker, Haddad, Lindgren, Öberg, & Öztekin, 2020/in press), but not when the full age range (4–7) was analysed in Bohnacker, Haddad, Lindgren, Öberg & Öztekin (2021/in press).

exposure to the minority language declines, partly as an effect of schooling in the majority language (Montanari, Abel, Graßer, & Tschudinovski, 2018). However, the situation for the Arabic-speaking population in Sweden is different from many other groups. The Arabic-speaking population is the largest linguistic minority in Sweden.[51] Because of this, Arabic-speaking children are likely to get the chance to speak the language in many other places outside the home. Second, the vast majority of children in Sweden (including bilinguals) attend preschool from an early age. Even though this is mostly conducted in the majority language Swedish there are some preschools and schools in Sweden with a bilingual Swedish/Arabic or Islamic profile. In fact, 23 children in the sample attended such a (pre)school. Even though Arabic is not the sole language of conversation, nor the language of instruction in such schools, it is likely that Arabic will be more prominent in the overall daily language exposure for these children compared to Arabic-speaking children who attend 'regular' Swedish schools. Third, differences in patterns of language input and proficiency may be related to migration patterns in the Arabic-speaking population compared to other minority language groups. The recent wave of migration to Sweden due to the war in Syria means that there are also a substantial group of Arabic-speaking children with a migrant background. This was reflected in the number of children in the sample who were born abroad (45%). Furthermore, 46% of the children had an age of onset to Swedish after age 3, and 20% had less than two years (24 months) of exposure to Swedish at the time of testing. As a comparison, in Öztekin's (2019) Turkish-speaking sample, 96% of the children were born in Sweden and 81% had an age of onset to Swedish before age 3. At group level, the children in the present study had a higher amount of cumulative exposure to their minority language compared to the children in Öztekin's sample, and they likely also had more sources of continuous input in the minority language, which could explain the stronger development with age for minority language vocabulary production in the current study.

Although there was a general development with age for comprehension and production in both Arabic and Swedish in the current study, there was much individual variation in performance even in the oldest age group. This was true particularly for production in Arabic and Swedish, but also for comprehension in Swedish, suggesting that input factors play an important role for performance on the CLTs. This will be discussed in the next section.

[51] There are no official statistics in Sweden on the number of speakers of different languages. According to an unpublished internal report by Bohnacker (2017), the number of Arabic speakers in Sweden likely exceeds 400 000, based on official statistics on the number of people originating from an Arabic-speaking country residing in Sweden.

4.4.2 Effect of input on vocabulary

In this study, input effects were investigated via one variable (percent daily exposure) for Arabic, and via two variables (length of exposure and percent daily exposure) for Swedish linearly via correlations as well as a part of the multivariate regression models. The analyses were conducted separately for comprehension and production in each language. Information about these input/background factors was obtained from the parental questionnaire.

First, percent daily exposure will be discussed. For both the majority language Swedish and the minority language Arabic, there was a positive correlation between reported daily exposure and vocabulary scores. This finding matches that of e.g. Prevoo et al. (2014), who found a positive correlation between reported daily exposure and vocabulary comprehension in the minority language (Turkish) and vocabulary production in the majority language (Dutch). It is also similar to Unsworth (2016), who found that a higher amount of current exposure to the majority language (Dutch) yielded higher vocabulary comprehension scores in that language. For both Arabic and Swedish, the correlation between daily exposure and vocabulary scores was stronger for production than comprehension. Although Thordardottir (2011) investigated the *cumulative effect* of relative exposure (i.e. relative exposure to a language over time), a comparison is still relevant to make. Thordardottir found that less relative exposure (40–60%) was needed in order to score high (i.e. in line with monolinguals) in *receptive* vocabulary compared to *expressive* vocabulary, where 80% relative exposure was needed to score high. In the current study too, relative daily exposure had a larger impact on vocabulary production than comprehension.

Several researchers have argued that there is a fundamental difference between simultaneous and sequential bilingualism (De Houwer, 2009, pp. 4–7; Meisel, 2009). However, recent studies indicate that a binary division of 'early' vs. 'late' onset of bilingualism does not seem to explain vocabulary performance of children in the preschool and early school years, when cumulative amount or current amount of exposure is also taken into account (Thordardottir, 2019; Unsworth, 2016). In the current study, length of exposure to Swedish (in months) was used to measure the effect of timing of bilingual onset. Length of exposure was significantly correlated with both comprehension and production scores. In the multivariate regression analyses, length of exposure to Swedish was a stronger predictor of vocabulary scores than chronological age, both for comprehension and production. Interestingly, the results go against those of Thordardottir (2019), who found that length of exposure was not a significant predictor of vocabulary comprehension scores in 7–9-year-olds when *cumulative* amount of exposure was also accounted for. Since only *current* amount of exposure was measured in the current study (and as a separate variable), the length of exposure variable is likely to capture length as well as cumulative amount of exposure, at least to some degree. In

order to investigate the relative effect of length of exposure vs. amount of exposure thoroughly, parents would have to be interviewed about cumulative amount of exposure, which was not done in this study.

Finally, it should be mentioned that only quantitative input measures were analysed in this study. As underscored by for instance Hart and Risley (1995), not only *quantitative* aspects of language input have an impact on vocabulary development, but also *qualitative* aspects play a role. As demonstrated by Rowe (2012), parental use of diverse and sophisticated vocabulary and decontextualised language are important predictors of children's receptive vocabulary development. One example of such decontextualised language is narrative and explanatory discourse, used for instance during book reading and storytelling activities. Increased amounts of book reading, media consumption and engaging in extracurricular activities in a certain language is likely to boost children's vocabulary as well as their overall language skills in that language. Another qualitative aspect of input that is likely to affect vocabulary development is which language is spoken with interlocutors such as siblings, parents, extended family and friends. For instance, Öztekin (2019, Chapter 4) found that children whose parents spoke only or mostly in the home language Turkish to them and with each other had higher vocabulary scores in comprehension as well as production in Turkish. Performance in the minority language (Turkish) was also higher for children who spoke only or mostly Turkish with their parents. In the minority language, which has a more vulnerable position compared to the majority language, the effect of interlocutors is likely to be particularly important if the amount of input is limited. This was documented by Bohnacker et al. (2016), who found that for children whose parents did not address them mainly in the home language (alternatively, when only one parent did so), having additional input providers (in this case, friends) had a positive effect on vocabulary scores in the minority language. The questionnaire used in the current study did contain questions about such qualitative aspects of input (e.g. book reading, extracurricular activities, interlocutors with whom the child spoke the minority language). Although not feasible within the confines of the current study, further studies should look into these qualitative aspects of input and their effect on vocabulary skills.

4.4.3 Effect of SES on vocabulary

In this study, SES was operationalised as parental level of education. SES effects were investigated linearly via correlation as well as a part of the multivariate regression models. The analyses were conducted for comprehension and production separately for each language.

There was no effect of SES on vocabulary comprehension or production, neither in the minority language Arabic, nor in the majority language Swedish. As previously stated (see section 4.1.2), numerous studies have found that at

group level, monolingual children from lower-SES backgrounds tend to have smaller vocabularies than children from higher-SES backgrounds (Hart & Risley, 1995; Rowe, 2012). The same pattern has also been established for bilingual children's vocabulary skills in the majority language (Buac et al., 2014; Calvo & Bialystok, 2014; Cobo-Lewis et al., 2002a; Prevoo et al., 2014). By contrast, SES does not seem to have a substantial effect on the vocabulary development in the minority language (Buac et al., 2014; Cobo-Lewis et al., 2002b; Prevoo et al., 2014).

Based on previous findings, the fact that no effect of SES was found for the minority language Arabic is not surprising, and it is also in line with the previous studies on Turkish-Swedish-speaking bilinguals by Bohnacker et al. (2016) and Öztekin (2019). Perhaps more unexpected is the fact that SES was not a significant predictor of vocabulary performance in the majority language. There may be several reasons for this. Öztekin (2019) compared vocabulary performance in the majority language in Turkish-Swedish-speaking bilinguals from high-SES vs. low-SES households, and found no effect of SES on majority language vocabulary performance.[52] As Öztekin (2019, p. 127) points out, other ways of operationalising SES (for instance parental occupation, income level, residence area or any combination of these) may be relevant in the Swedish context. Another possible explanation could be that SES simply is not a determining factor when it comes to language development in Sweden, at least not in the late preschool and early school ages. In contrast to many other countries where childcare is expensive and often exclusive to families with a higher income, Sweden has an affordable childcare system where preschool services are available to all. Because of this, most children (also bilinguals and children from lower-SES households) attend preschool from an early age. As such, all children, irrespective of SES status, have similar chances of receiving early-onset and regular exposure to Swedish via preschool, which may explain the lack of influence of SES on majority language vocabulary skills in the current study. However, a recent study by Andersson, Hansson, Rosqvist, Lyberg Åhlander, Sahlén and Sandgren (2019) investigated the effect of bilingualism, SES (parental education), school characteristics (percentage of parents with tertiary education; percentage of students with Swedish as a second language) and recreation centre enrolment on the performance on 'core language skills' from CELF-4 (Semel et al., 2013) in the majority language Swedish in a group of 224 7–8-year-olds in Sweden.[53] They used hierarchical regression analyses in order to investigate the unique and combined effects of all predictors. All variables had high levels of shared variance, and the predictive power of

[52] The division between high SES and low SES was ISCED level 3 (completed upper-secondary education and below) and 4 (post-secondary education, including university degrees).
[53] 'Core language skills' includes the subtasks 'Concepts and following directions', 'Word structure', 'Recalling sentences' and 'Formulated sentences'.

bilingualism alone decreased when SES (parental education) and school characteristics were also included in the model. Thus, children speaking Swedish as a second language, growing up in lower-SES households at the same time as attending a school with a high degree of pupils sharing that same background ran an increased risk of scoring low in core language skills in Swedish. Thus, SES is likely to have an impact also on vocabulary skills in Swedish bilingual children. In order to investigate the relationship between SES and vocabulary skills further, future studies should include other measures of SES than the one that was used in the current study, at the same time as examining (pre)school characteristics.

4.4.4 Concluding remarks

In the present study, vocabulary comprehension and production was investigated Arabic-Swedish-speaking bilinguals, using the Cross-linguistic Lexical Tasks (CLTs) in both languages. There was much individual variation in performance, which was related to language exposure patterns but unrelated to SES (parental education). The maximum score was reached for CLT comprehension in the majority language (Swedish, N=5) but not in the minority language (Arabic), despite the fact that comprehension scores were overall higher in Arabic. There were also more children who scored above 90% in Swedish comprehension (N=29) compared to Arabic comprehension (N=22). There may be several reasons for why the maximum score was not reached in Arabic. First, the children spoke different varieties of Arabic. Although the comprehension prompts were adapted to match the Arabic variety of each child, it may be the case that the test disadvantaged some children on certain vocabulary items. Furthermore, it may also be the case that there were certain items that were overall difficult, irrespective of the Arabic variety the child spoke (a so-called item effect). Finally, for some children assessment in Arabic was conducted in environments that were generally Swedish-speaking (i.e. (pre)school), which may have induced more instances of code-switching. Future studies could investigate the possible influence of these factors on the CLT comprehension scores in Arabic.

5 The cross-sectional study: Phonological working memory

Phonological processing involves making use of one's language knowledge (phonological skills in particular) in order to process spoken and written language. It is often described as having the following three main components, which are in turn reliant on lower-order functions that involve hearing, decoding, encoding and articulating speech sounds (Wagner & Torgesen, 1987, p. 192):

- Phonological awareness (PA): being aware of the sound structure of words, and being able to consciously analyse and manipulate phonological material, such as adding, deleting or moving segments or syllables in words or non-words.
- Phonological working memory (PWM): a temporary store of phonological information, which enables manipulation of such material – for instance during PA tasks. Non-word repetition (NWR), which entails repeating a series of phonological nonsense forms, is one example of a task that taps into PWM.
- Phonological retrieval (PR): recall of phonological information associated with specific graphemes or pictures (words). This ability is typically assessed via rapid naming tasks, where a person is confronted with a series of numbers or pictures (typically depicting figures of varying shape and colour), and is instructed to name them as accurately and quickly as possible.

Phonological processing skills are essential for the development of literacy proficiency (Melby-Lervåg, Lyster, & Hulme, 2012).[54] Also, deficits in phonological processing abilities are associated with developmental disorders of language and literacy in general; however, while weak PA and PR are more associated with reading disorders (RD), impaired PWM has a stronger association with DLD (Ramus, Marshall, Rosen, & van der Lely, 2013). NWR has been proposed as a clinical marker for DLD in many languages, as numerous studies have shown that children diagnosed with DLD have poor

[54] At least this is the case in languages that use alphabetic or syllabic writing systems. To which extent it is also true for languages that use e.g. logographic writing systems is less clear.

NWR skills compared to their typically developing peers (for a comprehensive overview, see Chiat (2015)). Furthermore, some studies indicate that there may be a hereditary component to deficient NWR skills (Bishop, North, & Donlan, 1995). Kalnak, Peyrard-Janvid, Forssberg, and Sahlén (2014) found that children with DLD whose parents had self-reported difficulties with language and literacy performed significantly worse on NWR than their peers with DLD who had unaffected parents, supporting the idea that there is a hereditary component to NWR performance.

The following research questions are asked in this chapter:

- How does performance on four different NWR tasks develop from age 4 to 7?
- Are there differences in performance between tasks?
- What is the relationship between participant-related factors such as age, vocabulary and length of exposure, and item-related variables such as type of task, item length, and syllabic complexity and performance on NWR items?

In section 5.1, earlier studies concerning non-word repetition and factors influencing performance will be summarised. Next, some considerations for choosing non-word tasks for the present study will be presented in section 5.2, followed by section 5.3 on transcription and scoring. After that, results will be reported with respect to age development (section 5.4.1) and the influence of item-related as well as participant-related factors (section 5.4.2). Finally, the findings will be discussed in section 5.5.

5.1 Background literature

5.1.1 Working memory

Working memory (WM) is a theoretical construct describing a cognitive system that enables individuals to simultaneously process and store information for short time periods. WM is thought to play a fundamental role in most higher-order cognitive tasks such as directing attention, monitoring behaviour, learning, reasoning, decision-making as well as processing and comprehending language (Diamond, 2013). Several models of WM have been proposed. These include models describing WM as an embedded part of long-term memory (Cowan, 2001; Ericsson & Kintsch, 1995) and models making a distinction between procedural (non-conscious) and declarative (conscious) working memory (Oberauer, 2009). However, Baddeley and Hitch's *multi-component model of working memory* (1974) and its subsequent revised versions (see for instance Baddeley (2000, 2002)) is currently the most established.

Baddeley and Hitch's multi-component model (depicted in Figure 5.1) introduced the notion of a *central executive*, in which storage and processing of information takes place. This central executive was described as having two submissive subsystems: the *articulatory loop* (later renamed *the phonological loop*), and the *visuo-spatial sketchpad*. The revised model also contains reciprocal connections between the central executive (via the two sub systems and the episodic buffer) and different aspects of the long-term memory (e.g. episodic long term memory, language, visual semantics, etc.), in the so-called *episodic buffer*, in which information can be stored and modified between the central executive and the episodic memory.

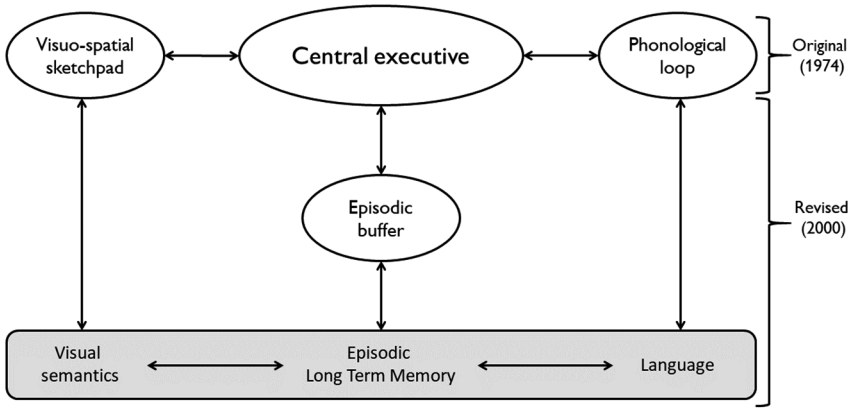

Figure 5.1. The multicomponent model of working memory. The top section ("original") represents Baddeley and Hitch's original model (Baddeley & Hitch, 1974), and the bottom section ('revised') represents the supplementary components that were added in the revised model (Baddeley, 2000, 2002). The parts of the model that are transparent represent systems that involve temporary activation only, whereas the grey-shaded area represents knowledge and skills that are fundamentally perpetual or stabilised in long-term memory.

5.1.2 Phonological working memory and factors influencing NWR performance

The phonological component of the working memory was further described in detail by Baddeley (2002). The phonological working memory, or phonological loop, is the part of WM that handles verbal and auditory information. This is the main entity for performing phonological processing operations, such as NWR. This phonological loop has two sub components: the *phonological short term store*, in which auditory memory traces (from perceptual input) are subject to rapid decay, unless this trace is repeated in the

articulatory rehearsal component, an 'inner voice' function that enables the memory trace to be refreshed in the phonological short term store.

Three main types of tasks are traditionally used to assess phonological working memory: *digit span*, *word repetition* and NWR.[55] However, all three tasks are to some degree dependent on long term memory knowledge. In the case of digit span, more frequent digit sequences yield better serial recall than do less frequent digit sequences (Jones & Macken, 2015). In the case of word repetition, factors such as word frequency, as well as neighbourhood density (the number of phonologically similar words) play an important role for performance (Goh & Pisoni, 2003). For NWR, a wider set of linguistic, cognitive and articulatory skills and abilities, as well as factors related to the phonological structure of the non-word items influence repetition performance. Gathercole (2006) suggests that NWR may provide a more genuine measure of phonological working memory, since lexical reconstructions cannot compensate for impaired phonological storage when non-words are used as stimuli.

The phonological loop has been proposed to have evolved as a language learning device, with the primary function of storing unfamiliar sound patterns during the process of establishing a stable memory record (Baddeley, Gathercole, & Papagno, 1998). As argued by Gathercole (2006), the ability to repeat non-words is closely associated with language learning, and vocabulary learning in particular. Since all words we know were first unknown to us, repeating a non-word can be likened to the experience of first encountering a new word. NWR performance has been found to correlate with vocabulary scores (S. E. Gathercole & Adams, 1993; S. E. Gathercole & Baddeley, 1989), but the causational relationship between NWR and vocabulary has been debated. Non-word repetition performance has been put forward as a predictor of vocabulary scores (Bowey, 1996), but the reverse pattern has also been proposed (Dollaghan, Biber, & Campbell, 1995; S. E. Gathercole & Adams, 1994). Coady and Evans (2008) reviewed 24 studies investigating the relationship between NWR performance and vocabulary scores in (primarily) English-speaking TD children (ages ranged between 1;8–12;5). They concluded that the relationship between NWR and vocabulary is likely reciprocal, since the direction of causality seems to shift over time.

Impairments in the phonological loop capacity and associated abilities has been suggested as one of the main causes for the word learning deficits that are associated with DLD/(S)LI (S. E. Gathercole & Baddeley, 1990a).

[55] Digit span is a task that requires listening to and repeating a series of number sequences, for instance *one–seven–four–six*. Typically, there is a stepwise increase in difficulty by adding an extra number. Word repetition tasks follow the same principle as digit span. They require listening to and repeating a sequence of real words, with a stepwise increase in difficulty by adding an extra word. Digit span and word repetition can be examined by either forward repetition or backward repetition, where backward repetition requires repeating the number or word sequence in reversed order.

Consistent with this hypothesis is for example studies that have found that children with poor NWR scores learn the phonological form of new words slower than children with higher NWR scores (S. E. Gathercole & Baddeley, 1990b; Michas & Henry, 1994), and that children diagnosed with DLD need more repetitions in order to learn the phonological form of new words compared to TD age matched peers (Nash & Donaldson, 2005). However, as emphasised by Gathercole (2006), poor phonological loop capacity is not likely to be the sole cause of the impaired language learning abilities that characterises DLD.

According to Graf Estes, Evans and Else-Quest (2007), the ability to repeat phonological nonsense forms may be influenced by any of the component skills that are involved in the process of hearing, encoding and producing a word or sound sequence. They point out that the impaired abilities present in the cognitive system of children with DLD may affect one single or many sub-elements in this process. These sub-elements include perception of speech distinctions, the organisation of phonological and morphological representations, storage of phonological material, as well as motor planning and articulation. In younger children who are still in the process of acquiring and automatizing their phonological system and articulation, this factor has more of an impact than in older children, whose phonological system is stabilised. Furthermore, Chiat (2015) mentions five main aspects of non-word characteristics that influence performance, namely: length, segmental complexity, prosodic structure, phonotactic probability and wordlikeness. Since languages vary with respect to these aspects, they play out differently in different languages. In conclusion, NWR cannot be viewed as exclusively targeting one single ability; rather it taps into a combination of cognitive and motor abilities and long-term linguistic knowledge.

5.1.3 Characteristics of non-word items affecting NWR performance

5.1.3.1 Item length

There are two ways of defining item length, which also to some extent coincide with item complexity. The most common way of defining item length is by the number of syllables the NWR item contains. An alternative, traditionally less common way of defining item length is by the number of phonemes it contains (Szewczyk, Marecka, Chiat, & Wodniecka, 2018). Some researchers argue that NWR tasks should only contain items of 1–3 syllables. According to dos Santos and Ferré (2018, p. 59) and Abed-Ibrahim and Hamann (2017, p. 5), NWR items containing more than three syllables tax WM, thus no longer tapping into phonological abilities. However, no arguments are provided that support why a maximum of three syllables should be the most suitable cut-off. Many commonly used NWR tasks have a range

of 2–4-syllable items (with some also employing monosyllabic items like the Non-word Repetition Task (NRT; Dollaghan & Campbell, 1998) or five-syllable items like the Children's Non-word Repetition test (CNRep; S. E. Gathercole, Willis, Baddeley, & Emslie, 1994), the Swedish non-word repetition test used in the present study (Radeborg et al., 2006), and the quasi-universal non-word repetition test (Chiat, 2015).

Several studies report a length effect, with a decrease in performance with longer non-word items in children with typical language development and DLD alike (Boerma et al., 2015; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Radeborg et al., 2006; Thordardottir & Brandeker, 2013; Topbaş et al., 2014). Kalnak et al. (2014) did not find a length effect in their study on 8–12-year-old children (N=147) with and without DLD. However, their task only consisted of 3–4-syllable non-words, and the children in their sample were older than the previously mentioned studies who did find length effects.

Although there is a general effect of length, it does not necessarily hold across the full range of number of syllables, and there may be different patterns for TD children and DLD children. In their study of a sample of TD pre-schoolers (age range 2;0–4;0) and a clinically referred group (age range 2;6–4;0), Chiat and Roy (2007) found that the clinically referred children had significantly poorer performance than the TD group on monosyllabic words and non-words compared to items containing three syllables. No significant differences were found between mono- and disyllabic or disyllabic and trisyllabic items. By contrast, in the clinically referred sample there was a significantly poorer performance on three-syllable items compared to disyllabic, and on monosyllabic words compared to disyllabic ones. Likewise, Dollaghan and Campbell (1998) found a significant difference in performance between their TD and DLD groups (age range 6;0–9;9) for items that were 3 or 4 syllables long, but not for 1–2-syllable items. Saiegh-Haddad and Ghawi-Dakwar (2017) found a significant difference in performance between TD and DLD children for non-words at all syllable lengths (1–4 syllables) in two age groups: children in senior kindergarten (mean age: 5;9) and first grade (mean age: 6;11).

In sum, several studies report an effect of length, with longer items generally being more difficult to repeat. Furthermore, while children with DLD generally perform more poorly on NWR tasks than their TD peers do, there is not always a significant difference for every syllable length, and group differences seem to appear at different item lengths at different ages. Very young children with DLD may perform more poorly than TD age peers at very short syllable lengths, but at older ages, there may not be a difference in performance on items with few syllables.

5.1.3.2 Item complexity: syllable structure

A syllable consists of two fundamental parts: the *onset* (a consonant or consonant cluster) and the *rime*. The rime can further be divided into two parts: the *nucleus* (a vowel or syllabic consonant), and the optional *coda* (a consonant or consonant cluster). The structure of a syllable is schematised in Figure 5.2. Syllable structure is commonly represented using capital letters: consonants with a capital C, and vowels with a capital V. Thus, a simple syllable with no consonant clusters and no coda would be represented as CV, and a complex syllable with an initial consonant cluster and coda could be represented as CCVC. These representations of syllable structure will be used in the following.

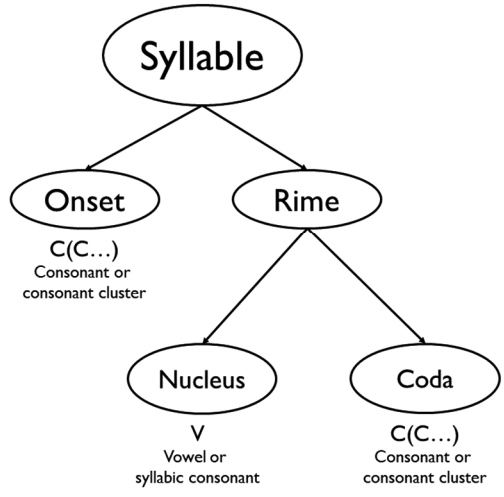

Figure 5.2. Schemata of the structure of a syllable.

NWR tasks vary regarding whether they contain consonant clusters or not. In some languages such as Turkish, consonant clusters are rare or even non-existent. Therefore, NWR tasks constructed for such languages naturally do not contain clusters, see for instance Topbaş et al. (2014). Other languages vary in their phonotactics, which means that there is a variation in the phonological structures of syllables. This means that non-word items with a variety of syllable structures are possible in these languages.

Many NWR tasks originally constructed for languages that allow clusters also contain cluster items such as Radeborg et al. (2006) for Swedish or the CNRep (S. E. Gathercole et al., 1994) for English. However, there are also tasks developed for languages allowing clusters that do not contain such structures, such as Rispens and Baker (2012) for Dutch and the NRT for English (Dollaghan & Campbell 1998). Some researchers report that their NWR task included items with clusters and without, but do not discuss any possible differences in performance (Kalnak et al., 2014; Meir & Armon-

Lotem, 2017). Other studies have investigated syllable structure by comparing results for items with different syllable structures (Abed Ibrahim & Hamann, 2017; dos Santos & Ferré, 2018), or by comparing two different NWR tasks where one contains clusters and the other does not (Boerma et al., 2015; Thordardottir & Brandeker, 2013). Abed Ibrahim and Hamann (2017) found that items containing clusters were more difficult than items without clusters in a task developed for German, and dos Santos and Ferré (2018, pp. 66–67) found that items with two clusters had a significantly worse performance rate than those with only one cluster in a task developed for French. Thordardottir and Brandeker (2013) found that a group of English-French bilinguals performed better on a French NWR task containing no clusters than on the English CNRep (S. E. Gathercole et al., 1994) that does contain clusters. In sum, there is much evidence that shows that items and tasks with a higher degree of syllabic complexity are more difficult to repeat than those with a lower degree of complexity.

Apart from general findings that syllable length and complexity affect the accuracy of repetition, *sublexical features* also play a role. These features are related more directly to inherent characteristics of the target language, and encompass *prosody*, *phonotactic probability* and *wordlikeness*. Depending on the child's language experience and the structure of the non-words pertaining to sublexical features, performance may be affected in different ways.

5.1.3.3 Prosody

Not many studies have explored the topic of prosody, but some papers should be mentioned here. While some languages like for instance Turkish and Hungarian are syllable timed with even stress on each syllable, other languages like Swedish and English have variable stress patterns and word prosody. In studies investigating NWR accuracy in Swedish and English, extrametrical syllables (i.e. syllables appearing before a stressed syllable) are more prone to deletion than syllables that fall within the common metrical template (Chiat & Roy, 2007; Sahlén, Reuterskiöld-Wagner, Nettelbladt, & Radeborg, 1999). Furthermore, Sundström, Samuelsson and Lyxell (2014) found that non-words carrying the grave tonal accent had a higher rate of correctly repeated segments compared to non-words with the acute tonal accent in a sample of monolingual Swedish-speaking 4–6-year-olds.[56] The authors speculate that since the grave accent is more salient and acquired earlier than the acute accent, it may provide an explanation for why children repeated non-words with a grave accent correctly to a higher degree.

[56] Swedish is a so-called pitch-accent language, where (disyllabic) words carry one of two lexical pitch contours, often called accent 1 (or acute accent: *stegen* /ˈstégen/ 'the steps') and accent 2 (or grave accent: *stegen* /ˈstègen/ 'the ladder'). There are roughly 350 minimal word pairs differing only via this distinction in Swedish (Riad, 2014, pp. 181–182).

5.1.3.4 Phonotactic probability

Phonotactic probability refers to the probability that a given phoneme sequence in a non-word would appear in that order in a real word in the target language. These calculations are most often conducted on bigram sequences (two adjoining phonemes) or trigrams, but could in theory be calculated for any number of phonemes in the non-word. A number of studies report that children are significantly better at repeating non-words that have higher phonotactic probabilities than those that have lower ones. This effect has been reported by Munson, Ewards and Beckman (2005) and Munson, Kurtz and Windsor (2005) for groups of English–speaking 3–6-year-olds and 11-year-olds, as well as by Rispens, Baker and Duinmeijer (2015) for Dutch-speaking 5–8-year-olds, Topbaş, Kaçar-Kütükçü and Kopkalli-Yavuz for a group of Turkish-speaking 4–7-year-olds (2014), and Szewczyk, Marecka, Chiat and Wodniecka for Polish-speaking 4–6-year-old monolinguals (2018).

The findings from studies investigating the role of phonotactic probability on performance of NWR seem to suggest that in general, children perform better on items with high phonotactic probability. Furthermore, results seem to indicate that children with DLD may perform disproportionally worse on items with low phonotactic probability than their typically developing peers (Munson, Kurtz, et al., 2005; Rispens et al., 2015), at least for some syllable lengths (Jones, Tamburelli, Watson, Gobet, & Pine, 2010). However, more research on larger populations is needed in order to determine whether items with high vs. low phonotactic probability can distinguish between TD and DLD.

5.1.3.5 Wordlikeness

Wordlikeness is a subjective measure. Contrary to phonotactic probability measures, it captures subjective ratings of the perceived wordlikeness vs. non-wordlikeness of a particular non-word. A common way of determining the likeness vs. non-wordlikeness in a set of non-words is to ask a group of native speakers of a language to rate the non-word items on a Likert scale, and ask them to rate each item as sounding more or less like a real word in that particular language. To complete the task, the rater may make use of (conscious or subconscious) linguistic knowledge such as presence or absence of real morphemes in that language, frequency of phoneme sequences, neighbourhood density (the number of phonologically similar words in the lexicon), non-word length, as well as suprasegmental features such as word prosody (stress patterns, pitch accent). Another approach is to construct one set of non-words that contain some characteristics of real words in that language, and another set of non-wordlike items that do not contain these features. These characteristics of real words may be grammatical and derivational morphemes such as in the CNRep in English (S. E. Gathercole et

al., 1994) or typical consonant roots and vowel patterns, such as in the Hebrew NWR task used by Meir and Armon-Lotem (2017).

A number of publications have investigated the effect of wordlikeness on performance on non-word repetition tasks. Gathercole (1995) found that there was a correlation between rated wordlikeness and performance on the CNRep in a group of 70 4–5-year-old monolingual English speaking children. Jones et al. (2010) investigated the effect of lexicality on performance on NWR items on a group of monolingual English speaking 6-year-olds, 18 with DLD (range 5;7–6;7) and 18 with TD (range 5;7–6;6). They found that the DLD group repeated high-lexicality non-words – a subset of the items from the CNRep – more accurately than a newly developed set of low-lexicality non-words, while the performance for the TD group was similar across the two non-word types. Meir and Armon-Lotem (2017) compared the performance on wordlike vs. non-wordlike items in a group of 88 bilingual Hebrew-Russian and 32 monolingual Hebrew speaking children between the ages of 5;7–6;7. They found that the bilingual group performed significantly poorer than the monolingual group on the non-wordlike stimuli, but not on the wordlike stimuli. In sum, the evidence suggests that there may be a negative effect of both bilingualism and DLD on the performance of non-wordlike stimuli, but not of word-like stimuli. This indicates that researchers should proceed with caution when choosing NWR items for maximal separation between DLD and TD groups in bilingual populations.

This section has presented an overview of how different characteristics inherent to NWR tasks or items may influence performance. The next section covers some child-internal factors that may influence performance on NWR tasks.

5.1.4 Participant-related factors influencing NWR performance

There are three areas of child-internal factors that have been extensively described in the literature: age effects, socioeconomic status (SES) and language exposure/vocabulary size.

5.1.4.1 Age effects

To investigate the effect of age on the performance on NWR, Chiat and Roy (2007) administered the Preschool Repetition Test (PRT) to a group of British 2;0–4;0-year-old children with typical language development (N=315) and a group of 2;6–4;0-year-old clinically referred children (N=168). The PRT is a test with both real word and non-word items. The TD children were divided into three age groups: 2;0–2;5, 2;6–2;9 and 3;0–4;0. The clinical group was also divided into three age groups but with slightly different cut-offs: 2;6–2;9, 3;0–3;5, 3;6–4;0. Analyses showed that performance increased with age in the TD group, but no significant difference was found between the two older age groups. In the clinical group scores generally increased with age, but contrary

to the TD group, there was a significant difference between the two older age groups. Radeborg et al. (2006) investigated non-word repetition in a group of 200 4–6-year-old Swedish speaking monolinguals. In a mixed ANOVA – analysing effects of gender, age and number of syllables in the non-words (2–5 syllables) – they found an effect of age which was significant between all three age groups (4;0–4;11, 5;0–5;11 and 6;0–6;11). In another Swedish study, Kalnak et al. (2014) investigated non-word repetition as a potential clinical marker for Swedish speaking 8–12-year-olds. They administered a set of non-words with 24 items, three and four syllables in length, to a group of children with DLD (N=61) and a control group (N=86). In the control group, there was generally better performance with increased age. By contrast, no significant age effect was found in the DLD group. Topbaş et al. (2014) investigated the performance on a Turkish NWR test on a group of 150 TD children (age 3–7). They found a significant age effect; an increased performance rate was found for items of all syllable lengths as age increased.

In sum, age effects on NWR tasks have been attested for typically developing preschool and school age children. For children at risk or with a diagnosis of DLD, the literature indicates that there may also be similar patterns, but the evidence base is more limited than for TD children.

5.1.4.2 NWR and SES

In a seminal paper, Dollaghan and Campbell (1998) addressed the issue of potential cultural bias in NWR tasks. They state on p. 1137:

> "Unless non-words are designed to ensure that they are equally unfamiliar to children with LI and LN, the poor repetition of children with LI could be attributed to their reduced language knowledge rather than to be a fundamental psycholinguistic deficit".[57]

Dollaghan and Campbell were particularly critical towards the widely used English CNRep, which had been shown in earlier studies to be heavily dependent on previous language knowledge (S. E. Gathercole et al., 1991). As a response to this, Dollaghan and Campbell (1998) constructed a new NWR task, the Non-word Repetition Test (NRT). The test was administered along with other norm-referenced language tasks to two groups of English-speaking children in the United States ranging in age between 5;8 to 12;2. The first group were in language therapy (N=44), and the second group had typical language development (N=41). The children were from predominantly low-SES backgrounds, as determined by parental educational levels and current occupation. The authors conclude that the NRT did reliably differentiate between the TD and the DLD group. Furthermore, a comparison between performance on the NRT and a spoken language composite from the Test of

[57] Here, 'LI' is short for Language Impairment, and 'LN' is short for Language Normal.

Language Development-2 (TOLD) revealed that the NRT could reliably distinguish between the TD and the DLD group, while the TOLD could not. The authors claimed that while the TOLD was biased in favour of children with majority backgrounds, the performance on the NRT was free from such cultural bias.

Since Dollaghan and Campbell's (1998) paper, a number of studies have investigated whether there is an effect of SES on the performance on other non-word repetition tasks. In the previously described study by Chiat and Roy (2007) of English-speaking 2;0–4;0-year-olds, the authors asked school heads of nurseries to make judgements on the SES of the children in their TD sample, based on the head's impression of parental background, occupation and home environment. The researchers found no effect of SES on the repetition test for the TD group. For the clinically referred sample in the same study, SES was operationalised as the father's occupation, income and education level. Analyses revealed that neither of these measures had any impact on the child's performance on the repetition task. It is important to note is that the PRT includes 1–3-syllable words and non-words. The non-words in these tasks should be considered as fairly language dependent, since the non-words are based on real words, with metatheses of phonemes or syllables (for example *banana*/*nabana*, *balloon*/*laboon*).

Engel, Santos and Gathercole (2008) aimed to explore to which extent different measures of working memory and language measures (vocabulary) were dependent on the nature of the child's language environment. They studied native Brazilian Portuguese children between 6;3–7;6 years (M = 6;11) with reported typical language development. The children were divided into two groups (N=20 in each group); one high and one low SES, based on an amalgamation of family monthly income, educational level as well as the occupation of the main caregiver. On average, the monthly family income in the high-SES group was ten times higher than in the low-SES group. The children in the two groups were matched for age, gender, and nonverbal ability. The children were then tested on a Brazilian Portuguese NWR that was based on the CNRep (S. E. Gathercole et al., 1994), but adapted according to Portuguese phonotactic rules, and adhering to Portuguese stress patterns. While the tasks that assessed receptive and expressive vocabulary clearly disadvantaged the children in the low-SES group, no significant differences were found between SES groups with respect to performance on the NWR task.

Kalnak et al. (2014) examined the performance on NWR relative to parental education in their sample of Swedish monolinguals (age 8–12 years). The parents were divided into three groups based on their level of education: elementary school (14%), upper secondary school (48%) and higher education (university studies, 38%). Although the parents of the children diagnosed with DLD had disproportionally lower education levels compared to the parents of the children in the control group, Kalnak et al. did not find a correlation

between the parents' educational level and the child's performance on the NWR task.

While the aforementioned studies have explored non-word repetition performance in relation to SES in monolingual populations, the association between SES and performance on NWR tasks has also been investigated in studies including both monolingual and bilingual groups.

Boerma et al. (2015) examined the performance on two different NWR tasks in a sample of monolingual and bilingual Dutch 4–7-year-old children (N=120) with and without a diagnosis of DLD. The children were divided into four groups: TD monolinguals (N=30), monolinguals with DLD (N=30), TD bilinguals (N=30), and bilinguals with DLD (N=30). The children in the bilingual groups had varying first languages, but all spoke Dutch as a second language. All children were tested on two NWR tasks, the Dutch version of Chiat's quasi-universal task (2015) as well as on a language-specific task with items of both high and low phonotactic probability (Rispens & Baker, 2012). The authors found that performance was not correlated with SES as operationalised by parental education on any of the two NWR tasks within any of the four groups. Therefore, SES was not investigated further. However, the BiTD group came from significantly lower SES backgrounds than the MoTD group did. This difference is important to keep in mind when interpreting the results, as there was no difference in performance between the two TD groups in performance on the quasi-universal NWR task, while the MoTD group performed significantly better than the BiTD group on the language-specific NWR task. Possible explanations for this result will be discussed in the next section.

To summarise, the articles reviewed in this section suggest that there is no direct influence of SES on performance on NWR. However, in the case of more language-like items, it seems that there may be an effect of SES and/or bilingualism. Several researchers have suggested that this effect is actually not a direct effect of SES or bilingualism, but rather mediated by vocabulary, which in turn is correlated with language exposure. Research on this topic will be discussed in the next section.

5.1.4.3 NWR, language exposure and vocabulary

As described in section 5.1.2, the relationship between NWR and vocabulary has been demonstrated in many studies investigating (primarily English-speaking) monolinguals (Coady & Evans, 2008). In this section, studies that have looked into the relationship between vocabulary size, language exposure, and performance on NWR tasks in bilinguals will be reviewed.

Thordardottir and Brandeker (2013) explored the effect of varying degrees of cumulative language exposure to English and French on the performance on a French and an English NWR task for a group of 84 5-year-old Canadian children. Sixteen were monolingual English, 19 were monolingual French, and the remaining 49 were bilingual with varying degrees of exposure to each

language. The French NWR task included 40 items, ranging between 2–5 syllables, contained no clusters and carried equal stress on each syllable. The English task used was the CNRep. Analyses showed that there was no correlation between performance on the French task and exposure to French. For the English task a different pattern appeared, where there was a significant correlation between the amount of exposure to English and performance on the NWR task. Comparing performance for children with different levels of exposure, it seemed that there was a threshold effect at around 30% exposure to English before the bilingual children could be expected to perform on par with their monolingual English speaking peers.

Gibson et al. (2015) administered two NWR tasks, one in English (the NRT) and one in Spanish, to 52 English-Spanish-speaking 5-year-olds. Half of the children were dominant in English, and the other half were dominant in Spanish, as reported by the parents. The researchers wanted to examine whether performance on the two NWR tasks was influenced by phonological patterns in each language, and whether performance was dependent on language experience. Furthermore, they wanted to explore whether the effect of language experience was the same in each language, and for all syllable lengths. Gibson et al. found that the children who were dominant in Spanish performed better than the English-dominant group on both NWR tasks. There was also an overall effect of task, where both groups performed better on the Spanish task than on the English task. The authors claimed that Spanish has more long words than English, and that the Spanish-dominant children performed better also on the English task because of transfer from Spanish to English.

The previous section mentioned the article by Boerma et al. (2015), who found that bilingual children with typical language development performed significantly worse than the MoTD group on a language-specific Dutch NWR task. The BiTD group had significantly lower SES than the MoTD group, but they also had lower scores on a receptive vocabulary test. Boerma et al. did not analyse this difference between the two TD groups statistically. However, the BiTD group may perform more poorly because they are at a disadvantage compared to the MoTD group in the language-specific task, since the bilingual children are likely to have smaller vocabularies in the majority language compared to the monolinguals.

Engel de Abreu (2011) examined the performance on a language-specific NWR task (adhering to the phonotactical rules of Luxembourgish) in 44 monolingual and bilingual Luxembourgish children (age range 5;9–6;8). While the bilinguals performed significantly lower than the monolinguals, the group difference disappeared when expressive vocabulary was added as a covariate.

Szewczyk et al. (2018) investigated the effects of both item-related and participant-related factors that could influence performance on a Polish NWR task in a group of 57 monolingual children between the age of 4;5 to 6;10. The

task was designed to include items that had high as well as low phonotactic probability, but all were phonotactically legal in Polish. Szewczyk et al. found that of the six participant-related predictors (age, receptive vocabulary, non-verbal IQ (NVIQ), sex, and parental educational level), receptive vocabulary was the predictor that could best explain performance on the NWR task, overshadowing all other participant-related predictors including NVIQ and age. Szewczyk et al. speculate that vocabulary size, a measure of vocabulary richness, may be functioning as a proxy for the range and depth of sublexical representations (i.e. patterns of phoneme combinations occurring at different frequencies in a given language), and thereby overshadowing the effect of age (a measure that is also correlated with vocabulary size) as a significant predictor.

In conclusion, there is evidence for a significant effect of language exposure on the performance on NWR tasks. However, this effect is dependent on the characteristics of the non-words. Performance is more affected the more language-like the items are (Boerma et al., 2015; Engel de Abreu, 2011). The findings of Thordardottir and Brandeker (2013) suggest that there may be a threshold effect, where children with very little exposure to one of their languages perform significantly worse than their peers who receive low-medium to high levels of exposure, but this effect only seems evident for tasks/items with a high level of language specificity. The work of Engel de Abreu (2011) as well as Szewczyk et al. (2018) suggests that the effect of exposure may actually be a proxy for the influence of sublexical knowledge, which in turn is directly related to vocabulary size.

5.1.5 Comparing different scoring methods

This section contains a review of different ways of scoring the NWR tasks. Scoring methods will be discussed in terms of feasibility (in the clinical setting), and accuracy (i.e. to which extent there is a difference between the accuracy of different scoring methods to distinguish between DLD and TD groups).

A handful of different approaches to scoring NWR tasks have been described in the literature. One common and straightforward way of scoring is *whole item correct vs. incorrect*. This scoring method is also called *binary*, because each item is scored as either correct (1 p) or incorrect (0 p). This approach is favoured by many authors and used in tasks such as the CNRep, as well as in many other studies (Abed Ibrahim & Hamann, 2017; Chiat & Roy, 2007; Engel et al., 2008; S. E. Gathercole et al., 1994; Meir & Armon-Lotem, 2017; Radeborg et al., 2006; Saiegh-Haddad & Ghawi-Dakwar, 2017; Szewczyk et al., 2018). Another, more fine grained, way of scoring is by counting percent phonemes correct (PPC), which is applied in the NRT and also used in other studies (Dollaghan & Campbell, 1998; Ellis Weismer et al.,

2000; Gibson et al., 2015).[58] Although the PPC scoring method gives a more fine-grained result, it is not necessarily the case that it is more informative in distinguishing between DLD and TD groups. Some studies comparing these different ways of scoring on the same data set will be reviewed next.[59]

Graf Estes et al. (2007) conducted a meta-analysis concerning the difference in performance in non-word repetition for TD children and children with a DLD diagnosis. The analysis included studies on monolingual English-speaking populations, and the majority of the studies used the CNRep or the NRT. Remember that the CNRep is scored whole item correct vs. incorrect, while the NRT is traditionally scored as PPC. Graf Estes et al. found that across studies, the CNRep reported higher effect sizes for distinguishing TD groups from DLD groups compared to the NRT. They hypothesised that the difference in effect size could in part be attributed to the different scoring methods. Therefore, they decided to test the two scoring methods on the same dataset. They collected data from a group of monolingual English-speaking children with DLD and an age-matched group (N=68) on the NRT. They found that the whole item correct vs. incorrect scoring produced lower scores overall, and also gave higher standard deviations compared to PPC. However, the effect size between the TD and DLD groups was twice as large when applying the PPC scoring method compared to the whole item correct vs. incorrect. The authors attributed the larger effect size for PPC scoring to differences in standard deviations between the groups, and concluded that using binary scoring should not result in a larger effect size, thus not providing an explanation for their original finding. Hence, effect size between TD and DLD groups may depend on a number of different factors, such as task type and group selection (such as age, and inclusionary vs. exclusionary criteria), in addition to which scoring method is utilised.

Kalnak et al. (2014) compared the binary scoring method and PPC for a group of 8–12-year-old Swedish-speaking children with a DLD diagnosis (N=61) and an age-matched control group (N=86). The results mirrored those of Graf Estes et al. (2007), where the binary scoring method produced lower overall scores and higher standard deviations. However, both scoring methods could reliably distinguish between the two groups, and contrary to Graf Estes et al.'s findings, the binary scoring method had a higher effect size compared to the PPC method in Kalnak et al.'s study. One can only speculate on the reason for the difference in effect size between the two studies, but one possible explanation may be that there was a ceiling effect in the NWR task in Kalnak et al.'s study when using PPC scoring. The mean score differed by

[58] Yet other studies score percent consonants correct (PCC), or percent vowels correct (PVC), see for instance Sundström et al. (2014).

[59] Regardless of the overall scoring method (binary or any variety of PPC), the majority of the studies mentioned in this section apply lenient scoring regarding minor articulation deviances and phonological processes that are consistent in the child's speech. Some studies also disregard errors of voicing (Abed Ibrahim & Hamann, 2017).

20% between the TD group and the DLD group with PPC scoring, while the score differed by 40% between the groups when using binary scoring.

Another study also examining the difference between scoring methods is Boerma et al. (2015), who administered a language-specific and a quasi-universal NWR task to 120 bilingual and monolingual Dutch-speaking children with and without a DLD diagnosis (30 in each group). They found that the two scoring methods were similar in distinguishing between the TD vs. DLD groups, although *specificity* was slightly better with the PPC scoring, while *sensitivity* was higher when applying the binary scoring method.[60] This was true for both the language-specific and the quasi-universal task. In general, Boerma et al. found that the binary scoring method gave similar results to the more time-consuming PPC scoring. Therefore, they proposed that the binary scoring should be used, since this approach is more feasible in a clinical setting.

In sum, the evidence seems to suggest that although PPC and other fine-grained methods for scoring NWR performance give overall higher scores, binary scoring methods may be just as accurate in separating TD and DLD groups.

5.1.6 Background literature: summary

The item-related factors reviewed were item length, item complexity, prosody, phonotactic probability and wordlikeness. With respect to item length (number of syllables), in general findings suggest that the level of difficulty increases with an increased number of syllables, which is reflected in lower accuracy rates. However, there is not always a significant difference in performance between TD children and children with a DLD diagnosis for every syllable length (and it also seems to vary by age). Similarly to item length, a higher degree of syllabic complexity (i.e. consonant clusters) leads to lower accuracy rates. The role of prosody has not been studied to the same extent as syllable count or complexity, but the research that exists suggests that in languages with variable stress patterns, features with a lower degree of saliency (e.g. prestressed syllables or non-words with Swedish acute tonal accent) are more difficult to repeat. Regarding phonotactic probability, children generally seem to perform better on items with high phonotactic probability. Some evidence also seems to suggest that children with a DLD diagnosis may perform disproportionally worse than their TD peers on items with low phonotactic probability, at least for some syllable lengths.

[60] The concepts *sensitivity* and *specificity* are frequently used in clinical studies, and may be helpful in evaluating the accuracy of a diagnostic test. Sensitivity refers to how accurately the test (e.g. NWR) picks up all individuals in a sample with a particular condition (i.e. DLD); while specificity refers to how accurately the test can identify all individuals in the same sample without the condition (i.e. TD) (Loong, 2003, pp. 716–717). A diagnostic test is only clinically informative if both sensitivity and specificity rates are sufficiently large.

Furthermore, word-like items have been shown to be more accurately repeated than non-wordlike items. Finally, some studies imply that there may be a negative effect of *both* bilingualism and DLD on the performance on non-wordlike stimuli.

The participant-related factors reviewed were age, SES, language exposure and vocabulary. Numerous studies report a general (and expected) pattern of better performance with increased age. With respect to SES, no direct effect can be attested. However, for more language-like NWR items or tasks, several studies have reported an effect of language exposure and/or vocabulary (where language exposure is likely to act as a proxy for vocabulary size). This means that the more language-like the items in a particular NWR task, the greater the risk of children with a smaller vocabulary (such as bilinguals or children from lower SES backgrounds) being disadvantaged.

The last section of this literature review concerned different ways of scoring NWR tasks. Findings from studies that have evaluated different scoring methods suggest that the binary scoring method is often as informative as percent phonemes correct or other fine-grained measures, as well as being less time-consuming. Therefore, binary scoring is preferred in a clinical setting.

5.2 Some considerations for selecting NWR tasks for the current study

In light of the findings from the literature review, this section contains a brief discussion about the rationale for selecting the NWR tasks that were included in the current study. First, the tasks should include items with a variety of length (number of syllables). The reason for this is that performance may be different between TD and DLD groups for items of some lengths, but not others (e.g. performance may be similar between groups on shorter items, but not on longer items). Likewise, the discriminatory power may not be the same for items with simple syllable structures and items with complex syllable structures. A variety of items of different syllable lengths increases the likelihood of items being included that best differentiate between TD and DLD. Furthermore, the tasks should include items with different levels of syllabic complexity, i.e. both items with simple CV structures, as well as items containing consonant clusters and codas.

In the current study, four different NWR tasks are employed to investigate phonological working memory in Arabic-Swedish-speaking bilinguals. The tasks vary in phonological setup and complexity. Three of these tasks were developed with multilingual assessment in mind, whereas one is Swedish language-specific. First, there are two versions of the quasi-universal non-word repetition task (Chiat, 2015). These tasks consist of NWR items that do

not adhere to the phonological rules of a specific language. The task contains items of 2–5 syllables, with a simple CV syllable structure and phonemes that are common in many languages. For each language version the items are selected from a list of candidates, avoiding any items that are real words in that language and items containing phonemes that do not exist in that language. For the current study, the quasi-universal task was adapted to Swedish (QU-Swe) by the author and to Arabic (QU-Ara) by Rima Haddad, PhD student of Linguistics. The third task is a task with 1–3 syllable items with a restricted phoneme inventory and mixed simple and complex syllables is used, the Non-word Repetition Task-Lebanese (NWRT-Leb; Abou Melhem et al., 2011). This task was developed to be used with multilingual Lebanese children. It was not constructed to adhere to the phonotactics of Arabic, but the phoneme inventory was controlled so that only phonemes that exist in Arabic are present in the items. Some items contain syllables with clusters or codas and others containing no clusters or codas in various positions. Finally, a language-specific task (LS-Swe) adhering to the phonological rules of Swedish is used (Radeborg et al., 2006). The task contains items of 2–5 syllables, with mixed simple and complex syllables, and a wide phoneme inventory with speech-sounds and stress patterns that are typical of Swedish. The NWR tasks are described in further detail in the Methods chapter (section 3.2.2).

The inclusion of the Swedish language-specific task (LS-Swe) allows for an investigation of whether language-specific knowledge (vocabulary skills) and exposure (length of exposure to Swedish) have an impact on repetition accuracy on this task. Also, including two language versions of the quasi-universal NWR task (Chiat, 2015), allows investigating whether accuracy is the same or different in the two versions of this task that is supposed to be equivalent across languages. Finally, the NWRT-Leb is a task with items shorter in length than in many other tasks (1–3 syllables), but where syllabic complexity is specifically probed.

5.3 Transcription and scoring

All NWR tasks were audio recorded for later transcription and analysis. The responses were transcribed phonemically by a native speaker of Swedish (LS-Swe and QU-Swe: the author of the present thesis) and Arabic (QU-Ara and NWRT-Leb: Rima Haddad, PhD student of Linguistics) respectively. Scoring was then conducted as follows. First, allowances were made for minor articulation deviances such as non-adultlike/indistinct pronunciation of /r/ and /s/.[61] Second, any phonological substitution processes that were consistent in

[61] These phonemes are challenging to articulate and may be difficult to pronounce for children in the target age and even some adults.

the child's speech were disregarded.[62] Third, errors of voicing (/p/ vs. /b/) and minor vowel deviations (e.g. /œ/ vs. /ø/) were also disregarded. However, major vowel substitutions such as substituting /a/ for /i/ were not allowed. Finally, any additions of syllables or phonemes before or after the otherwise correctly repeated item were also disregarded (i.e. children were not penalised for hesitation noises). After making these allowances, scoring was conducted according to the *whole item correct vs. incorrect* approach, where a child received 1 point for each correctly repeated non-word, and 0 points for any repetition containing an error.

5.4 Results

In this section, scores from the NWR tasks are reported in relation to age (section 5.4.1). Furthermore, the results are explored with respect to participant-related and item-related factors, and the performance on the two quasi-universal tasks are compared (section 5.4.2).

5.4.1 NWR performance and age

The following specific research questions are asked with respect to NWR performance and age:

- How does performance on the four NWR tasks develop from age 4 to 7?
- Are there differences in performance between tasks?
- What characterises the children performing low on NWR tasks?

5.4.1.1 Statistical analyses

The total scores for all age groups combined are reported in absolute scores as well as percentages. Since the number of items are different for the different tasks, overall task performance was compared for percent correctly repeated items.

The relationship between age and performance on the NWR tasks are investigated in two ways: first, via age groups using one-way ANOVAs, and second via age in months using Pearson correlation. For all statistical analyses, the level of significance was consistently set to $p < .05$ (two-tailed). The analyses were made in R (R Core Team, 2020).

[62] One example of such a phonological substitution process is fronting – substituting phonemes with a velar place of articulation with their corresponding alveolars, e.g. substituting /k/ with /t/ and /ŋ/ with /n/.

5.4.1.2 Total scores for all NWR tasks

In this section, the total scores for the language-specific Swedish task (LS-Swe), the quasi-universal Swedish task (QU-Swe), the quasi-universal Arabic task (QU-Ara), and the Non-word repetition task-Lebanese (NWRT-Leb) are reported. Table 5.1 shows the mean scores, standard deviations (SD), ranges, as well as mean scores and SDs in percentages (%) for all NWR tasks.

As can be seen in Table 5.1, performance on the LS-Swe task was the lowest, with a mean score of 54.7%. The two quasi-universal tasks were in the middle, with mean scores of 76.0% (QU-Swe) and 69.5% (QU-Ara). Mean scores for the NWRT-Leb task were proportionally the highest at 83.7%. There was considerable variation in performance for all tasks. At least one child obtained the maximum score in all tasks apart from the LS-Swe task, where the highest score was 22 (92%).

A one-way ANOVA showed that there was a significant difference in performance between the four tasks ($F(3,388) = 61.34$, $p < .001$, $\eta_p^2 = .322$). Post hoc analyses (with Bonferroni correction) revealed that the difference in performance was significant between all tasks (LS-Swe vs. QU-Swe: $p < .001$, LS-Swe vs. QU-Ara: $p < .001$, LS-Swe vs. NWRT-Leb: $p < .001$, QU-Swe vs. QU-Ara: $p < .05$, NWRT-Leb vs. QU-Ara: $p < .001$, NWRT-Leb vs. QU-Swe: $p < .01$).

Table 5.1. Scores for all ages combined (4;0–7;11). Language-specific Swedish task (LS-Swe), Quasi-universal Swedish task (QU-Swe), Quasi-universal Arabic task (QU-Ara), and the Non-word repetition task-Lebanese (NWRT-Leb).

	LS-Swe (max=24) N=97	**QU-Swe (max=16) N=98**	**QU-Ara (max=16) N=98**	**NWRT-Leb (max=30) N=99**
Mean	13.1	12.1	11.1	25.1
SD	4.0	2.2	2.8	4.2
Range	2–22	5–16	2–16	7–30
Mean (%)	54.7	76.0	69.5	83.7
SD (%)	16.6	14.0	17.2	14.0

In order to investigate whether there were correlations between performances on the different tasks, multiple Pearson correlation analyses were carried out between task pairs. As shown in Table 5.2, all tasks correlate with each other moderately to strongly.

Table 5.2. Correlation coefficients (Pearson) between NWR tasks. Language-specific Swedish task (LS-Swe), Quasi-universal Swedish task (QU-Swe), Quasi-universal Arabic task (QU-Ara), and the Non-word repetition task-Lebanese (NWRT-Leb).

	LS-Swe	QU-Swe	QU-Ara
LS-Swe	---	---	---
QU-Swe	.64***	---	---
QU-Ara	.62***	.62***	---
NWRT-Leb	.61***	.50***	.68***

Note. ***$p < .001$.

In the next section, age development will be investigated for each task separately.

5.4.1.3 Age development for each NWR task

In this section, scores are reported for each NWR task for each age group, via descriptive statistics (means, SDs, ranges, as well as means and SDs in percentages (%)). As shown in Table 5.3, and illustrated in Figure 5.3, the scores increase for all tasks between age groups (i.e. the 5-year-olds score better than the 4-year-olds, etc.), and the ranges become smaller, except for the 7-year-olds on the QU-Ara task. Notably, mean scores are consistently higher in the QU-Swe task than in the QU-Ara task, but the difference is particularly large (2.2 points) among the 4-year-olds.

Next, development with age will be investigated separately for each NWR task. After that, some children who scored relatively low compared to their peers in the same age group will be discussed in relation to the background information provided by their parents in section 5.4.1.4.

Table 5.3. Means, standard deviations (SD), ranges, means and standard deviations in percentages (%) for each task by age groups. Language-specific Swedish task (LS-Swe), Quasi-universal Swedish task (QU-Swe), Quasi-universal Arabic task (QU-Ara), and the Non-word repetition task-Lebanese (NWRT-Leb).

	4-year-olds	**5-year-olds**	**6-year-olds**	**7-year-olds**
LS-Swe (Max=24)	N=20	N=24	N=29	N=24
Mean	10.5	12.3	14.0	15.0
SD	3.7	3.5	3.9	3.4
Range	2–16	6–18	6–21	8–22
Mean (%)	43.8	51.1	58.4	62.8
SD (%)	15.5	14.6	16.3	14.4
QU-Swe (Max=16)	N=21	N=24	N=29	N=24
Mean	11.0	12.3	12.2	12.9
SD (%)	2.5	2.0	2.4	1.8
Range	5–14	8–15	8–16	10–16
Mean (%)	68.9	76.8	76.6	80.8
SD (%)	15.4	12.7	15.1	11.0
QU-Ara (Max=16)	N=22	N=23	N=29	N=24
Mean	8.8	11.7	11.7	12.0
SD	3.1	2.0	2.2	2.6
Range	2–14	6–14	7–16	4–16
Mean (%)	54.9	72.9	73.2	75.1
SD (%)	19.4	12.7	13.7	15.9
NWRT-Leb (Max=30)	N=22	N=24	N=29	N=24
Mean	21.3	25.3	26.2	27.0
SD	4.9	3.8	3.4	2.6
Range	7–28	15–30	19–30	19–30
Mean (%)	71.0	84.5	87.4	89.8
SD (%)	16.3	12.7	11.3	8.7

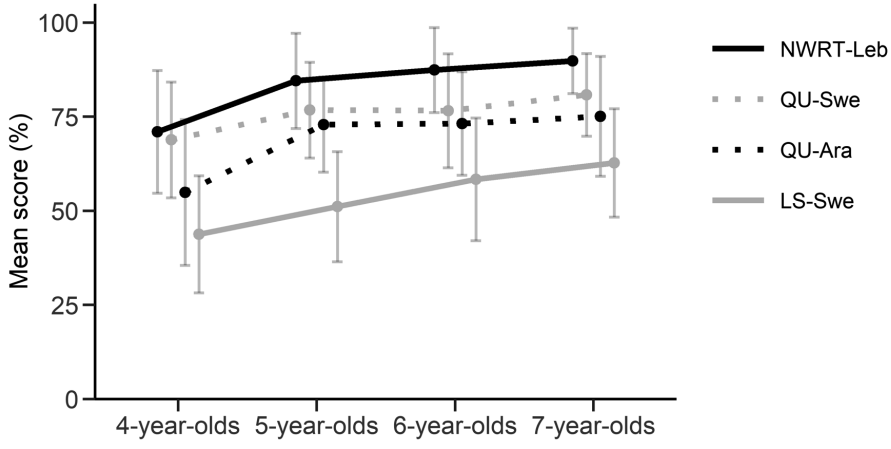

Figure 5.3. Line chart of mean scores (%) by age groups for the NWRT-Leb task, the QU-Swe task, the QU-Ara task and the LS-Swe task. Error bars indicate +/- 1 SD.

5.4.1.3.1 The language-specific Swedish (LS-Swe) task and age

As can be seen in Table 5.3, scores for the LS-Swe task increased with age. A one-way ANOVA showed that there was a significant difference between the age groups on performance on the LS-Swe task ($F(3,93) = 6.650$, $p < .001$, $\eta_p^2 = .177$). A post hoc analysis (with Bonferroni correction) revealed that the difference between the 4-year-olds and the 6-year-olds was significant ($p < .01$), as well as between the 4-year-olds and the 7-year-olds ($p < .001$). No other groups differed significantly from each other. The lowest score was found in the youngest age group, and the highest in the oldest age group. The variation was similar across age groups, as determined by similar SDs in all age groups.

Figure 5.4 plots scores on the LS-Swe task against age in months. There is a moderate positive correlation (Pearson) between age in months and scores on the LS-Swe task (N = 97, $r = .422$, $p < .001$).

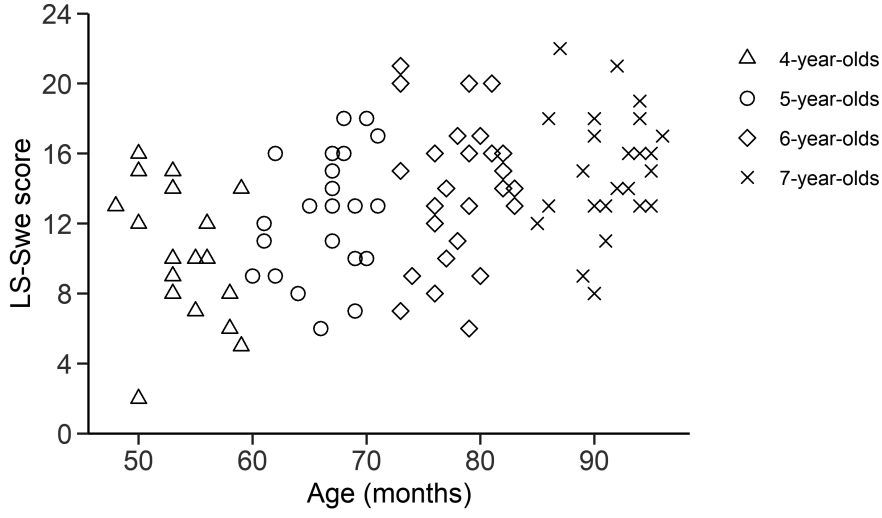

Figure 5.4. Scatterplot of scores on the LS-Swe task and age in months.

Of the 97 children who did the LS-Swe task, a substantial part of the group, 39%, scored 50% or lower (12 points).[63] This group consisted of a majority of the 4-year-olds, as well as many of the 5-year-olds and the 6-year-olds. Four 7-year-olds were also found in this group. The majority of the children, 60%, performed between 50% and 90%. Only one child, BiAra7-23, scored above 90% (22 points).

5.4.1.3.2 The quasi-universal Swedish (QU-Swe) task and age

As can be seen in Table 5.3, scores for the QU-Swe task increased slightly between ages four and five. There was no difference in means between the five- and 6-year-olds, and only a slight increase between the 6-year-olds and the 7-year-olds. A one-way ANOVA showed that there was a significant difference between the age groups on performance on the QU-Swe task ($F(3,94) = 2.952$, $p < .05$, $\eta_p^2 = .086$). A post hoc test (with Bonferroni correction) revealed that there was only a significant difference between the 4-year-olds and the 7-year-olds ($p < .05$). No other age groups differed significantly from each other. The lowest score was found in the youngest age group. The highest performance was found in the two oldest age groups, two 6-year-olds and two 7-year-olds received the maximum score (16 points). The variation was similar across all age groups.

Figure 5.5 shows scores on the QU-Swe task plotted against their age in months. A Pearson correlation showed that there is a weak positive correlation

[63] Two 4-year-olds, BiAra4-14 and BiAra4-18, did not do the LS-Swe task. In both cases, the LS-Swe task was the very first task in the first session, they were perceived as a bit shy and unwilling to cooperate. However, both children completed the vocabulary and the narrative tasks that were also part of the test procedure.

between age in months and scores in the QU-Swe task (N = 98, r = .273, p = .006).

Out of the 98 children who did the QU-Swe task, ten children, mainly 4-year-olds and 6-year-olds, scored at 50% or below.[64] No 7-year-olds scored that low. The majority of all children, 78%, scored between 50 and 90% correct. Twelve percent, mainly 6-year-olds and 7-year-olds, scored above 90%.

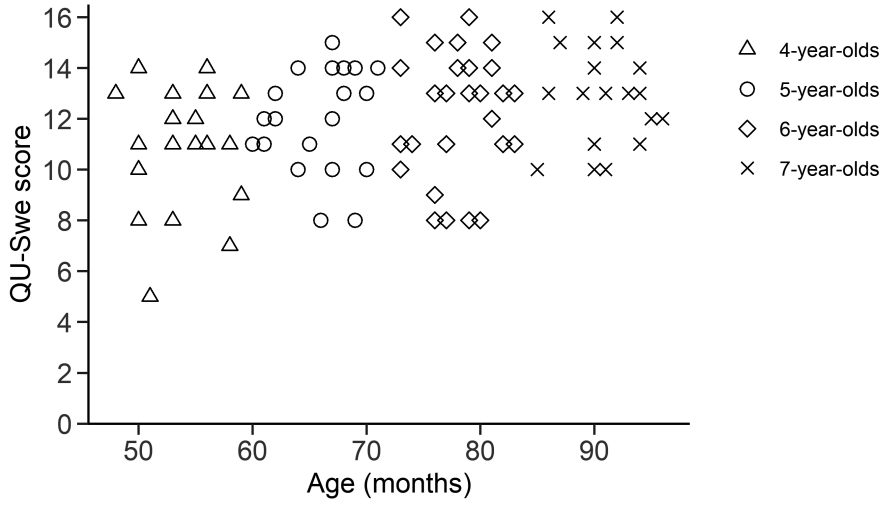

Figure 5.5. Scatterplot of scores on the QU-Swe task and age in months.

5.4.1.3.3 The quasi-universal Arabic (QU-Ara) task and age

As can be seen in Table 5.3, scores for the QU-Ara task increased between age 4 and age 5. A one-way ANOVA showed that there was a significant difference between the age groups in performance on the QU-Ara task (F(3,94) = 8.480, p < .001, η_p^2 = .213). A post hoc analysis (with Bonferroni correction) confirmed the pattern suggested in Table 5.3, as there was only a significant difference between the youngest age group and the three older age groups (4-year-olds vs. 5-year-olds: p < .01; 4-year-olds vs. 6-year-olds: p < .001; 4-year-olds vs. 7-year-olds: p < .001). No other groups differed significantly from each other. The lowest score was found in the youngest age group. The highest performance was found in the two oldest age groups, with children among both the 6-year-olds and the 7-year-olds scoring maximum (16 points). However, the range was rather large in the oldest age group, since one 7-year-old scored very low.

[64] One child (BiAra4-18) did not do the QU-Swe task, due to shyness (same as for the LS-Swe task). However, she did complete the other tasks (vocabulary and narrative) that were part of the test procedure.

Figure 5.6 shows scores on the QU-Ara task plotted against age in months. A Pearson correlation showed that there is a moderate positive correlation between age in months and scores on the QU-Ara task (N = 98, r = .400, p < .001).

Ninety-eight children did the QU-Ara task.[65] Out of these, 15 children scored at 50% or below. The majority of them (9) were 4-year-olds, but there were children from all age groups represented here, including two 7-year-olds. Five children, three 6-year-olds and two 7-year-olds, performed at 90% or better.

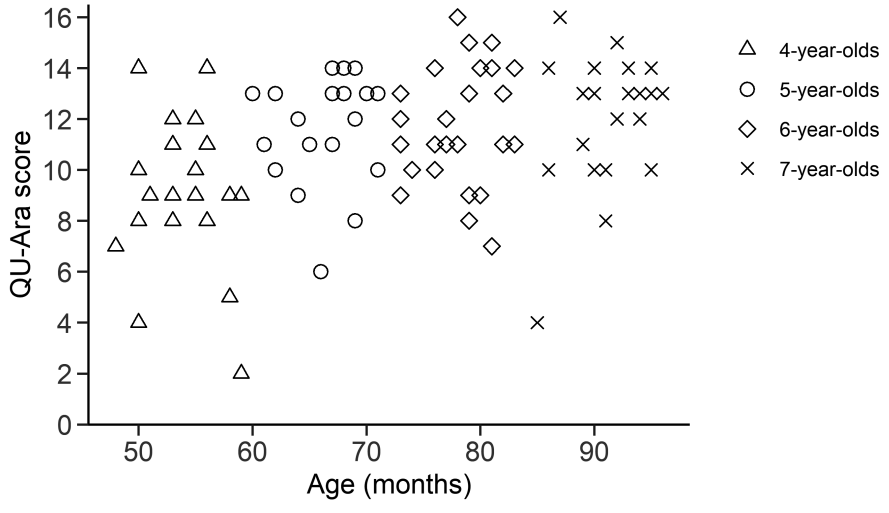

Figure 5.6. Scatterplot of scores on the QU-Ara task and age in months.

5.4.1.3.4 The Non-word repetition task-Lebanese (NWRT-Leb) and age

As can be seen in Table 5.3, scores for the NWRT-Leb increased from the 4-year-old group to the 5-year-old group, but did not increase substantially between other age groups. As determined by a one-way ANOVA, there was a significant difference between the age groups on performance on the NWRT-Leb (F(3,95) = 10.63, p < .001, η_p^2 = .251). A post hoc analysis (with Bonferroni correction) revealed that there was a significant difference between the 4-year-olds and the three older age groups (4-year-olds vs. 5-year-olds: p < .01; 4-year-olds vs. 6-year-olds: p < .001; 4-year-olds vs. 7-year-olds: p < .001). No other groups differed significantly from each other. The lowest score was found in the youngest age group. There were children in the three oldest age groups who all scored maximum (30 points).

Figure 5.7 shows scores on the NWRT-Leb plotted against age in months. Investigating the relationship between age in months and scores on the NWRT-Leb task with a Pearson correlation, we see that there is a moderate

[65] One child, BiAra5-25, did not do the QU-Ara task due to experimenter error.

positive correlation between age and scores in the NWRT-Leb task (N = 98, $r = .445$, $p < .001$).

All 99 children did the NWRT-Leb. Only three children, two 4-year-olds and a 5-year-old, scored below 50% on the task. The majority, 55%, scored between 50 and 90%. A substantial group of children, 41%, scored at 90% or above. Children from all age groups were represented here, including two 4-year-olds.

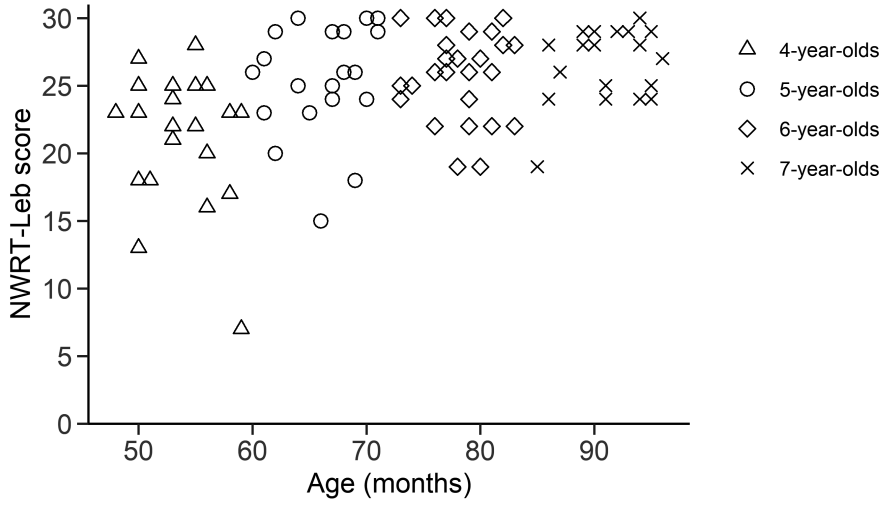

Figure 5.7. Scatterplot of scores on the NWRT-Leb and age in months.

5.4.1.4 Low performers

In this section, the children who received low scores (z-scores below –1.25) on any of the NWR task are described.[66] They are discussed in light of information from the parental questionnaire as well as their vocabulary skills in both languages (CLTs). A table with all children who received a z-score of –1.25 in at least one NWR task is available in the Appendix (Table A5.1), with z-scores for all NWR tasks and CLT tasks.

Many children received z-scores below –1.25 in one NWR task only. In some instances they had low performance in one or two vocabulary tasks as well.[67] For some other children, a low score on one NWR task coincided with the parents reporting some reason for concern regarding language

[66] In the current study, a z-score below –1.25 is considered as potentially clinically informative threshold. See Chapter 6, section 6.2.1, for a more in depth-description of z-scores and cut-offs.

[67] The children who also scored –1.25 or below on vocabulary tasks were: BiAra4-02 (Arabic comprehension: –1.64, Arabic production: –2.02), BiAra4-14 (Arabic comprehension: –1.36), BiAra5-06 (Arabic comprehension: –2.98, Arabic production: –1.81), BiAra6-03 (Arabic comprehension: –1.55), BiAra6-08 (Arabic production: –1.47), BiAra6-10 (Arabic comprehension: –2.20, Arabic production: –1.40), BiAra6-29 (Swedish comprehension: –1.96), and BiAra7-05 (Swedish comprehension: –3.09, Swedish production: –2.44).

development, such as having late language development in one language (BiAra4-10, BiAra5-06, BiAra6-19), the first word or word combination appearing late (BiAra6-22), having had anxiety for their child's language development at some point (BiAra5-06, BiAra6-10), having language problems in the family (BiAra6-19), or having consulted an SLP at some point (BiAra5-06). A low result on one task alone need not necessarily be a cause for concern; however, children who score low on two or more tasks are worth looking into more in detail.

Among the 4-year-olds, two children (BiAra4-24 and BiAra4-11) received z-scores below –1.25 on three NWR tasks. As previously mentioned, BiAra4-24 was reported to have had temporary hearing problems that were now resolved. The parents reported no other warning signs associated with late or deviant language development. CLT scores were above average in Arabic, and well above average in Swedish. For BiAra4-11 there was nothing in the parental questionnaire that indicated any problems relating to language difficulties; she was reported to be an early talker, learning languages quickly, having no heredity for language problems, and normal hearing. CLT scores were within one SD below the mean (z-scores between –0.21 and –0.74) for both tasks in both languages.

Among the 5-year-olds there were two children who had z-scores below –1.25 on all NWR tasks (BiAra5-01 and BiAra5-03). BiAra5-01 was reported by the parents to have normal language development in both languages, having produced the first word and word combination within the expected time frame. The parents stated that they had concern about the child's language development, because they did not speak Swedish very well (this is interpreted as the parents being concerned because they could not provide their child with high-quality input in Swedish). Concerning CLT, the Swedish scores were close to the mean for the child's age group, however the z-score for Arabic production was below –1.25. Regarding BiAra5-03, parents reported their child to have a normal language development in both languages, however, the first word appeared a bit later than expected (at 20 months). The parents report concern, but not for the child's language development per se, but rather because of the move to Sweden and the implications for their child growing up as a bilingual (being different from the majority). No other warning signs associated with a late or deviant language development were reported. CLT scores were within one SD below the mean (z-scores between –0.99 to –0.14) for both tasks in both languages.

Within the 6-year-old-group there was one child who had z-scores below –1.25 on two NWR tasks (BiAra6-05), and three who had z-scores below –1.25 on three tasks (BiAra6-06, BiAra6-23, BiAra6-26). BiAra6-05 was reported by the parents to having been a late talker, saying his first word at 35 months. However, looking at the CLT scores for this child, he performed 0.5–1 SDs (z-scores between 0.55–1.00) *above* the mean on comprehension and production in both languages. The parents of BiAra6-06 reported that he had

normal language development in both languages. The first word appeared at 18 months, and the first word combination at 30 months. CLT scores in Arabic were slightly above the mean, but the Swedish production z-score was below –1.25. There was nothing in the background information of BiAra6-23 that is associated with late or deviant language development; the parents report that language development was normal in Arabic and even early in Swedish. The first word appeared at 10 months, and the parents do not report any concern or having consulted an SLP. Z-scores for Arabic comprehension were well above average for his age, while Arabic production and the two Swedish scores were slightly below average (–0.41 to –0.72). BiAra6-26 has been mentioned in the Vocabulary and Methods chapters, having a brother with a DLD diagnosis, and the parents also report that his language development was late in both languages. As mentioned in Chapter 4 (Vocabulary), he also scored low on all vocabulary tasks, both comprehension and production in both languages, with scores well below the mean for comprehension in both languages (z-scores between –2.33 to –0.79).

There were two 7-year-olds who had z-scores below –1.25 on two tasks (BiAra7-16 and BiAra7-21), and one 7-year-old who had z-scores below –1.25 on three tasks (BiAra7-19). There was nothing in the parental questionnaires associated with late or atypical language development; all three children were reported to have normal or even early language development, the first word and word combination appeared within the expected time frame, and no parents reported any concern, language problems in the family or having consulted an SLP. All children were also reported to have normal hearing. BiAra7-19 had CLT scores slightly above average, while BiAra7-21 had scores slightly below average in comprehension and production in both languages. BiAra7-16 had z-scores below –1.25 for both comprehension and production of Swedish.

5.4.1.5 NWR scores and age development: summary

In this section, a summary of the age development in the different NWR tasks is provided. For all tasks, age effects were investigated in two ways. First, a one-way ANOVA was conducted in order to investigate whether there were significant differences between age groups. Second, a Pearson correlation was conducted between scores and age in months for each task.

Investigating the age development between age groups with ANOVAs revealed that there were differences between age groups for all tasks. However, post hoc analyses showed that there were only significant differences between the 4-year-olds and the three older age groups (QU-Ara and NWRT-Leb), between the 4-year-olds and the two oldest age groups (LS-Swe), and between the 4-year-olds and the oldest age group (QU-Swe). No other age groups differed significantly from each other. Concerning development with age in months, correlation analyses revealed a moderate positive correlation between age and NWR performance for three tasks (LS-

Swe, NWRT-Leb and QU-Ara), and a weak positive correlation between age and NWR performance for the QU-Swe task. The weaker age effect in the QU-Swe task can be explained by the fact that performance was already high among the youngest children.

There were differences in the proportion of the children who scored high or low in each task, reflecting different overall difficulty levels of the tasks. For instance, there was a striking difference between the LS-Swe task and the NWRT-Leb, where 39% of the children scored below 50% on the LS-Swe task, but only 3% did so on the NWRT-Leb). The reverse pattern emerged when investigating the proportion of children who scored 90% or better; only one child did so on the LS-Swe task, but 41% did so on the NWRT-Leb. For the two quasi-universal tasks, most children scored between 50–90% correct, with fewer children scoring below 50% or above 90% (QU-Swe: 12% scored > 90%, 10% scored < 50%; QU-Ara: 5% scored > 90%, 15% scored < 50%).

Children who scored low on one or more NWR tasks were described in light of the background information provided by the parents in the parental questionnaire. In some instances, a low score on one or more NWR tasks coincided with parents reporting cause for concern regarding language development, such as having a late language development, first words or word combinations appearing late, hearing problems, having consulted an SLP, etc. Although scoring low on one task alone should not be a cause for concern, some children scored low on multiple NWR tasks, as well as scoring unexpectedly low on one or more CLT tasks compared to their age peers (considering their language exposure patterns). These children may be candidates for undiagnosed DLD, and will be discussed further in Chapter 6 (the clinical study).

5.4.2 NWR performance in relation to task, items, and previous language experience

As described in the literature (see section 5.1), a number of item-related as well as participant-related factors have been shown to influence performance on NWR tasks. In this section, the aggregated effect of a number of participant-related as well as item-related factors on the probability of correct repetition will be investigated. These factors were type of task, item length (number of syllables), presence of consonant clusters, and for the LS-Swe and QU-Swe tasks length of exposure to Swedish, estimated daily exposure to Swedish and Swedish (CLT) vocabulary scores.

The NWRT-Leb was not included in these analyses, since it was not comparable with the other tasks in some important aspects. The NWRT-Leb items were generally shorter than the other tasks (1–3 syllables vs. 2–5 syllables), thus there would be a 100% overlap between task and item length for items with one syllable (NWRT-Leb) as well as four and five syllables

(QU tasks, LS-Swe task). The specific research questions asked in this section were the following:

- What are the effects of task (LS-Swe vs. QU-Swe), item length (number of syllables), presence of consonant clusters, exposure to Swedish, SES and Swedish vocabulary on the performance on the LS-Swe and the QU-Swe tasks?
- What are the effects of type of task (QU-Swe vs. QU-Ara) and item length (number of syllables) on the performance on the QU-Swe and the QU-Ara tasks?

5.4.2.1 Statistical analyses

First, exploratory analyses were conducted on the LS-Swe task (in which the items were constructed with adherence to Swedish phonotactics) and the QU-Swe (in which the items were constructed to be quasi-universal) in order to investigate whether there was a correlation between performance and (1) length of exposure to Swedish, (2) current daily exposure to Swedish, (3) Swedish (CLT) vocabulary, and (4) SES. The same analyses were conducted on both tasks (LS-Swe and QU-Swe) in order to investigate whether patterns were the same or different for the two tasks, thus forming a comparison between language-like (LS-Swe) and non-language-like (QU-Swe) test items.

Next, participant-related and item-related effects on accuracy of repetition were analysed statistically by using logit mixed-effects models, in R (R Core Team, 2020) using the function *glmer* from the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). This type of logistic regression model is suitable for repeated measurement data, when observations are not independent from each other (e.g. when participants and test items are occur more than once in a data set). The dependency structures are accounted for in the mixed models by so-called *random effects*.[68] In mixed-effect models, independent variables are referred to as *fixed effects*. The dependent variable in all analyses was a categorical (i.e. binary) variable, accuracy, where the data used was the participants' answers to all items (where every answer was coded as either correct or incorrect). The number of answers (total amount of data) was 2 328 for LS-Swe (97 participants x 24 test items), 1568 for QU-Swe (98 participants x 16 test items), and 1 568 for QU-Ara (98 participants x 16 test items). Hence, the analyses investigated which of the independent variables could significantly predict whether an answer was correct or not.

[68] There are two types of random effects: random intercepts (that takes into account estimates for individual measurements) and random slopes (that takes into account that the measured effect may be different for individual measurements).

The non-categorical variables (i.e. age, item length, Swedish vocabulary scores and length of exposure to Swedish) were standardised prior to modelling.[69]

Model 1 investigated the effects of task for the LS-Swe vs. QU-Swe, namely language-like vs. language-unlike test items, non-word length (number of syllables), Swedish vocabulary scores and length of exposure to Swedish. Model 2 investigated the effects of consonant clusters, non-word length and Swedish vocabulary scores for the LS-Swe task. Model 3 investigated the effects of task (QU-Swe vs. QU-Ara) test items and non-word length (number of syllables). For all models, random intercepts were included for participants and test items, since they are repeated many times in the data set (Baayen, Davidson, & Bates, 2008). Additionally, for models comparing the accuracy of two different tasks (Model 1 and Model 3), by-participant random slopes were included for task.[70] Age was included as a control variable in all analyses.

For each model, the following is reported. For fixed effects, the coefficient (coef.), standard error (SE), z-value and p-value are reported. Variance (s^2) and standard deviations (SD) are reported for random effects, and additionally, a correlation estimate between the slope and the intercept is reported for models that include random slopes. Finally, pseudo-R-squared (obtained by using the *r.squaredGLMM* function in the *MuMIn* package (Bartoń, 2020) is reported for each model. The *marginal* R^2 expresses the amount of variance that is explained by the fixed effects alone, while the *conditional* R^2 expresses the amount of variance explained by the full model, including random effects. For all statistical analyses, the level of significance was consistently set to $p < .05$ (two-tailed).

5.4.2.2 Swedish tasks: effect of task, number of syllables, exposure to Swedish and Swedish vocabulary

The exploratory analyses showed that here were positive correlations between performance on both tasks (LS-Swe and QU-Swe) and length of exposure to Swedish (LS-Swe: N = 96, $r = .34$, $p < .001$; QU-Swe N = 97, $r = .26$, $p < .05$) and Swedish vocabulary comprehension (LS-Swe: N = 97, $r = .46$, $p < .001$; QU-Swe N = 98, $r = .37$, $p < .001$), but no relationship between performance on the NWR tasks and daily exposure to Swedish (LS-Swe: N = 96, $r = .02$, $p = .82$; QU-Swe N = 98, $r = –.08$, $p = .44$) nor SES (parental education; LS-

[69] Standardising a variable means that it is rescaled to have a mean of 0, and a standard deviation of 1. Standardising may be helpful for making variables more comparable if the scales are of different magnitude (in this case, number of syllables: 2–5, age in months: 48–96, length of exposure in months: 7–95 and Swedish vocabulary comprehension: 0–60 points) (Winter, 2019, p. 89).

[70] This was done in order to account for individual variation between particiants with respect to their performance on the two tasks.

Swe: N = 96, r = .03, p = .75; QU-Swe N = 98, r = –.05, p = .65).[71] Scatterplots are available in the Appendix (Figure A5.1–A5.4). The correlation between length of exposure to Swedish and Swedish vocabulary and performance on the LS-Swe task and the QU-Swe task could potentially be related to age (since both length of exposure to Swedish and vocabulary scores generally increased with age). Therefore, the effect of length of exposure to Swedish and Swedish vocabulary was further analysed in multivariate analyses.

Figure 5.8 shows the accuracy for the LS-Swe and the QU-Swe tasks by item length (number of syllables). Accuracy was analysed for the two Swedish tasks (LS-Swe vs. QU-Swe) and non-word length (number of syllables), and the interaction between the two, as well as Swedish vocabulary and the interaction with task, and length of exposure to Swedish and interaction with task. The result from the statistical analysis is presented in Table 5.5.[72]

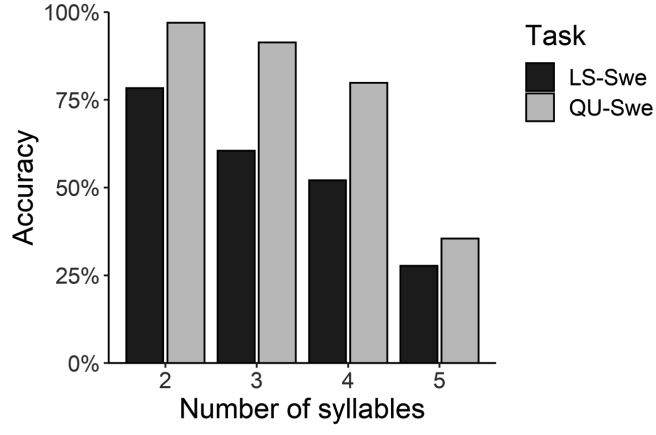

Figure 5.8. Accuracy (% correct answers), LS-Swe (language-specific Swedish task) and QU-Swe (quasi-universal Swedish task) by number of syllables.

[71] Separate analyses were conducted for vocabulary comprehension scores and production scores, with the same overall results. Therefore, only analyses with the comprehension scores are reported here.

[72] An analysis that included vocabulary comprehension instead of production showed the same overall results.

Table 5.5. Model 1: Accuracy, Swedish NWR tasks (LS-Swe and QU-Swe).

Model Summary				
Random effects	***(s²)***	***SD***	***corr***	
Participant (intercept)	.64	.80		
Task LS-Swe (slope)	.22	.47	.25	
Item (intercept)	1.20	1.10		
Fixed effects	***Coef.***	***SE***	***z***	***p***
Intercept	.31	.25	1.27	.204
Age (months)	.29	.12	2.48	.013*
Task (LS-Swe vs. QU-Swe)	1.76	.38	4.63	<.001***
Length	–1.11	.23	–4.75	<.001***
Swe_vocab	.31	.14	2.25	.025*
LoE_Swe	.06	.13	.51	.609
Task (LS-Swe) x Length	–.78	.38	–2.06	.039*
Task (LS-Swe) x Swe_vocab	.09	.14	.63	.526
Task (LS-Swe) x LoE_Swe	–.07	.14	–.51	.608
Model evaluation	***Marginal R^2***		***Conditional R^2***	
	.38		.61	

Note. *p < .05, ***p < .001. Logit mixed effects model with random effects: random intercepts for participant and test item, and by participant random slopes for task. Model fit with maximum likelihood (Laplace approximation). The reference level for the categorical variable is the first category. The values have been rounded off to two decimals. Length = item length (number of syllables), Swe_vocab = Swedish vocabulary comprehension, LoE_Swe = Length of exposure to Swedish (months).

The analysis showed significant effects of task, with higher accuracy on QU-Swe, and of item length, with lower accuracy for items with more syllables. There was also an interaction between task and item length. Accuracy rates dropped for both tasks with increased number of syllables. However, the difference between accuracy rates for four syllables and five syllables were larger for the QU-Swe task than for the LS-Swe task. Accuracy rates for QU-Swe items with 2–4 syllables were above 80%, while accuracy for five syllable items was 35%. For the LS-Swe task, accuracy rates for items with 2–4 syllables was between 78–52%, while accuracy for the five-syllable-items was 28%. As can be seen in Figure 5.8, the decrease in scores for items with five syllables is steeper for QU-Swe than for LS-Swe, which is likely to drive the interaction between task and item length. There was no effect of length of exposure to Swedish; however, there was a positive effect of Swedish vocabulary. Furthermore, there was no interaction between task and Swedish vocabulary. That is, the positive effect of higher Swedish vocabulary scores was not significantly better for the language-specific items.

5.4.2.3 Swedish language specific-task: effect of clusters, number of syllables and Swedish vocabulary

Figure 5.9 shows the accuracy for the LS-Swe task by item length (number of syllables) and presence vs. absence of consonant clusters. An analysis was conducted of accuracy on the LS-Swe task by presence of consonant clusters, item length (number of syllables), and Swedish vocabulary scores, as well as the interaction between clusters and syllable length, vocabulary scores and clusters and vocabulary scores and syllable length. The result from the statistical analysis is presented in Table 5.6.[73]

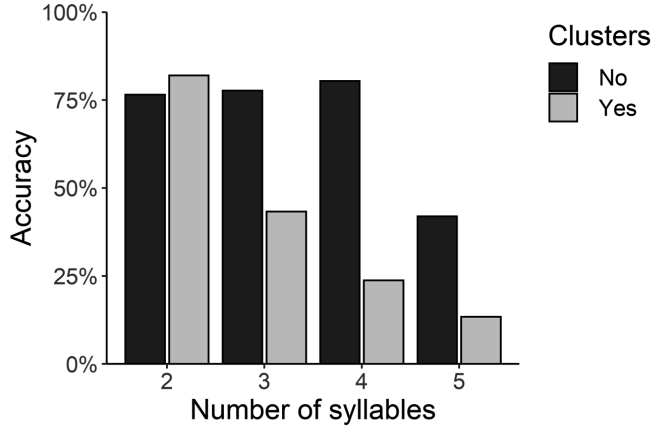

Figure 5.9. Accuracy (% correct answers), LS-Swe (language-specific Swedish task) items with and without consonant clusters by number of syllables.

[73] A separate model that included vocabulary comprehension instead of production yeilded the same overall results.

Table 5.6. Model 2: Accuracy, LS-Swe (language-specific Swedish) NWR task.

Model summary				
Random effects	***(s²)***	***SD***		
Participant (intercept)	.63	.80		
Item (intercept)	.75	.87		
Fixed effects	***Coef.***	***SE***	***z***	***p***
Intercept	1.14	.27	4.26	<.001***
Age (months)	.28	.11	2.46	.014*
Clusters (no vs. yes)	–1.70	.38	–4.52	<.001***
Length	–.68	.25	–2.74	.006**
Swe_vocab	.38	.12	3.13	.002**
Clusters (no) x Length	–.79	.38	–2.08	.038*
Clusters (no) x Swe_vocab	–.03	.06	–.29	.77
Length x Swe_vocab	–.05	.06	–.81	.42
Model evaluation	***Marginal R^2***		***Conditional R^2***	
	.34		.54	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Logit mixed effects model with random effects: random intercepts for participant and test item. Model fit with maximum likelihood (Laplace approximation). The reference level for the categorical variable is the first category. The values have been rounded off to two decimals. Length = item length (number of syllables), Swe_vocab = Swedish vocabulary comprehension.

The analysis showed a significant effect of length, with lower accuracy rates for items with increased number of syllables, and an interaction between clusters and item length. As can be seen in Figure 5.9, accuracy rates for items with clusters were on par with those without clusters at two syllables, but they dropped with an increased number of syllables. By contrast, accuracy for items without clusters was slightly above 75% for items with two to four syllables, but dropped to below 42% for items with five syllables. There was also a positive effect of Swedish vocabulary scores. The control variable age was also significant, with higher accuracy rates with increased age. There were no interactions between clusters and vocabulary, or item length and vocabulary.

5.4.2.4 Quasi-universal tasks: effect of task and number of syllables

Finally, accuracy was analysed for the two quasi-universal tasks (QU-Swe vs. QU-Ara) and non-word length (number of syllables), and the interaction

between the two (as shown in Figure 5.10). The results from the statistical analysis are presented in Table 5.7.[74]

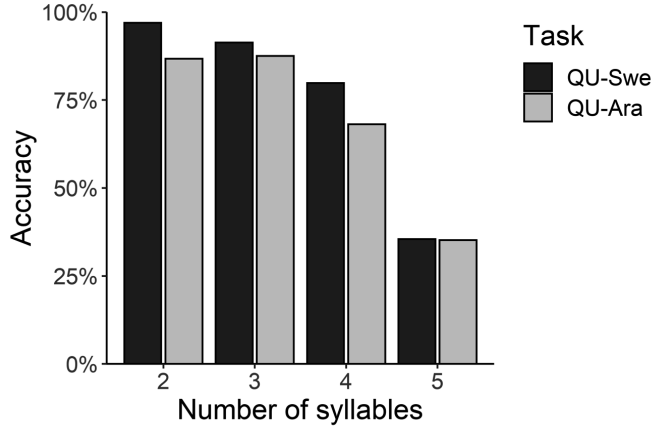

Figure 5.10. Accuracy (% correct answers), QU-Swe (quasi-universal Swedish task) and QU-Ara (quasi-universal Arabic task) by number of syllables.

Table 5.7. Model 3: Accuracy, quasi-universal NWR tasks (QU-Swe and QU-Ara).

Model summary				
Random effects	***(s^2)***	***SD***	***corr.***	
Participant (intercept)	.88	1.10		
Task QU-Ara (slope)	.25	.50	.08	
Item (intercept)	.64	.80		
Fixed effects	***Coef.***	***SE***	***z***	***p***
Intercept	1.22	.23	5.20	<.001***
Age (months)	.46	.11	4.12	<.001***
Task (QU-Ara vs. QU-Swe)	.86	.32	2.68	.007**
Length	–1.24	.22	–5.73	<.001***
Task (QU-Ara) x Length	–.65	.32	–2.04	.04*
Model evaluation	***Marginal R^2***		***Conditional R^2***	
	.37		.58	

Note. *$p < .05$, **$p < .01$, ***$p < .001$. Logit mixed effects model with random effects: random intercepts for participant and test item, and by-participant random slopes for task. Model fit with maximum likelihood (Laplace approximation). The reference level for the categorical variable is the first category. The values have been rounded off to two decimals. Length = item length (number of syllables).

[74] A model that included vocabulary comprehension instead of production gave the same overall results.

The analysis showed a significant effect of task, with lower accuracy for the QU-Ara task, and of item length, with lower accuracy rates with increased number of syllables. There was also a significant interaction between task and length. As can be seen in Figure 5.10, the accuracy rate decreased for both tasks with increased item length, but the effect was slightly stronger for the QU-Swe task; there was a larger difference between four-syllable items and five syllable items compared to the QU-Ara task. Accuracy rates for the QU-Swe task were nearly at ceiling for items with two syllables (97%), and then gradually decreased slightly at three (91%) and four syllables (80%). In the QU-Ara task, the pattern was similar, with a high accuracy rate (87%) at two syllables, staying constant at three syllables (88%), and decreasing slightly at four syllables (68%). Accuracy was slightly lower for the QU-Ara task compared to the QU-Swe task for all syllable lengths apart from five syllables, for which levels of accuracy were the same (35%). The control variable age was also significant; accuracy was better with increasing age.

5.4.2.5 Analysis of difference of accuracy for items in the two quasi-universal tasks

Model 3 in the previous section showed significant effects of task between QU-Swe and QU-Ara. This difference is unexpected, considering that the tasks are nearly identical in item setup, differing only in phoneme realisation, pronounced with a Swedish and Syrian Arabic accent respectively. Possible explanations for this discrepancy will be addressed in this section. First, as noted in Table 5.3, the difference in performance between the two tasks was particularly prominent among the 4-year-olds (2.2 points vs. 0.4–0.9 points in the other age groups). The possible interaction between age and task was explored in a logit mixed effects model, with age being operationalised as a categorical variable (4-year-olds vs. the three older age groups). There was no significant interaction effect between age group and task performance (see Figure A5.5 and Table A5.2 in the Appendix). Next, the difference in performance was explored separately for each item. Figure 5.11 shows accuracy rates for all items on the QU-Swe and the QU-Ara task.[75]

[75] Figures showing overall accuracy for performance on individual items on the LS-Swe task and the NWRT-Leb are available in the Appendix (Figure A5.6– A5.7).

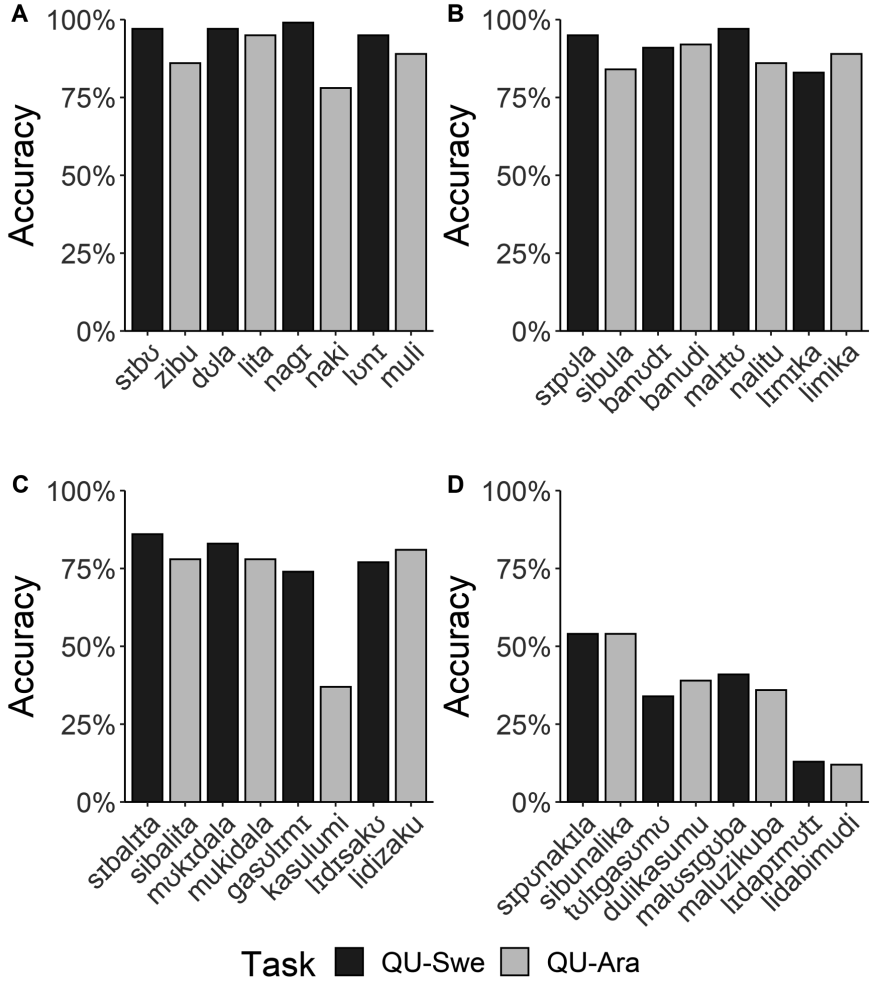

Figure 5.11. Overall accuracy (% correct answers) for items in QU-Swe (Swedish Quasi-universal) and QU-Ara (Arabic Quasi-universal), by number of syllables: two syllables (A), three syllables (B), four syllables (C), and five syllables (D).

An analysis of the differences in performance was conducted between corresponding items in the two language versions. Apart from items two (/dʊla/; /lita/) and four (/lʊnɪ/; /muli/), the items were nearly identical across tasks, apart from minor differences in voicing or place of articulation in one phoneme. In order to investigate whether there was a significant difference in accuracy for near-identical items between tasks, paired samples t-tests were conducted for all item pairs except number two and four. As shown in Table 5.8, significant differences in performance were found for item pairs one, three, five, seven and eleven. In all five cases, performance was significantly better on the Swedish item than the corresponding Arabic item.

Table 5.8. Difference in accuracy for item pairs in the Swedish quasi-universal task (QU-Swe) and the Arabic quasi-universal task (QU-Ara).

	QU-Swe	QU-Ara	*t*	*d*	*p*
1	/sɪbʊ/	/zibu/	$t(97) = -2.95$	.41	.004**
2	/dʊla/	/lita/	---	---	---
3	/nagɪ/	/naki/	$t(97) = -4.86$	.70	<.001***
4	/lʊnɪ/	/muli/	---	---	---
5	/sɪpʊla/	/sibula/	$t(97) = -2.76$	.37	.007**
6	/banʊdɪ/	/banudi/	$t(97) = .26$	.07	.798
7	/malɪtʊ/	/nalitu/	$t(97) = -2.95$	.40	.004**
8	/lɪmɪka/	/limika/	$t(97) = 1.35$	.17	.181
9	/sɪbalɪta/	/sibalita/	$t(97) = -1.65$	.21	.103
10	/mʊkɪdala/	/mukidala/	$t(97) = -.93$	.13	.356
11	/gasʊlʊmɪ/	/kasulumi/	$t(97) = -6.05$	.82	<.001***
12	/lɪdɪsakʊ/	/lidizaku/	$t(97) = .75$	.10	.453
13	/sɪpʊnakɪla/	/sibunakila/	$t(97) = 0$	0	1
14	/tʊlɪgasʊmʊ/	/dulikasumu/	$t(97) = .90$	.11	.372
15	/malʊsɪgʊba/	/maluzikuba/	$t(97) = -.90$	.10	.372
16	/lɪdapɪmʊtɪ/	/lidabimudi/	$t(97) = -.23$	.03	.820

Note. **$p < .01$, ***$p < .001$.

5.4.2.6 Item-related characteristics, participant-related factors and performance: summary

This section contains a summary of the results for the analyses concerning item-related characteristics and participant-related factors. Accuracy was analysed for a number of item-related characteristics (number of syllables, presence or absence of clusters, type of item by task) and participant-related factors (age, Swedish vocabulary scores, and length of exposure to Swedish).

First, pairwise comparisons between the LS-Swe task and the QU-Swe task, and the QU-Ara task and the QU-Swe task revealed a task effect; performance on the QU-Swe task was significantly better than performance on both the LS-Swe task and the QU-Ara task. Second, performance was investigated for items with and without consonant clusters in the LS-Swe task. The analyses revealed a significant effect of presence of clusters, with lower accuracy for items with clusters. Third, age was a significant predictor for accuracy rates for all tasks.

Across all tasks, there was an effect of the number of syllables (item length), with lower accuracy for items with more syllables. However, there were interaction effects between syllables and tasks as well as syllables and the presence of clusters. The analyses revealed the following. In the Swedish tasks (LS-Swe and QU-Swe), there was an interaction between task and syllables. Accuracy levels diminished with similar distance between 2–4 syllables, but then decreased proportionally more so for the (otherwise easier) QU-Swe task at five syllables. The same pattern was evident when comparing

accuracy levels for the two QU tasks (QU-Swe vs. QU-Ara): accuracy for the QU-Swe task was slightly higher at every syllable length besides five syllables, where accuracy was the same. Analysing the difference in performance on the LS-Swe task for items containing clusters vs. items not containing clusters revealed an interaction between the presence of clusters and number of syllables. Accuracy rates for items without clusters stayed constant between 2–4 syllables, but then dropped steeply at five syllables. By contrast, there was a steady decrease in accuracy rates for each additional syllable for items that contained clusters.

Finally, for the Swedish tasks (LS-Swe and QU-Swe), analyses were conducted to see whether there was an effect of vocabulary size (Swedish vocabulary comprehension) or length of exposure to Swedish. The analyses revealed a positive effect of vocabulary, but no effect of length of exposure to Swedish. Furthermore, there was no interaction effect between vocabulary and item length (number of syllables), vocabulary and presence of clusters in the LS-Swe items, nor between vocabulary and task. That is, there was a *general* positive effect of Swedish vocabulary scores on performance on the Swedish tasks, but children did not perform significantly better on (1) (the language-like) LS-Swe items, (2) longer items, or (3) items with clusters if they had higher Swedish vocabulary scores.

5.5 Discussion

This chapter investigated Arabic-Swedish-speaking bilingual children's phonological working memory with four different NWR tasks: one language-specific Swedish task (LS-Swe; Radeborg et al., 2006), one Swedish (QU-Swe) and one Arabic (QU-Ara) version of Chiat's (2015) quasi-universal task, as well as the Non-word Repetition Test-Lebanese (Abou Melhem et al., 2011). The following research questions were investigated:

- How does performance on the four different NWR tasks develop from age 4 to 7?
- Are there differences in performance between tasks?
- What is the relationship between participant-related factors such as age, vocabulary and length of exposure, and item-related variables such as type of task, item length, and syllable complexity and performance on NWR items?

In this section, the findings in the current study will be discussed in relation to the literature regarding age and task effects on NWR (section 5.5.1), and the effect of participant-related (age, language exposure and vocabulary scores) and item-related factors (item length in syllables, item complexity as

operationalised by the presence or absence of clusters, as well as task (LS-Swe vs. QU-Swe; QU-Swe vs. QU-Ara; section 5.5.2 and section 5.5.3).

5.5.1 NWR performance: age and task effects

The first issue to be discussed is age development. In all tasks, there was an increase in scores with age, mirroring the results of several previous studies which included TD children in a wide age range (Chiat & Roy, 2007; Kalnak et al., 2014; Radeborg et al., 2006; Topbaş et al., 2014). The effect of age in months was smaller in the QU-Swe task than the other tasks. Age development was also investigated via age groups. Generally, there were significant differences between the youngest and the oldest age group, but fewer differences between the age groups in the middle. For the NWRT-Leb, QU-Ara and QU-Swe tasks, there was a leap between the 4-year-olds and the 5-year-olds, but for LS-Swe a steadier pattern of increase between all age groups was found. Furthermore, there were striking differences in the proportional accuracy between tasks, with very high overall accuracy for the NWRT-Leb task, and very low overall accuracy for the LS-Swe task. The performance on each task will be discussed in the following.

In the Swedish language-specific task, the children in the current study scored on average two points below the monolinguals in the study by Radeborg et al. (2006). The same pattern was seen when comparing results between age groups in the current study and Radeborg et al.'s study. That is, 4-year-olds, 5-year-olds and 6-year-olds in the current study scored on average two points below the corresponding age group in Radeborg et al.'s study, which featured 200 children in age groups four, five and six. Radeborg et al. (2006) provide scarce information about how their scoring was conducted. However, since the scoring in the current study was lenient in order not to disadvantage children who had non-native-like pronunciation, it is likely that the scoring procedure in the current study was less strict. The scoring in the Radeborg et al. study was done online, whilst in the current study all items were recorded and transcribed in full before conducting the scoring.[76] What is more, the non-word task was presented live in Radeborg et al.'s study, whereas in the current study the tasks were pre-recorded and presented via headphones. It is hard to tell whether the difference in scoring (online or offline) or presentation mode (live or pre-recorded) could affect overall accuracy (beneficial or non-beneficial). Nevertheless, compared to the monolingual children in the Radeborg et al. study, the Arabic-Swedish-speaking bilinguals in this study scored on average two points less. This indicates a possible bilingual disadvantage on this language-specific task,

[76] All sessions were audio recorded in the Radeborg et al. (2006) study, however, only a small sample was transcribed afterwards in order to check inter-rater reliability, as described in the unpublished MA thesis on which the publication is based (Barthelom & Åkesson, 1995).

mirroring the findings in Thordardottir and Brandeker (2013) who found that the bilinguals in their sample of Canadian children scored significantly below the monolinguals on an English NWR task, and Boerma et al. (2015) who identified the same pattern when comparing the performance of monolinguals and bilinguals on a language-specific Dutch NWR task. One possible explanation for the overall lower scores of the participants in this study compared to the monolinguals in Radeborg et al. could be related to the fact that the bilinguals in the current study probably have smaller vocabularies in Swedish compared to the monolinguals in the Radeborg et al. sample. As reported by Gathercole (1995) and Szewczyk et al. (2018, p. 29), larger vocabularies are associated with better performance on NWR tasks, especially so if the non-word items are very word-like and language specific. There are no measures of vocabulary in the Radeborg et al. paper, but since the children in that sample were monolinguals, they had more overall exposure to Swedish and it is safe to assume that their Swedish vocabulary skills were higher at group level than the Arabic-Swedish-speaking bilinguals in the current study. In order to investigate this further, a monolingual control group would have to be compared to the Arabic-Swedish-speaking children in the current study, using the same assessment materials. There were also differences in performance between age groups in the current study compared to Radeborg et al. While the only significant difference between age groups in the current study was between the youngest (4-year-olds: M = 10.50, SD = 3.71) and the two oldest age groups (6-year-olds: M = 14.00, SD = 3.91 and 7-year-olds: M = 15.04, SD = 3.43), there were significant differences between all age groups in Radeborg et al. (4-year-olds: M = 12.70, SD = 3.80 vs. 5-year-olds: M = 14.70, SD = 3.70 vs. 6-year-olds: M = 16.40, SD = 3.70). This is unexpected, since the mean differences between age groups (1.7–2.0 points) are similar across studies, as are the SDs (3.43–3.91). However, there were more children per age group in the Radeborg et al. study, so these differences could be due to lack of power in the current study.

Second, the performance on the two language versions of the quasi-universal task (QU-Ara and QU-Swe) will be discussed. As shown in the analyses conducted in section 5.4.2.4, there was a significant difference in accuracy between the QU-Swe and the QU-Ara task in the current study, with better performance on the QU-Swe task. This difference in performance was surprising, since the QU task was developed to be comparable across language versions. Thus, when a group of bilingual children are tested on two different language versions, it is expected that performance would be the same. However, further analyses (section 5.4.2.5) showed that the effect was rather small, and to some extent driven by item effects. It is also interesting to compare the performance of the children in the current study with that of other children that were tested with other language versions of the task. Boerma et al. (2015) used a Dutch version of the quasi-universal task with monolingual and bilingual children in the same age range as in the current study (age range

54–87 months, corresponding to 4–7 years). Performance was assessed for percent phonemes correct as well as percent items correct (which is also reported in the current study). Interestingly, accuracy rates for the Arabic-Swedish-speaking bilinguals were much higher on both QU tasks (QU-Ara mean: 69.5%, QU-Swe mean: 76.0%) compared to the TD monolinguals as well as the TD bilinguals in Boerma et al. (monolingual-TD QU-Dut mean: 36.6%, bilingual-TD QU-Dut mean: 23.9%). Furthermore, the accuracy levels even for the youngest children in this study were higher than the mean accuracy of the children in Boerma et al. There are many possible reasons for the differences in performance between the studies, for instance differences in assessment procedure, transcription and scoring. Nevertheless, comparisons of the findings in the current study to accuracy levels reported for other language versions indicate that the QU task – although constructed with the intention of being equivalent across languages – cannot be assumed to be directly comparable across language versions, as anticipated by Chiat (2015, p. 143).

Third, the children's performance on the NWRT-Leb will be discussed. As previously mentioned, overall accuracy on the NWR-Leb was high (83.7%), the highest proportional accuracy of all tasks in this study. Overall accuracy was smaller in this sample compared to a sample of Lebanese children (N=35, mean accuracy: 92%) on which the task was piloted, however the age range was considerably smaller in that sample (5;5–6;8) (Abou Melhem, 2017).[77] The effect of syllabic complexity and item length (number of syllables) was not analysed statistically for this task. However, one can speculate that the main reason for the overall better performance was that the task only consisted of items with 1–3 syllables, in contrast to the other tasks, which contained items of 2–5 syllables.[78]

Finally, the characteristics of the low-performing children will be discussed briefly. There were several children who received a low score in one task only. Of the children who received a low score in two or more tasks, some children were reported to have had hearing problems, being a late talker, or having heredity for DLD, characteristics which are associated with language difficulties. Importantly though, not all low-performing children met these criteria, which points to the importance of interpreting the results of poor NWR scores in combination with the performance on other language measures and reports of early language development, heredity and hearing problems. This issue will be further developed in the clinical chapter, where some low-performing children are discussed with respect to poor NWR and vocabulary

[77] As a comparison, the mean accuracy for the five- and 6-year-olds in the present study (a larger sample, N=53, with a wider age range, 5;0–6;11) was 86%.
[78] Accuracy rates for individual items on the NWRT-Leb are available in the Appendix (Figure A5.7).

performance and the background information provided by the parents in the questionnaire (section 6.3.4.2).

5.5.2 Language experience and performance on language-like vs. non-language-like tasks: SES, language exposure and vocabulary

In the current study, the possible effect of previous language experience on performance on the LS-Swe task was investigated in two ways.[79] For this purpose, the QU-Swe task was used as a comparison, since this task contained items that were not constructed to adhere to Swedish phonotactics, but the items contained phonemes that exist in Swedish, and they were recorded by a native speaker of Swedish. First, exploratory analyses were conducted, investigating whether there were correlations between any of the tasks and: (1) SES, (2) percent current exposure to Swedish, (3) length of exposure to Swedish, and (4) Swedish vocabulary size. For both tasks, length of exposure to Swedish and Swedish vocabulary were positively correlated with performance, but neither SES nor percent daily exposure were. A multivariate analysis that took age into account in addition to the other variables (Model 1, Table 5.5) revealed that Swedish vocabulary was a significant predictor of accuracy on both tasks, but length of exposure to Swedish was not. Thus, the correlation between length of exposure to Swedish and NWR scores likely reflects an age effect, since at group level older children generally had longer exposure to Swedish and had higher NWR scores. Contrary to the expected pattern, Swedish vocabulary size was not a stronger predictor of accuracy for LS-Swe items (language-like) than QU-Swe items (non-language-like). However, none of the LS-Swe items contained syllables that were real morphemes in Swedish, which may explain why children were not more helped by having a larger Swedish vocabulary when repeating the LS-Swe items than the QU-Swe items. As demonstrated by Dollaghan, Biber and Campbell (1993), repetition accuracy increases if non-word items contain real morphemes. Furthermore, phonotactic probability and wordlikeness are two other factors known to influence NWR performance (S. E. Gathercole, 1995; Jones et al., 2010; Munson, Edwards, et al., 2005; Rispens et al., 2015; Szewczyk et al., 2018; Topbaş et al., 2014). In the current study, neither of these measures were investigated for the items in the LS-Swe task. Thus, it may be the case that such factors could have an impact on repetition accuracy for individual items in the LS-Swe task. Future studies including NWR items carefully controlled for these factors could answer whether children with

[79] The LS-Swe task was the only task in which the items were constructed to adhere to the phonological rules of Swedish. Since previous research has shown that performance on NWR tasks can be affected by previous language experience (in particular, vocabulary size) if the items are word-like, this was only analysed for the LS-Swe task.

smaller vocabularies (e.g. bilingual children) are disadvantaged on language-specific Swedish tasks. In the current study, overall accuracy was higher for the non-language-like and phonologically less complex QU-Swe task compared to the LS-Swe task. Interestingly, for monolingual Swedish-speaking adults who were tested on the LS-Swe and the QU-Swe tasks, the pattern was reversed; overall accuracy was slightly higher for the LS-Swe task than the QU-Swe task (Öberg & Lindgren, 2019). This indicates that very extensive exposure to Swedish and/or large vocabulary may facilitate accurate repetition of the language-like items in LS-Swe.

5.5.3 Item length, syllabic complexity and NWR performance

Repetition accuracy was also investigated in terms of item length (number of syllables) for both QU tasks and the LS-Swe task, and syllabic complexity (presence or absence of consonant clusters) for the LS-Swe task. For all tasks, accuracy generally decreased as a function of number of syllables, reflecting findings from multiple previous studies (Boerma et al., 2015; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Radeborg et al., 2006; Thordardottir & Brandeker, 2013; Topbaş et al., 2014).

The results of Model 2 (see table 5.6), which investigated the effect of clusters in the LS-Swe task, mirrored previous findings that items containing clusters were more difficult to repeat than those without clusters (Abed Ibrahim & Hamann, 2017; dos Santos & Ferré, 2018). Interestingly though, the patterns of performance for syllables containing clusters and those not containing clusters were not the same at all syllable lengths. While the children performed slightly better on items with clusters at two syllables, performance was clearly poorer for items with clusters at three and four syllables, but less so at five syllables. This indicates that consonant clusters may contribute more to item difficulty than do number of syllables, at least at some syllable lengths. In the current study, only presence vs. absence of clusters was tested (and not number of clusters), since there were only two items in the LS-Swe task containing two clusters, and they were both trisyllabic. Another aspect of syllable complexity is presence vs. absence of coda, which was not analysed in the current study. In order to exhaustively investigate the effect of syllable complexity on repetition accuracy, future studies should make use of a task that is controlled for the number of items containing codas as well as consonant clusters (both in onset and coda position), at every syllable length.

Although not tested statistically, an interesting pattern emerged with respect to non-word length: At five syllables, the effect of other item-related factors such as item type (LS-Swe vs. QU-Swe, QU-Swe vs. QU-Ara) and presence or absence of clusters (in the LS-Swe task) diminished. This pattern could tentatively be interpreted as a threshold, or a ‘five-syllable-effect’, where the number of syllables contributes stronger to item difficulty than any

other factor. It is not clear whether this effect holds across the whole age range. In order to investigate this further, it would have to be analysed separately for the different age groups.

5.5.4 Concluding remarks

In the current study, phonological working memory was assessed in Arabic-Swedish-speaking 4–7-year-olds, using four NWR tasks.

NWR tasks were included in the present study since they are a proposed clinical marker of DLD, at the same time as being minimally affected by previous language experience compared to other measures targeting language knowledge, for instance vocabulary. Four different tasks differing with respect to item length, phonological complexity and language-likeness were employed in order to investigate whether certain tasks or types of items that are more suitable for a certain population (in this case: Arabic-Swedish-speaking 4–7-year-olds).

The analyses showed that length of exposure to Swedish was not a significant predictor of performance on the Swe-LS task (nor the QU-Swe task, analyses that were made for comparison), but contrary to the expected pattern, Swedish vocabulary was a predictor of performance in *both* the LS-Swe and the QU-Swe tasks, and the effect of vocabulary was not stronger for LS-Swe items compared to QU-Swe items. That is, children with a lower vocabulary score were equally disadvantaged on the LS-Swe task and the QU-Swe task compared to children with higher vocabulary scores. Thus, quasi-universal tasks are not necessarily less biased for bilingual populations compared to language specific tasks, and it remains an open question which task or types of items that are best suited for identifying DLD while having minimal influence of bilingualism. In Chapter 6, the diagnostic usefulness is explored for the four NWR tasks through comparing the performance of children with a DLD diagnosis to the children in the cross sectional study.

6 The clinical study

Bilingualism poses a challenge for Swedish health care professionals when assessing language and communication skills in children with suspected Developmental Language Disorder (DLD). Misconceptions among child healthcare nurses that bilingualism causes language delay lead to simplified screening processes and delayed referrals for language assessment in bilinguals (Nayeb et al., 2015; Salameh, Nettelbladt, Håkansson, et al., 2002). Moreover, bilingualism has been reported by practising SLPs as a confounding factor making it difficult to make clinical judgements about the presence and severity of DLD (SOU 2016:46, pp. 222–223). Consequently, bilingual children are at risk of being both over- and underdiagnosed with language disorders (Grimm & Schulz, 2014). This can largely be attributed to a lack of sufficient assessment materials, reference data and overall knowledge about developmental trajectories concerning bilingual language development (in this case, for Arabic-Swedish-speaking children). It is important that DLD is identified early, so that language intervention can be offered from an early age. At the same time, wrongfully labelling a child with DLD when poor language skills are in fact attributed to insufficient language input can be stigmatising.

The aim of the present chapter is to better understand how DLD manifests in bilingual children, with a particular focus on vocabulary and NWR skills in Arabic-Swedish-speaking children. This is done by comparing the performance of Arabic-Swedish-speaking bilinguals who received a DLD diagnosis to their typically developing peers in the cross-sectional study. The second aim of this chapter is to investigate the early language development, language skills, communicative behaviour and social characteristics of the children with a DLD diagnosis as reported by parents, SLPs and teachers, and to explore how this information can be used in order to better understand how DLD manifests in bilinguals.

The following general research questions are asked:

- What are the reported early language development, current language skills, communicative behaviours and social characteristics of Arabic-Swedish-speaking children with a DLD diagnosis, as described by parents, SLPs and teachers?

- How does DLD manifest in Arabic-Swedish-speaking bilingual children aged 5–7, with respect to vocabulary skills and non-word repetition performance?

The setup of the chapter is as follows. Section 6.1 provides an overview of DLD (focusing particularly on DLD in the context of bilingualism), and vocabulary and NWR performance in bilinguals with DLD. The analyses and results of the clinical study are reported in section 6.2. The chapter closes with a discussion of the findings in light of previous studies in section 6.3.

6.1 Developmental Language Disorder in the context of bilingualism

6.1.1 DLD: An overview

DLD refers to a developmental disorder in which an individual has pronounced delayed or deviant language development, which cannot better be explained by hearing impairment, intellectual disability (ID), medical syndromes or neurological problems (Bishop, 1997, pp. 21–23; Bishop et al., 2017; Leonard, 2014, p. 3). Common early signs, typically manifesting in young childhood, include a late onset of first words, deficient expressive morphosyntactic skills and poor language comprehension. Estimates of prevalence of DLD differ depending on which definition is used. Two major epidemiological studies have determined the prevalence to be 7.4% in a cohort population of 5–6-year old children (N=1 502) in Illinois, USA (Tomblin et al., 1997), and 7.58% in a cohort population of 4–5-year-old children (N=529) in Surrey, UK (Norbury et al., 2016), thereby estimating the prevalence to be the same for the whole population. DLD may affect receptive and expressive language skills in isolation or in combination, and usually more than one linguistic domain (vocabulary, morphosyntax or discourse) is affected (Tomblin et al., 1997). The severity of impairment differs between individuals. Furthermore, the manifestation of DLD also changes with age. During the school years, spoken language may no longer be the primary difficulty, but children with DLD often face difficulties with literacy achievement (Catts, Fey, Tomblin, & Zhang, 2002; Snowling, Duff, Nash, & Hulme, 2016), and many also struggle to acquire academic language, e.g. expository discourse with increased morphosyntactic complexity (Nippold, Mansfield, Billow, & Tomblin, 2008), specialised vocabulary, etc.

Historically, many different terms have been used to describe children who struggle to acquire language in the expected way. The oldest terminology was based on concepts used in aphasiology and included terms such as *developmental aphasia* and *developmental dysphasia* (Leonard, 2014, p. 9). Until recently, *Specific Language Impairment* (SLI) has been the most

widespread and widely used term (Bishop, 2014; Leonard, 2014, pp. 10–11), although terms such as *Primary Language Impairment* (PLI) have also been used (Kohnert, 2010). The prefix 'specific' in SLI has been subject to criticism since it implies that the child has no difficulties other than those affecting their language abilities. Originally, the 'SLI' label was used to identify a narrow group of children with normal to high non-verbal IQ (NVIQ) who had a large discrepancy between non-verbal and language skills, and no other conditions such as attention deficit hyperactivity disorder (ADHD), developmental dyslexia or hearing loss. The definition was originally intended to be used mainly in research settings (Stark & Tallal, 1981).

However, since language disorders often do co-occur with neurodevelopmental disorders, in particular developmental dyslexia and ADHD, and the new term *Developmental Language Disorder* (DLD) was proposed in order to move away from the focus on 'specific' language problems (Bishop et al., 2017). In the previously mentioned epidemiological study by Norbury et al. (2016), children with low-normal NVIQ (i.e. 70–85) were not excluded from the sample. No significant differences were found in the language profiles and severity of language disorder (LD), emotional, social and behavioural problems or academic performance in children with average NVIQ compared to children with low-average IQ. Furthermore, children with LD and an additional medical diagnosis and/or intellectual disability (ID) often had more severe problems in many of these areas compared to the children with no medical diagnosis and average or low-average NVIQ. Norbury et al. conclude that the absence of ID and/or a medical diagnosis should not be a prerequisite for diagnosis or intervention regarding LD. Additionally, they argue that a discrepancy between verbal and non-verbal IQ should not be used as an exclusionary criterion for (D)LD in clinical decision making. In this thesis, the term 'Developmental Language Disorder' (DLD) will be used, and is considered to be synonymous with other used terms such as '(Specific) Language Impairment' ((S)LI) and 'Primary Language Impairment' (PLI).

The direct cause of DLD is not known, but there is much evidence that suggests that heredity plays a part. Several studies have shown that it is more common for monozygotic twins than dizygotic twins to both have a language disorder (Bishop et al., 1995; DeThorne et al., 2006; Tomblin & Buckwalter, 1998). Furthermore, there is also evidence that children who are diagnosed with DLD are more likely to have a grandparent, parent or sibling with language, literacy or social communication problems (Kalnak et al., 2012). Further evidence for heritability lies in the fact that several genes have been linked to language disorders (Newbury & Monaco, 2010). However, no gene has been identified as solely associated with DLD. Rather, several genes are associated with multiple developmental disorders (DLD, dyslexia, ADHD, autism), indicating that the genetic basis for DLD is complex (Bishop, 2009a).

6.1.1.1 Assessment and diagnosis

When diagnosing DLD, a classification manual is usually utilised. Currently in Sweden, the World Health Organisation's International Classification of Diseases-10 (ICD-10) is used in clinical practice (World Health Organization, 1992).[80] According to the ICD-10, a language disorder can be diagnosed when a child's language ability is significantly below the appropriate level of that of his or her mental age. Although the ICD-10 does not provide any precise demarcations between disordered and non-disordered language abilities, it is suggested that four main criteria should be considered during assessment: severity (in statistical terms), course (development over time), pattern (delay or deviance), and associated problems (co-occurring conditions or psychosocial problems) (World Health Organization, 1992, pp. 234–236). The WHO definition does not provide exact statistical cut-offs to determine the presence or severity of DLD, nor does it give any suggestion as to which language abilities should be assessed, but it proposes that a case falling below 2 SD below the age mean should be considered as severe. However, it is unclear whether performance on one test or language domain should be considered in isolation, or on a test battery targeting multiple domains.

When conducting the previously mentioned epidemiological study of 5–6-year old children in Illinois, USA, Tomblin, Records and Zhang (1996) constructed a system for identifying DLD (or in their words, SLI) as well as diagnosing sub-categories of the condition. Notably, exclusionary criteria for participating in the study were: bilingualism, blindness, hearing loss, 'mental retardation' (i.e. NVIQ < 70), autism, neurological problems, as well as having a NVIQ below 85. The authors used a language test battery based on the Test of Language Development-2: Primary (TOLD-2:P) as well as a narrative telling task. The diagnostic system (labelled the 'EpiSLI' system) encompassed five composite scores in three language domains (vocabulary, grammar and narration), as well as two modalities (production and comprehension). All children who received a composite score that fell –1.25 standard deviations below the mean on at least two of the five composite scores were considered to meet the criteria for SLI. Norbury et al. (2016) emphasised that existing criteria and cut-offs are completely arbitrary in the sense that it has not been systematically investigated in the general population to which extent certain diagnostic criteria or cut-offs correspond to functional communicative abilities. Norbury et al. used a similar setup as Tomblin et al. (1996), with a language assessment battery consisting of five composite indexes: production and comprehension of vocabulary, production and comprehension of grammar, retelling and comprehension of narratives, as

[80] Another commonly used classification manual is the Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 1995). According to a survey reported by Thordardottir (2015), the ICD-10 and the DSM-4 (fourth edition) are used in a number of European countries (either on their own or in combination) when diagnosing DLD.

well as expressive and receptive composites of vocabulary, grammar and narrative abilities respectively. A child was considered to fulfil the criteria of DLD if they received scores of –1.5 SD or below on at least two of the five composite indexes, a cut-off that closely matched functional impairment in school attainment.

To summarise, epidemiological studies have given insights on which the relevant cut-off scores may be for identifying children with DLD. Furthermore, the ICD-10 stresses that the process of assessing and diagnosing DLD should take into consideration background information in addition to language test scores, as well as the clinician's observations. In their clinical guidelines concerning assessment of language disorders, the American Speech-Language-Hearing Association (ASHA) also stress this, and argue that an assessment should include a case history, where information is gathered with respect to the medical and educational history, as well as the socioeconomic, cultural and linguistic background of the individual. If possible, information should be gathered from teachers and other related service providers in addition to parents (American Speech-Language-Hearing Association, 2004, p. 27).

With this information in mind, which are the most essential questions to ask parents that can guide the clinician when making diagnostic considerations? Several studies indicate that the *early language development* and *family history (heredity)* can be informative about the possible presence of DLD. Parents are an important source of information, and can provide details about onset of the first words and multi-word utterances. They can also tell the clinician whether there are any family members or close relatives who have (had) problems with language or literacy development. Furthermore, *parental concern* about the child's language development has also been shown to be related to the prevalence of DLD. All three factors have been associated with DLD (Tuller, 2015), and will be described in more detail in the following. Additionally, research on teacher reports will be reviewed.

One early sign of impaired language abilities is the late emergence of first words and first multi-word utterances. Typically, onset of the first word is considered to be late if it comes after 18 months, and toddlers who do not yet produce multi-word utterances at 24 months are commonly regarded as late talkers. Trauner, Wulfek, Tallal and Hesselink (2000) asked parents to report the age of onset of the first word and multi-word utterance (in months) in a group of monolingual American children with DLD (N=72) and without DLD (N=82).[81] They found that at group level, onset of the first word and multi-word utterance differed significantly between the DLD group ($M_{\text{first word}}$=22.7, SD = 11; $M_{\text{first multi-word utterance}}$=36.5, SD = 13) and the control group ($M_{\text{first word}}$=10.3, SD = 3.1; $M_{\text{first multi-word utterance}}$=17.1, SD = 5.2). Several studies using parental reports have shown that at 24 months, 10–20% of the participating

[81] Trauner et al. (2000) use the term (Developmental) Language Impairment, (D)LI.

children did not yet express multi-word utterances (Rescorla & Alley, 2001; Zubrick, Taylor, Rice, & Slegers, 2007). Although late talkers are at greater risk of developing persistent language disorders, far from all children do so (Ellis Weismer, 2007). Dale, Price, Bishop and Plomin (2003) conducted a study with the MacArthur-Bates CDI checklist with English and Welsh twin children, investigating long-term language outcomes in children with a late language development vs. children who had normal onset of language development. Parental assessments of 8 386 children's development in various linguistic and cognitive domains were gathered at age 2, 3 and 4. The children who had poor vocabulary scores (below the 10th percentile) at age 2 were assigned to the 'early language delay' group (ELD), and the rest of the children were assigned to the 'typical language' group (TL). Dale et al. found that the children with an early language delay were overrepresented in the group of children who were judged to have persistent language difficulties at age 3 (ELD: 44.1% vs. TD 7.2%) and age 4 (ELD: 40.2% vs. TD 8.56%), but the majority did not develop persistent language difficulties. Rice, Taylor and Zubrick (2008) found that children with a history of late language emergence (less than 70 words in their expressive vocabulary or no multi-word utterances at 24 months) were overrepresented in a group of children who performed low on language tests at age 7. Twenty percent of the late talkers performed below 1 SD on a test of general language ability, vs. 11% of the children in the control group. In sum, early language delay is considered to be a risk factor for developing persistent language difficulties, but not all children with DLD have a history of language delay.

The hereditary component of DLD has been attested in twin studies (Bishop et al., 1995), so-called family aggregation studies (Choudhury & Benasich, 2003; Spitz, Tallal, Flax, & Benasich, 1997; Stromswold, 1998), and studies investigating molecular genetics in language disorders (Anthoni et al., 2012). In an interview study conducted with parents of Swedish monolinguals with DLD (N=61) and without (N=100), Kalnak et al. (2012) explored the prevalence of language-related diagnoses and problems (LRDP) in three generations (grandparents, parents and siblings). Reading and writing difficulties were the most common symptoms reported in all generations of relatives of children with DLD. Among the parents, presence of language and literacy problems were five times as high in the DLD group compared to the control group. Similarly, language problems were five times as frequent, and literacy problems three times as frequent in siblings of children with DLD compared to the controls. Thus, a family history of language or literacy difficulties is considered a risk factor for DLD.

Apart from providing information about early language development and family history, parents should also be asked to evaluate their child's current language skills. Restrepo (1998) interviewed parents of bilingual English-Spanish children aged 5–7 with (N=31) and without (N=31) a DLD diagnosis, and asked them questions concerning their child's current language skills. She

found that parental reports of the children's language skills could significantly discriminate between the two groups. In a Swedish context, Salameh, Nettelbladt and Gullberg (2002) investigated parental concern regarding language development in 131 bilingual children who had been referred during the course of 12 months to an SLP clinic in Malmö, Sweden. They found that a majority of the parents of children who received a DLD diagnosis had expressed concern about their child's language development. Furthermore, the proportion of parents who expressed concern increased with the level of severity of the disorder (mild DLD: 60%, moderate DLD: 72.2%, severe DLD: 82.6%).

In addition to parental reports, useful information about the child's language and communication skills can also be obtained from teachers and preschool staff. Teachers see the child every day, know about their learning outcomes, and observe them in interaction with peers and adults. Teacher reports may also be less biased than parents', since teachers can compare the child to other age peers, and know how they fare compared to the group. Purse and Gardner (2013) showed that teacher evaluations of children's comprehension skills correlated with standardised measures of comprehension of instructions and paragraphs in a group of monolingual English 6–8-year-olds (N=30). Botting, Conti-Ramsden and Crutchley (1997) demonstrated that teacher descriptions of children's language difficulties correlated with the child's performance on standardised measures of phonology and morphosyntax in a study of 242 second-graders (mean age: 7 years) with language impairment attending special language school units in England. However, semantic-pragmatic difficulties described by the teachers did not become evident the standardised tests used in the study. Thus, teacher evaluations provide ecologically valid reports of children's functional language skills, and may also reveal difficulties that are not always straightforwardly captured by standardised language tests. In the present study, teachers of the children with a DLD diagnosis were interviewed about the child's language and social communication skills.

6.1.1.2 Common characteristics

As previously mentioned, DLD may affect different linguistic domains: phonology, morphology, syntax, vocabulary, discourse and pragmatics. DLD also affects language processing skills, such as non-word and sentence repetition. The linguistic manifestations of DLD will vary in different languages due to typological differences, in particular when it comes to morphology and syntax (see Leonard (2014, Chapter 4) for an overview). Therefore, the linguistic symptoms of DLD must be established for each language separately. This section contains a brief overview of some common linguistic manifestations of developmental language disorder that have been reported crosslinguistically in the literature.

6.1.1.2.1 Language processing/working memory

Non-word repetition and sentence repetition are both tasks that do not directly target explicit language knowledge but rather implicit skills: the ability to process linguistic stimuli. Non-word repetition entails repeating a series of phonological nonsense forms as accurately as possible. Sentence repetition involves repeating a series of clauses, varying in word length and morphosyntactic setup and complexity. Several studies have found that children with DLD have difficulties with non-word repetition (see section 6.1.5 for a detailed overview) and sentence repetition (Conti-Ramsden, Botting, & Faragher, 2001; Thordardottir & Brandeker, 2013) compared to their age peers.

6.1.1.2.2 Phonology

Although not a core feature of DLD, many children exhibit delayed or deviant phonological development. During childhood, these difficulties typically appear as phonological processes in their speech (i.e. systematic substitutions, metatheses or deletions of speech sounds or syllables) which can make them difficult to understand to others (Sices, Taylor, Freebairn, Hansen, & Lewis, 2007). As children grow older, phonological deficits may manifest as poor phonological awareness, which in turn may hamper early literacy development (Vandewalle, Boets, Ghesquière, & Zink, 2012).

6.1.1.2.3 Vocabulary

Several symptoms of DLD have been attested in the vocabulary domain. A late emergence of first words has been described crosslinguistically for children with DLD (see section 6.1.2). Children with DLD also tend to have smaller vocabularies, as determined by standardised vocabulary tests (see section 6.1.4 for an overview), and they often show a slower rate of vocabulary growth compared to their typically developing peers (Rice & Hoffman, 2015). A possible reason for this is the fact that children with DLD have shown to be slower at acquiring new words, needing a higher amount of repetitions in order to learn them (Nash & Donaldson, 2005). Further evidence indicates that children with DLD often have weak lexical processing skills and experience lexical retrieval problems, which can manifest in several ways. For instance, they may have longer reaction times and make more errors compared to typically developing peers when performing tasks that tap into lexical retrieval such as rapid naming (Coady, 2013; Sheng & McGregor, 2010). At the discourse level, children with lexical deficits may produce narratives with reduced lexical diversity or have more disfluencies (e.g. ‘um’, ‘eh’, ‘what is it called’) in connected speech compared to their age-matched peers (Thordardottir & Ellis Weismer, 2002).

6.1.1.2.4 Morphology and syntax

The morphological and syntactic development of children with DLD has been researched extensively. The way in which DLD manifests morpho-syntactically varies substantially between different languages, due to the vast typological disparity of the languages in the world. Although the specific constructions that are affected vary between languages, some common patterns can be identified crosslinguistically.

Concerning comprehension, children with DLD often have a weaker understanding of syntactically complex structures compared to their typically developing peers. For instance, children with DLD have been shown to have a weaker comprehension of object relative clauses in Hebrew (Friedmann & Novogrodsky, 2004), and clauses with a non-canonical word-order (e.g. passives) in English (Marshall, Marinis, & van der Lely, 2007).

With respect to expressive language skills, individuals with DLD are prone to make morphological errors, e.g. incorrect use of tense inflections in English and Swedish (Conti-Ramsden et al., 2001; Hansson, Nettelbladt, & Leonard, 2003) or case marking in Hungarian and Hebrew (Lukács, Kas, & Leonard, 2013; Rom & Leonard, 1990). Regarding syntax, children with DLD are more prone to make word order errors (Hansson & Nettelbladt, 1990; Lely & Battell, 2003) and omit copulas and function words (Hansson & Leonard, 2006; Leonard, Eyer, Bedore, & Grela, 1997). Often they form less syntactically complex clauses compared to their TD peers (Topbaş, Güven, Aydin Uysal, & Kazanoglu, 2017), with fewer arguments and fewer types of argument structures (Thordardottir & Ellis Weismer, 2002). This deficit often persists into adolescence and adulthood. Adolescents with (a history of) DLD may have difficulties learning and making use of a more diverse syntactic repertoire, i.e. such that is needed for acquiring academic style writing skills (Nippold, Mansfield, Billow, & Tomblin, 2009).

6.1.1.2.5 Discourse and Pragmatics

Another area of difficulty for individuals with DLD is discourse and pragmatics. Concerning discourse, narrative abilities have been identified as a problem area. Producing a narrative is a complex cognitive task, since it entails bringing together linguistic skills from different domains (vocabulary, morphology, syntax) at the same time as having to produce a coherent storyline, and taking the perspective of the listener into account (Botting, 2002). Often children with DLD will produce shorter narratives than their peers, with reduced structural complexity (Botting, 2002; Fichman, Altman, Voloskovich, Armon-Lotem, & Walters, 2017; Schneider, Hayward, & Dubé, 2006), less varied vocabulary (Paul & Hernandez, 1996), and have difficulties using suitable referring expressions (Miranda, McCabe, & Bliss, 1998; Norbury & Bishop, 2003; Schneider & Hayward, 2010).

Some, but not all, children with DLD may have difficulties at the general pragmatic level. This is thought to be linked to difficulties integrating word and clause level comprehension with the overall situation, and a poor ability to understand and interpret the thoughts, needs and intentions of others (Bishop, 2000; Ryder & Leinonen, 2014). Common signs include difficulties with turn taking during conversations, topic maintenance (topic shift or topic drift), making inferences and interpreting figurative speech (Botting & Conti-Ramsden, 1999). Pragmatic difficulties in children with DLD can be difficult to capture with norm-referenced language tests, but using questionnaires with parents and teachers may uncover these communicative difficulties (Norbury, Nash, Baird, & Bishop, 2004). Social relations may be affected as a consequence of pragmatic language difficulties, e.g. peer conflicts (Ketelaars, Cuperus, Jansonius, & Verhoeven, 2010).

6.1.2 Identifying DLD in bilingual children

When a bilingual child does not develop their language skills in the expected way, parents, clinicians and educators are faced with the difficulty of trying to figure out – disentangle – whether this delay is due to not receiving enough high-quality input or whether there is a language disorder. Goral and Conner (2013) highlight two key issues that are often cause for confusion. First, there is an overlap in the language characteristics of typically developing bilinguals and the linguistic markers of DLD. Second, 'bilinguals' are a heterogeneous group of children, differing immensely amongst each other regarding which language combination they speak, at which time they became bilinguals (age of onset), the quantity and quality of input they have received in each language, and which socioeconomic background they have. Letts (2013) also points out this predicament, further underscoring that the heterogeneity among bilinguals leads to difficulties identifying a relevant peer group for comparison. Armon-Lotem (2012) argues that the effects of bilingualism can be untied from those of DLD by integrating knowledge about bilingualism and knowledge about language disorders, i.e. a bilingual child's language performance needs to be interpreted in light of information about their linguistic and social background.

In the previous section, some common linguistic characteristics associated with DLD were outlined. Morphosyntax is perhaps the linguistic domain that has been most extensively researched in monolinguals with DLD. As described earlier, numerous studies have found that monolingual children with DLD may have difficulties with for instance temporal marking, case marking, word order, copula verbs and other function words. Such 'errors' in expressive morphosyntax are sometimes referred to as *clinical markers* of DLD. However, evidence from several languages shows that there is extensive overlap in the morphosyntactic features of normal bilingual language acquisition and morphosyntactic clinical markers of DLD (Boerma, Wijnen,

et al., 2017; Paradis & Crago, 2000). In fact, bilinguals with typical language development may score significantly below monolingual age norms on standardised language tests not only in morphosyntax but in several language domains (Barragan et al., 2018). In Swedish, comparisons of morphosyntax between monolinguals with DLD and typically developing bilinguals have shown an overlap in the use of constructions with the verb in third position (XSV), instead of the target structure with the verb in second position (Håkansson, 2001; Håkansson & Nettelbladt, 1993). A recent study explored the performance of 224 bilingual and monolingual Swedish 7–8-year-olds on the 'core language skills' from the language test battery Clinical Evaluation of Language Fundamentals (CELF-4) (Andersson et al., 2019).[82] Andersson et al. found that as many as 80% of the bilingual children scored 1 SD below the mean, which is the cut-off recommended for further assessment (i.e. indicating possible DLD).

Kohnert (2010) argues that comparing bilinguals to monolingual norms is precarious for knowledge-based language measures as well as tasks relying on language processing, since even language processing tasks are to some degree influenced by earlier language experience. So how should we go about the task of assessing bilinguals in a valid way? It is often claimed that DLD should manifest in both languages (Kohnert, 2010; Salameh, Nettelbladt, & Gullberg, 2002). However, in reality, typological differences between the languages and differences in proficiency between the child's languages may lead to a language disorder not manifesting in the same way or to the same extent in both languages. Furthermore, it may be difficult to assess whether language skills are affected in both languages, particularly in languages where little is known about developmental trajectories and characteristics of typical and atypical language development. Thus, in order to solve the 'diagnostic dilemma' in bilingual DLD in a satisfactory way, it is advised to (i) identify and consider the signs of DLD that hold for both monolinguals and bilinguals alike, (ii) compare the child's language skills to those of a suitable group of comparison, and (iii) evaluate a child's language performance in light of information about language input.

Keeping in mind the overlap in linguistic features of typically developing bilinguals and monolinguals with DLD, there are some areas of language acquisition that may be informative when assessing bilinguals with potential DLD, for instance the timing of early language development. Paradis, Emmerzael and Sorenson Duncan (2010) interviewed Canadian parents of bilingual children with and without a DLD diagnosis. They found that the mean age for the debut of the first word was 12.96 months in bilinguals without DLD (N=139) and 21.96 months (N=29) in bilinguals with a DLD diagnosis, with a significant difference between the two groups. Hoff, Core,

[82] 'Core language skills' include the subtasks 'Concepts and following directions', 'Word structure', 'Recalling sentences' and 'Formulated sentences'.

Place, Rumiche, Señor and Parra (2012) used parental reports to investigate the onset of multi-word utterances in a group of bilingual Spanish-English children (N=47) compared to a monolingual English speaking control group (N=56). Parental reports on vocabulary and early syntactic development were collected at 1;10, 2;1 and 2;6 years of age (years;months). The researchers found that at age 1;10, significantly fewer of the bilinguals used multi-word utterances in the majority language English compared to the monolinguals. However, when counting the number of bilinguals who used multi-word utterances in at least one of their languages, there was no difference between them and the monolinguals concerning the proportion of children using combinatorial speech. Thus, bilingual children should not be expected to have an overall delay in the early language development, but it is important to keep in mind that the early milestones may not appear at the same age in both languages.

In the current study, the following considerations were made, taking into account knowledge about language disorders in the context of bilingualism:

- Both languages are assessed.
- The children in the DLD group are compared to a group of TD children with the same linguistic and social background (i.e. Arabic-Swedish-speaking children).
- In addition to asking parents about the child's early language development and family history, information is also gathered concerning bilingual experience, such as age of onset for both languages and exposure patterns.

6.1.3 DLD in Swedish bilingual children

In the Swedish context, DLD in bilinguals has been explored in two epidemiological studies that investigate age at referral, predictors of severe DLD and the proportion of children who were considered to have severe DLD in monolinguals and bilinguals. Research on the linguistic features of DLD in bilingual children has mostly concerned phonology, morphosyntax and vocabulary in the Arabic-Swedish-speaking population. These studies will be reviewed in the following.

Salameh, Nettelbladt, Håkansson and Gullberg (2002) compared the characteristics of monolingual (N=246) and bilingual (N=192) children who were referred for language assessment to an SLP clinic in southern Sweden during the course of 12 months. The monolingual and the bilingual group were compared with respect to source of referral, estimated severity of DLD, parental concern and therapy attendance. Although the proportion of bilinguals who were referred for assessment matched the proportion of bilinguals in the preschool population, the bilingual children were overrepresented in the group who were considered to have severe DLD. Notably, the bilingual children had a lower likelihood than monolinguals of

being referred before age five, but a higher likelihood of being referred after age five. Bilinguals were also less likely to be referred by the child healthcare centres compared to monolinguals, but more likely to be referred by preschools or other sources. Salameh, Nettelbladt and Gullberg (2002) investigated risk factors for severe DLD in monolinguals (N=446) and bilinguals (N=252), this time in children who had been referred over the course of 24 months to the same SLP clinic as in the previous study. In both groups, significant predictors of the severity of DLD were parental reports of comprehension problems, having a short attention span, and concern about language development.

Håkansson, Salameh and Nettelbladt (2003) investigated expressive morphosyntax in 20 Arabic-Swedish-speaking bilinguals with and without DLD (N=10 in each group). The children were 3;10–6;7 years at the time of testing, and they were matched pairwise across groups according to chronological age, gender, length of exposure to Swedish via preschool and Arabic variety. First, all children were tested in Swedish with standardised language tests probing expressive phonology and morphosyntax – Nya Lundamaterialet (the Lund test of phonology and grammar; Holmberg & Stenkvist, 1983), and comprehension – SIT (the Impressive language test for children; Hellqvist, 2011). The DLD group scored below the TD group on both measures, but there was only a significant difference between the groups on the comprehension task. Second, the children's expressive morphosyntax was assessed in both Arabic and Swedish with a task probing different morphosyntactic constructions. At group level, the children in the TD group scored higher in both languages. A longitudinal follow-up was conducted six and twelve months after the first session (Salameh et al., 2004), showing that the TD group scored higher than the DLD group on the probed morphosyntactic structures in both languages at all three sessions. Salameh, Nettelbladt and Norlin (2003) investigated the phonological development in the same 10 TD and 10 DLD children that participated the two previous studies. The children were assessed on phonology, using picture naming tasks eliciting 58 words in the Swedish phoneme test (Hellqvist, 1991) and 28 words in an Arabic phoneme test that had been constructed for the study. At group level, the TD children had higher accuracy in both Arabic and Swedish. An error analysis showed that the children with DLD had a wider range of phonological processes (e.g. segment substitutions, assimilations and cluster reductions) in both languages compared to the TD children.

In her doctoral thesis, Holmström (2015) investigated the vocabulary skills of 88 monolingual and bilingual Arabic-Swedish-speaking children with and without DLD between the ages of 5;11–9;3. Twenty-four children had a DLD diagnosis (15 bilingual, 9 monolingual) and 64 children had typical language development (29 bilingual, 35 monolingual). Vocabulary skills were assessed for comprehension (a task probing school-related words), production (a task developed for the study probing nouns) and lexical organisation (the Kent and

Rosanoff word association task). One study compared the vocabulary size of Arabic-Swedish-speaking bilinguals with DLD (N=15, M_{age}=7;0) and without DLD (N=15, M_{age}=7;4). The children were assessed in both Arabic and Swedish with tasks probing vocabulary comprehension and production. The groups did not differ significantly with respect to age, but there was a significant difference in NVIQ (the average score was higher in the TD group). The children with a DLD diagnosis performed significantly lower than the TD children on Arabic and Swedish comprehension and Swedish production, but not on Arabic production. In another study in her thesis, Holmström explored the lexical organisation over time in Arabic and Swedish in 20 children ($M_{age\ Time\ 1}$=7;1) with and without DLD. The children were matched for age and gender, with 10 children in each group. Lexical organisation was assessed twice (twelve months apart Time 1 and Time 2) using the Kent and Rosanoff word association task in Arabic and Swedish (Kent & Rosanoff, 1910). The task entails hearing a word, and making an association to the first word one comes to think of. Holmström classified the associations as sound-based, syntagmatic, paradigmatic or 'other', where paradigmatic associations were considered to be the most advanced. The TD children increased the number paradigmatic associations between T1 and T2 in both Arabic and Swedish, while the children in the DLD group produced the same proportion of paradigmatic associations at T1 and T2.

Recently, Öztekin (2019) investigated the vocabulary and narrative skills in both languages of six bilingual Turkish-Swedish-speaking children (aged 4;8–8;2) who had received a DLD diagnosis. Their vocabulary comprehension and production was assessed using the CLT (Haman et al., 2015), and comprehension and production of narrative macrostructure was assessed using the Multilingual Assessment Instrument for Narratives, MAIN (Gagarina et al., 2012). Each child's performance was compared to a reference group from a cross-sectional study, consisting of circa 25 children in each age group (4-, 5-, 6- and 7-year-olds) with typical language development (TD) speaking the same language combination (Turkish-Swedish). Thus, a child who was 4;8 was compared to 4-year-olds, a child who was 6;9 was compared to 6-year-olds and so on. Concerning vocabulary, Öztekin found that most children in the DLD sample had negative z-scores in both vocabulary comprehension and production in both languages, although two children had positive z-scores in one language (these two children had a higher degree of daily exposure to the language they scored better in). Additionally, most children had very low z-scores, below –1.25 in two or more CLT subtasks. For comprehension of narrative macrostructure, the majority of the children with DLD performed low compared to their age peers; z-scores were generally negative, with occasional exceptions. Most children in the DLD group had z-scores below –1.25 in at least one language. The children with DLD generally had negative z-scores also concerning production of narrative macrostructure, but the

difference in performance between the children with DLD and the TD group was less pronounced than for comprehension.

To conclude, comparisons between bilinguals and monolinguals in Sweden have shown that several common signs associated with language disorder in monolinguals (e.g. parental reports of comprehension problems, and concern regarding language development) are also predictors of DLD in bilinguals. Even so, bilinguals are more likely to be referred for language assessment at a later age compared to monolinguals. Although previous studies on the linguistic manifestations of DLD in Swedish-speaking bilinguals do investigate both languages, there is a lack of studies that compare the performance of children with DLD to a larger reference group (Öztekin's study of Turkish-Swedish-speaking children being a notable exception to this). In the current study, Arabic-Swedish-speaking children with a DLD diagnosis (N=11) are being compared on vocabulary and NWR performance against a larger reference group of children (N=99) speaking the same language combination.

6.1.4 Vocabulary skills in bilinguals with DLD and studies using Cross-Linguistic lexical tasks (CLTs) to investigate vocabulary in children with DLD

This section contains an overview of previous studies that have investigated vocabulary performance in monolingual and bilingual children with DLD. Additionally, studies that have used the CLTs to investigate vocabulary skills in monolingual and bilingual children with DLD will be reviewed.

As previously stated, a slow vocabulary development can be an early sign of delayed or deviant language development. However, while many children with DLD have a delayed onset of the first word, a slower vocabulary growth curve compared to age peers, and score lower on norm-referenced tests, not all of them perform poorly on vocabulary tasks. Spaulding, Hosmer and Schechtman (2013) investigated the diagnostic utility of the third and fourth editions of the PPVT (Dunn & Dunn, 2007, 1997) in monolingual English-speaking preschool children in the USA with DLD (M_{age}=51.77 months, N=40) and a TD control group (M_{age}=52.35 months, N=40). The groups differed significantly with respect to their general language abilities as determined by the Clinical Evaluation of Language Fundamentals, Preschool (CELF-P; Wiig, Secord, & Semel, 2004), but not regarding age, SES (maternal education) or NVIQ. Sensitivity was at 80% for both versions of the PPVT but specificity rates differed somewhat, being 75% in PPVT-III and 70% in

PPVT-IV.[83] Gray, Plante, Vance and Henrichsen (1999) investigated the diagnostic accuracy of four vocabulary tests in a group of monolingual English-speaking preschool children aged 4–5 with DLD (N=31) and a TD control group (N=31) in the USA. Two tests targeted vocabulary comprehension via picture selection and two targeted vocabulary production via picture naming. Although the DLD group performed significantly below the TD group on all measures, there was substantial overlap between the groups. Discriminant analyses were performed in order to investigate how well each test could correctly identify the children as impaired or non-impaired. Sensitivity and specificity rates were found to be moderate for all tasks, ranging between 68–74%, far too low in order to be utilised for clinical decision-making. As Gray et al. concluded, vocabulary tests are an important part of linguistic assessment of DLD, but they should not be used as a sole measure to determine clinical status.

Although there has been an increase during the later years in the number of studies concerning vocabulary skills in bilingual children with DLD, it is still rare to find studies where both languages are assessed. Thordardottir and Brandeker (2013) compared the performance of 14 Canadian bilinguals with DLD and 14 bilinguals with TD (with a mean age close to 5;0 years) on a receptive vocabulary test in their L2 French (Échelle de Vocabulaire en Images de Peabody, EVIP; Dunn, Theriault-Whalen, & Dunn, 1993). They found that while the EVIP had excellent sensitivity (100%), specificity was very poor (43%). In other words, all children who did have a DLD diagnosis were correctly classified as such at the utilised cut-off score, but more than half of the TD children were incorrectly classified as having DLD. Boerma, Leseman, Wijnen and Blom (2017) investigated Dutch vocabulary comprehension in L2 speakers of Dutch with DLD (M_{age}=71.3 months, N=32) and TD (M_{age}=72.4 months, N=32). They used a longitudinal design where the children were tested three times with 12 months between each session. The analyses revealed that the TD children scored significantly higher than the DLD children at group level, but there was considerable overlap between the ranges; at Time 2 the top score in the DLD group even exceeded that of the TD group.

Peña, Bedore and Kester (2016) point out that even in situations where vocabulary assessment materials and reference data are available for both languages of bilingual children, there are no empirically validated procedures or cut-offs that can offer a guide in clinical decision-making. Therefore, Peña et al. investigated whether cut-offs established for monolingual populations

[83] As mentioned in chapter 5 (section 5.1.5), *sensitivity* and *specificity* are frequently used concepts in clinical studies. They are used when evaluating the accuracy of a diagnostic test. Sensitivity refers to how accurately the test (e.g. a vocabulary or NWR test) identifies all individuals in a sample with a particular condition (i.e. DLD); while specificity refers to how accurately the test can identify all individuals in the same sample who do not have the condition (i.e. TD) (Loong, 2003).

could be informative when assessing bilingual children. Seventy-eight Spanish-English-speaking children (aged 4;0–6;11) who had balanced exposure to both languages (40–60% according to parent and teacher reports) participated in the study, 15 with DLD and 63 with TD. All children were assessed in Spanish and English on the semantics index of the Bilingual English Spanish Assessment (BESA; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014).[84] When taking both languages into account, the BESA had excellent accuracy, correctly classifying 96.2% of the sample, with 93.3% sensitivity and 96.8% specificity. However, there was a considerably large group of TD children who scored below the (monolingual) cut-off score in one language, and thus would have been misclassified if assessment had been conducted in one language only.

In the current study, Cross-Linguistic Lexical Tasks (CLTs) are used to assess vocabulary comprehension and production in the minority and the majority language of bilingual Arabic-Swedish-speaking children with a DLD diagnosis. The CLT was constructed specifically with bilingual assessment in mind, with the expectation that it would be able to distinguish between monolinguals vs. bilinguals, and TD vs. DLD (Haman et al., 2015). To date, a handful of studies have compared vocabulary comprehension and production scores on the CLT of children with DLD and typically developing children. Kapalková and Slančová (2017) compared the performance of 15 monolingual DLD children (M_{age} = 66.33 months, range = 51–80) to 15 monolingual age-matched TD children using the Slovak version of the CLT. The TD group scored significantly higher than the DLD group on all four subtests (production and comprehension of nouns and verbs). Khoury Aouad Saliby, dos Santos, Kouba Hreich and Messarra (2017) compared the performance of 32 TD children (age range 5;7–6;9) to 10 children with a DLD diagnosis (age range 5;9–7;10) on the Lebanese version of the CLT. Conceptual scoring was used in the expressive task (i.e. correct answers in Arabic, French or English were accepted). The TD group scored consistently higher than the DLD group on the production of nouns and verbs as well as on the comprehension of verbs (noun comprehension was close to ceiling for both groups). As previously mentioned, Öztekin (2019) compared CLT performance of six bilingual Turkish-Swedish-speaking children (age range: 4;8–8;2) with DLD to a cross-sectional TD reference group of circa 25 children in each age group (in total 102 children aged 4–7) speaking the same language combination. Most children in the DLD sample had negative z-scores in both comprehension and production in both languages (although two children had positive z-scores in one language, and most children had very low z-scores, below –1.25 in two or more subtasks).

[84] Note that the semantics index in the BESA consists of six subtests that do not probe vocabulary size, but rather vocabulary depth, e.g. verbal analogies, similarities and differences, describing the function of nouns, etc.

To summarise, vocabulary is an important part of language assessment when suspecting DLD. However, not all children with DLD will have low vocabulary scores. Thus, vocabulary tasks should not be used in isolation in order to determine clinical status. When assessing bilinguals, both languages should be evaluated, since this gives a more accurate image of a child's full lexical abilities. Finally, the CLTs have been shown to distinguish between TD and DLD groups in both monolingual and bilingual populations.

6.1.5 Clinical application of NWR tasks in monolingual and bilingual populations

The clinical usefulness of a NWR task is dependent on how well it identifies children with DLD. Studies investigating the use of NWR when screening for or diagnosing DLD have used different approaches. Many studies administer NWR tasks to two predefined groups of children: children with typical language development (TD) and children with a DLD diagnosis, and examine whether there is a significant difference in performance between the groups. Results from studies with this setup are not directly transferrable to a clinical setting, since there is not a binary distribution in the performance on NWR tasks between TD children and children with DLD. Rather, there is a spectrum of impairment vs. non-impairment, where many children perform in the 'grey area' of what should be considered a low score. From a clinical perspective, the primary concern is whether a particular NWR task is useful in identifying children with DLD and if so, which cut-off is best for avoiding both over- and underidentification. This section surveys research on the clinical application of NWR tasks. Most of these studies make use of predefined groups (typically developing, TD, and children with DLD), but there are also a couple of studies that investigate the performance of NWR in unselected populations (i.e. studies where the participants have not been recruited based on diagnostic status, TD vs. DLD, but rather reflect the spectrum of variation in the whole population). Some of the studies that compare the performance of a DLD group to that of a control group make use of sensitivity and specificity measures (see section 6.1.4 for a description of these concepts). First, an overview of studies investigating the performance on NWR tasks in monolinguals will be given, and after that follows an overview of studies concerning bilingual populations.

Several studies of predefined groups of monolinguals (TD vs. DLD) have found that DLD children perform significantly below TD children, and often with very little or no overlap in performance between groups. Dollaghan and Campbell (1998) found that English-speaking children (age 6;0–9;9) with DLD performed significantly lower than the children in the control group for non-words with three or four syllables but not on shorter item lengths. Furthermore, for the total score on the test (the Nonword Repetition Test,

NRT), the DLD group performed significantly poorer than the TD group, and there was no overlap between the groups even at 99% confidence intervals. Chiat and Roy (2007) found that a group of young children (aged 2;0–4;0) who were at risk for developing DLD performed significantly poorer than a control group on a test encompassing both repetition of non-words and real words (the PRT), although there was overlap in performance between groups. Looking at the performance on NWR alone, the mean score for the clinical group was within 1.5 to 2 SD below the mean for the TD group. Saiegh-Haddad and Ghawi-Dakwar (2017) compared the performance on an Arabic NWR test for a group of Palestinian-speaking pre-schoolers (M_{age} = 5;5) and first-graders (M_{age} = 6;11) with typical language development and DLD, and found a significantly poorer performance by the DLD group compared to their TD peers. Topbaş, Kaçar-Kütükçü, and Kopkalli-Yavuz (2014) investigated the clinical accuracy of a Turkish NWR task in 4–8-year-old children. They found sensitivity and specificity levels within the fair range (sensitivity 89%, specificity 86.5%). Furthermore, diagnostic accuracy was 90% at a cut-off of –2 SD as determined by a ROC curve analysis.[85] Kalnak, Peyrard-Janvid, Forssberg and Sahlén (2014) compared the performance on a NWR task in Swedish-speaking school-aged children (age 8–12 years) with and without DLD. They found that the diagnostic accuracy of the NWR test was high at a cut-off of –2 SD from the mean, with a sensitivity level of 90.2% and a specificity level of 97.7%. Additionally, a ROC curve analysis revealed that the probability of a random participant from the TD group scoring higher than one from the DLD group was 97.7%.

In a large population study, Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth and Jones (2000) extended the research previously focused on smaller samples and predefined groups. The aim was to evaluate whether a NWR test (the NRT; Dollaghan & Campbell, 1998) could accurately distinguish between TD and DLD groups based on two measures: (a) a DLD diagnosis according to the 'EpiSLI' criteria, and (b) treatment status (currently in language therapy or not).[86] Participants were 581 monolingual English-speaking children in second grade, (M_{age} = 95 months, range: 85–107), with a non-verbal intelligence quotient (NVIQ) > 70. Children who scored below 1.25 SDs on at least two combined language indices (expressive language, receptive language, combined vocabulary, combined grammar or combined narrative score) were assigned to the DLD group. The children were further

[85] Receiver Operating Characteristic (ROC) curves are graphic visualisations of the performance of a binary classifier system (Fawcett, 2006). They are frequently used in medical research when evaluating a method of clinical decision making that has a binary outcome, and visualises how accurately a clinical marker (e.g. NWR) can identify individuals with a condition (e.g. DLD).

[86] According to the EpiSLI criteria, a child (with a NVIQ score of > 84) meets the criteria for DLD if (s)he scores –1.25 SD or less on at least two composite scores (indexes) in three language domains (vocabulary, grammar and narrative ability) and two modalities (comprehension and production) (Tomblin, Records, & Zhang, 1996).

divided into sub-categories based on a matrix of language scores and NVIQ scores: Normal Language (NL: both language scores and NVIQ in the normal range), Specific Language Impairment[87] (SLI: poor language but normal-high NVIQ), Non-specific Language Impairment[88] (NLI: poor language and low NVIQ) and Low Cognitive (LC: normal language but poor NVIQ). Analyses showed that there was a difference in performance between groups, where the children in the DLD group performed significantly worse than the children who had language scores in the normal range. Similar findings were reported for categorisation based on treatment status; the children who were currently in language therapy performed significantly lower compared to their peers not receiving therapy. These results were promising for the prospect of the NRT being a useful diagnostic tool in identifying DLD. In order to investigate this further, Ellis Weismer et al. (2000) conducted likelihood ratio analyses based on three group divisions: (i) second-grade diagnosis (diagnostic category in second grade; all four categories), (ii) presence or absence of language disorder (regardless of NVIQ scores being in the low-normal or high-normal range), and (iii) treatment status.[89] The likelihood ratio analyses showed that results could at best reveal an intermediate to high positive result, and an intermediate to low negative result for both the diagnosis in second-grade categorisation and the presence/absence of language disorder categorisation. This means that both means of categorisation resulted in both over- and underidentification of DLD based on NWR scores, reflecting substantial overlap in performance between the groups. The analysis using treatment status as the gold standard resulted in the best predictive result, however still not with perfect division between the groups. The authors conclude that the NRT cannot be used as a sole measure for diagnostic decision-making.

In sum, several studies comparing the performance on NWR tasks between TD and clinical groups have found significant differences. Studies that make use of predefined groups of children with DLD and TD (based on diagnosis and current language therapy status) often find high effect sizes and little to no overlap in performance between groups. The abovementioned studies have all dealt with monolingual populations. In what follows, studies evaluating the diagnostic validity of NWR in bilinguals will be reviewed. A common denominator for these studies is that they all have four groups of comparison: (i) a group of bilingual children with DLD (or SLI) (BiDLD), (ii) a group of bilingual children with typical language development (BiTD), (iii) a group of monolingual children with DLD (MoDLD), and (iv) a group of monolingual

[87] Equivalent to the term "DLD", with NVIQ in the average/high-normal range.
[88] Equivalent to the term "DLD", with NVIQ in the low-normal range.
[89] Likelihood ratios are commonly used in medicine to evaluate the accuracy of a clinical finding. When used in this context, likelihood ratio is defined as the probability that an individual with a certain condition (i.e. DLD) has the clinical finding (i.e. poor NWR performance), divided by the probability of individuals without the condition having the same finding. Likelihood ratios above 1 are considered as clinically informative (McGee, 2002).

children with typical language development (MoTD). Note that although several studies (Abed Ibrahim & Hamann, 2017; Boerma, Leseman, et al., 2017; dos Santos & Ferré, 2018; Thordardottir & Brandeker, 2013) use the term SLI to refer to the group of children with a language disorder, DLD will be used here for consistency.

Thordardottir and Brandeker (2013) compared the performance on a French NWR task of four groups of Canadian children: BiDLD, BiTD, MoDLD and MoTD, all with a mean of 5;0 years, but with varying ranges. The monolingual children in both the DLD and the TD group spoke French. The BiTD group spoke both English and French, and the children in the BiDLD group spoke a variety of L1s and French as their L2. Both DLD groups performed significantly worse than both TD groups. The TD groups did not differ from each other, nor did the DLD groups. Thus, the NWR task did distinguish between TD and DLD children, regardless of language status (monolingual or bilingual). Clinical accuracy was established using sensitivity and specificity measures. While sensitivity and specificity rates were in the good to excellent range for the monolingual groups (specificity 100%, sensitivity 92%), they were not as high for the bilinguals, where specificity was at 79%, and sensitivity was 85%. Boerma, Chiat, Leseman, Timmermeister, Wijnen and Blom (2015) investigated the clinical accuracy of two NWR tests, a quasi-universal (QU) and a language-specific (LS), on monolingual and bilingual Dutch-speaking 5–6-year-old children. The two DLD groups performed significantly worse than the monolingual groups on both tasks, but there was a difference between the monolinguals and the bilinguals with respect to sensitivity and specificity rates. The LS task had high sensitivity and specificity rates (both 93%) for the monolinguals. For bilinguals, specificity was the same as for monolinguals, but sensitivity was only at 63%. This means that more than 35% of the bilingual children who had DLD were not identified as such by the LS task. For the QU task however, the specificity level was the same (93%), while the sensitivity rate was 83%, and thus 20% higher than the specificity of the LS task. The authors conclude that while the LS task had higher specificity and sensitivity rates for monolinguals than the QU task did, the QU task was a more accurate tool for identifying DLD in bilinguals. Abed Ibrahim and Hamann (2017) and dos Santos and Ferré (2018) both investigated the clinical accuracy of two language versions of a NWR task with 1–3 syllables. In Abed-Ibrahim and Hamann (2017), 54 children (monolingual German-speaking children and bilingual children with Turkish or Arabic as their L1 and German as their L2) between 5;6–9;3 participated in the study. For monolinguals, there was a significant difference in mean NWR scores between MoTD and MoDLD and no overlap between the groups. For the bilinguals however, although there was a significant difference in mean scores between BiTD and BiDLD, there was also a substantial overlap in range, where the lowest scoring child in the BiTD group performed close to the mean of the BiDLD group. With a cut-off

score of 60% items repeated correct, both sensitivity (100%) and specificity (95%) rates were high. In the study by dos Santos and Ferré (2018) a total of 67 children between 5;6–8;6 participated, French monolinguals as well as bilinguals speaking English or Arabic as their L1 and French as their L2. Although dos Santos and Ferré report a significant difference between the BiTD and the BiDLD group, diagnostic accuracy was poorer than for the bilingual groups in Abed-Ibrahim and Hamann (2017). One possible explanation for this difference was that dos Santos and Ferré's sample showed a clear tendency for a ceiling effect, which was not prevalent in Abed Ibrahim and Hamann's sample. At a cut-off of –1.28 SD, specificity was high (93%), while sensitivity was at an intermediate level (77%). This means that although the task rarely misclassifies TD children as DLD, it is not sufficiently sensitive to identify all children with DLD.

In sum, studies on predefined groups (DLD vs. TD) of monolingual children generally report high rates of sensitivity as well as specificity in NWR tasks. Evidence from a population study of monolinguals seems to confirm that NWR is a reliable tool for identifying DLD, however sensitivity as well as specificity rates are lower than in studies with predefined groups. Findings from studies of bilingual populations suggest that bilinguals may be at a disadvantage compared to monolinguals on language-specific NWR tasks. Sensitivity and specificity rates of NWR tasks are generally lower for bilinguals than to monolinguals, which increases the risk of both over- and underdiagnosis. Therefore, researchers and clinicians must be cautious when choosing NWR items and tasks for bilinguals, since NWR tasks are not equally unaffected by previous language experience. To conclude, although poor NWR performance may be promising as a clinical marker for DLD in bilinguals, more research is needed on bilingual populations. Additionally, NWR cannot be used as a sole diagnostic test for DLD, neither in bilingual or monolingual populations.

6.2 Results

This section starts with a description of the analyses (section 6.2.1), followed by relevant background information for the children in the clinical study (section 6.2.2), vocabulary performance (section 6.2.3), non-word repetition performance (section 6.2.4), and finally a summary of the results (section 6.2.5).

The following research questions are asked:

- What are the reported early language development, language skills, communicative behaviours and social characteristics of the Arabic-Swedish-speaking children with a DLD diagnosis, as described by parents, SLPs and teachers?

- What are the vocabulary skills in Arabic and Swedish of the children with a DLD diagnosis?
- How do the children with a DLD diagnosis perform on non-word repetition tasks?
- Are there similarities or differences between the children with a DLD diagnosis and the children in the cross-sectional study concerning vocabulary or non-word repetition performance?

6.2.1 Analyses

In this chapter, the performance of the children in the clinical group on the vocabulary and NWR tasks will be compared to that of the cross-sectional group. First, the raw score for each child will be compared to the range of the corresponding age group in the cross-sectional sample, to see whether they perform within the range of the age group. In addition, the children's raw scores will be transformed into z-scores, allowing further comparison with the performance of their peers. A more detailed description of z-scores and clinically informative thresholds will be provided below. Second, the performance on the NWR tasks of the DLD children will be compared to the TD children to see whether they perform similarly or differently on NWR items with different properties. The comparison will be made separately for each task, item length (number of syllables) as well as items with vs. without clusters (for the LS-Swe and NWRT-Leb tasks).

A *z-score* shows how much a particular score deviates from the sample mean. The standard deviation (SD) is a measure of the variation (or dispersion) of all scores obtained by the individuals in the group. In a normal distribution, 34% of the values fall within one SD from the mean, and 13.5% fall between 1 and 2 SDs from the mean. A child who receives a z-score of –1 performs one standard deviation below the mean of the group, meaning that they score lower than ca 85% of the whole sample. In this study, a z-score of –1.25 will be considered as a *potentially clinically informative threshold*. Receiving a z-score of –1.25 or lower means that the child performs within the lowest-scoring 10% of the group. The choice of threshold follows the cut-off proposed by Tomblin, Records and Zhang (1996) and Tomblin, Records, Buckwalter, Zhang, Smith and O'Brien (1997), stating that children who receive a z-score of –1.25 on at least two composite language scores meet the criteria for DLD (or in their words, SLI).[90] We need to bear in mind of course, that the cut-off scores recommended in these studies concern *composite measures* of language scores across language domains (morphosyntax, vocabulary, discourse) and modalities (comprehension, production).

[90] Yet other studies propose thresholds of –1.5 (Norbury et al., 2016), and the clinical guidelines in some countries state that a z-score of –2.0 is required to qualify for a DLD diagnosis (Thordardottir, 2015).

Furthermore, they are based on large sets of data of monolingual children, using standardised and norm-referenced language tests. Thus, the direct applicability of the same cut-off score for single language task scores for bilingual children is debatable. However, as discussed in the introduction, there is no evidence to date that can tell us where the relevant cut-off could be for bilinguals, let alone one that also takes exposure factors into account (such as age of onset and daily language exposure).

6.2.2 Early language development, current language proficiency and social characteristics of the children with a DLD diagnosis

This section contains an overview of the background information, language skills and social characteristics of the 11 children with a DLD diagnosis. The participants have also been described in more detail in section 3.1.2, Chapter 3. The information was obtained from parental questionnaires and interviews with parents, SLPs and (pre)school teachers. Table 6.1 provides information about their general language development, age at first word, age at first word combination and age of onset for Arabic and Swedish, as reported by the parents. Age of onset for Arabic was at birth for all children in the clinical study, which matches the pattern in the cross-sectional group where all children but one had an age of onset at birth. Age of onset for Swedish varied considerably, just as in the cross-sectional group. Six children in the clinical study (55%) started hearing Swedish before age three, and five children (45%) after the age of three, which matched the patterns found in the cross-sectional sample. According to parental reports, most children in the clinical group (10/11) had even patterns of relative exposure to each language – i.e. 40/60%, 50/50% or 60/40% daily exposure to Arabic and Swedish respectively. One child (BiAraLI-02) had more exposure to Arabic (80%) than Swedish (20%). These patterns were in line with the children in the cross-sectional study, where a majority (ca 75%) had even patterns of relative daily exposure. There was a substantial difference between the clinical group and the cross-sectional group with respect to the proportion of Arabic varieties though. While 6/11 (55%) of the children in the clinical study spoke an Iraqi variety, only 17.2% did so in the cross-sectional group. Five children (5/11, 45%) in the clinical group spoke a Levantine variety (Syrian or Palestinian), compared to almost 80% in the cross-sectional group.

As shown in Table 6.1, the parents of all children but two (BiAraLI-04 and BiAraLI-09) reported a late general language development in Arabic for their child. The information about general language development in Swedish is more varied. The parents of two children (BiAraLI-02 and BiAraLI-11) with delayed language development in Arabic stated that their children had a normal language development in Swedish. The larger variation and in some

cases, missing information about language development and developmental milestones in Swedish is likely a reflection of the fact that age of onset was generally later in Swedish. Furthermore, since all parents were L2 learners of Swedish, they may not be able to assess their children's language skills in Swedish as accurately as in the home language. As can be seen in Table 6.1, six children (54.5%) have a reported onset of the first word (BiAraLI-03), of the first word combination (BiAraLI-04, BiAraLI-07 and BiAraLI-11) or both (BiAraLI-06, BiAraLI-10) that falls outside the typical range. It is generally considered typical that the first word appears around 12 months, and the first word combination appears before 24 months (Paradis et al., 2010; Trauner et al., 2000). By comparison, for 24 (24.2%) children in the cross-sectional study, either the first word appeared after 12 months, the first multi-word utterance appeared after 24 months, or both.[91]

[91] For 21 children (21.2%) in the cross-sectional study the first word appeared after 12 months, and for nine children (9.1%) the first multi-word utterance appeared after 24 months. The number of missing values in the cross-sectional group was 7 (7.1%) for onset of the first word and 9 (9.1%) for onset of the first multi-word utterance.

Table 6.1. Age at testing, general language development, age at first word and first word combination and age of onset of Arabic and Swedish for the children with a DLD diagnosis according to the parents.

		Arabic				Swedish			
Child	**Age**	**General language development**	**1st word**	**1st word comb.**	**AoO Arabic**	**General language development**	**1st word**	**1st word comb.**	**AoO Swedish**
BiAraLI-01	6;8	late	0;9	---	at birth	late	5;0	6;0	5;0–6;0
BiAraLI-02	6;1	late	1;0	2;0	at birth	normal	1;6	2;0	1;0–2;0
BiAraLI-03	5;7	late	1;6	2;0	at birth	late	1;6	2;0	1;0–2;0
BiAraLI-04	6;0	normal	0;9	2;9	at birth	early	---	---	4;0–5;0
BiAraLI-05*	7;3	late	1;0	1;6	at birth	---	2;0	4;0	4;0–5;0
BiAraLI-06	5;4	late	2;0	3;0	at birth	late	2;0	3;0	1;0–2;0
BiAraLI-07	6;1	late	0;8	3;6	at birth	late	1;3	4;6	1;0–2;0
BiAraLI-08	7;1	normal	1;3	2;0	at birth	normal	3;6	---	3;0–4;0
BiAraLI-09	6;7	late	1;2	2;0	at birth	late	1;2	2;0	1;0–2;0
BiAraLI-10	5;0	late	2;0	4;6	at birth	late	2;6	4;6	1;0–2;0
BiAraLI-11	6;4	late	1;2	3;6	at birth	normal	4;10	5;0	4;0–5;0

Note. '---' = missing information. *The first word in Swedish is reported to have appeared before the age of onset of regular exposure to Swedish. There could be different explanations for this seemingly contradictory information. BiAraLI-05 was reported to not having attended preschool and Arabic was the language used in the home. Even though age of onset to regular exposure to Swedish was not until between age 4 and 5, the child may have received some exposure to Swedish words for instance via media or in the local community and started to express words in Swedish as a result of this.

The parents were also asked whether anyone in the family currently or previously had difficulties with language and/or reading and writing. Six children (55%) were reported to have a relative with such difficulties, which was a considerably higher proportion than that of the cross-sectional sample (7%). BiAraLI-03 had a younger brother who had received an autism diagnosis. BiAraLI-04 had an unnamed relative who did not finish school. The twin sister of BiAraLI-05 has autism, and does not speak. The parents of BiAraLI-06 say that they think that his younger brother has similar difficulties as he does (delayed language development). BiAraLI-08 has an aunt with reading and writing difficulties. According to the parental interview, the older sister of BiAraLI-10 previously had language difficulties and used to see an SLP, but any (possible) diagnosis is not known. She does not currently have difficulties and does well in school, according to the parents.

During the interviews, parents, SLPs and teachers were asked to characterise the children in terms of their language abilities, difficulties and social behaviour. Detailed reports for each child can be found in the methods chapter (section 3.1.2.3 from parents, section 3.1.2.4 from SLPs and section 3.1.2.5 from teachers). The following contains a brief summary of some common characteristics and behaviours. Many children in the clinical sample are reported to have poor language comprehension, for instance having difficulties understanding instructions and needing repetitions and rewordings. Many children have difficulties expressing themselves, for instance because of deficient morphosyntactic skills, difficulties with explaining things or having limited narrative abilities. A couple of children are described as having pronunciation difficulties. Vocabulary is also an area for concern, as many children are reported to have a weak vocabulary, and in some cases, word finding difficulties. A couple of children have severe communication problems, which often lead to misunderstandings, which has a considerable negative effect on their social relations and leads to many conflicts. Many of the children are also reported to have attention difficulties, particularly during therapy/assessment or during activities at school that require listening. Some problematic behaviours include the use of nonsense words that no one else understands, being meticulous about routines and playing exclusively with Arabic-speaking children. Four children (BiAraLI-02, BiAraLI-03, BiAraLI-09 and BiAraLI-10) have undergone assessment for suspected neuropsychiatric conditions (autism or ADHD) or intellectual disability.

6.2.3 Vocabulary

The vocabulary skills of the children in the clinical study were assessed using the Cross-linguistic Lexical Task (Haman et al., 2015) in both Arabic and Swedish, mirroring the setup of the cross-sectional study.

Most children in the clinical study perform within the range of their age peers, however some children scored below the minimum score of their age peers in the cross-sectional sample: BiAraLI-05 for Arabic comprehension, BiAraLI-03 and BiAraLI-10 for both Arabic comprehension and production and BiAraLI-08 for Swedish production.[92] No children in the clinical sample reached the maximum score obtained in their age groups. In fact, most children scored far below it. Furthermore, most children in the clinical sample scored below the mean of the corresponding age group, except for some children who scored at or above the mean in one or more tasks: BiAraLI-08 for Arabic comprehension, BiAraLI-06 and BiAraLI-09 for Swedish comprehension and production, BiAraLI-04 for Swedish comprehension, and BiAraLI-11 for all tasks apart from Swedish production.

Next, all raw scores were converted into z-scores based on the scores for each respective age group in the cross-sectional study. This way it is possible to compare the performance of each child in the clinical study to see how much it diverges from the mean score of the respective age group in the cross-sectional study.

As shown in Table 6.2, most children in the clinical study had negative z-scores in both comprehension and production of both Arabic and Swedish. However, there were some exceptions: BiAraLI-08 had positive z-scores in Arabic comprehension, BiAraLI-04 scored at the mean in Swedish comprehension. BiAraLI-06 and BiAraLI-09 had positive z-scores in both comprehension and production of Swedish. Finally, BiAraLI-11 had positive z-scores in comprehension and production of Arabic, and scored at the mean for her age group in Swedish comprehension. Furthermore, seven of the children had z-scores in one or more subtasks that were at or below –1.25 (BiAraLI-01, BiAraLI-02, BiAraLI-03, BiAraLI-05, BiAraLI-06 and BiAraLI-10. For the children in the clinical study as a whole, z-scores were generally lower in Arabic than in Swedish.

[92] For a detailed comparison of vocabulary scores of the children in the clinical sample in comparison to the cross-sectional sample (means and ranges), please refer to Table A6.1 in the Appendix.

Table 6.2. Ages and z-scores of the children in the clinical sample in Arabic and Swedish CLTs in relation to the cross-sectional means. Z-scores at –1.25 or below are marked with '*' and boldface.

		Arabic		Swedish	
	Age	**Comp**	**Prod**	**Comp**	**Prod**
BiAraLI-01	6;8	**–1.53***	**–1.25***	**–1.76***	**–1.62***
BiAraLI-02	6;1	**–1.40***	–0.72	–1.07	–0.66
BiAraLI-03	5;7	**–3.28***	**–2.40***	–1.13	–0.58
BiAraLI-04	6;0	–0.20	–0.49	0.01	–0.31
BiAraLI-05	7;3	**–2.60***	**–1.53***	–0.86	–0.40
BiAraLI-06	5;4	**–2.07***	**–1.73***	0.41	0.33
BiAraLI-07	6;1	–0.34	–0.64	–0.78	–0.49
BiAraLI-08	7;1	0.71	–0.23	**–2.03***	**–2.62***
BiAraLI-09	6;7	–0.07	–0.26	0.99	0.64
BiAraLI-10	5;0	**–3.89***	**–2.73***	–1.13	–0.69
BiAraLI-11	6;4	0.33	0.49	0.01	–0.66

Figure 6.1 shows z-scores for Arabic and Swedish comprehension (A) and z-scores for Arabic and Swedish production (B) for the children in the clinical sample in relation to the children in the cross-sectional study. As shown in Figure 6.1A, a majority (6) of the children in the clinical sample had a z-score below –1.25 for comprehension in one language, but only one child (BiAraLI-01) received a comprehension z-score below –1.25 in both languages. It was similar for the production scores; as shown in Figure 6.1B, five of the children in the clinical sample received a z-score below –1.25 for production in one language, but only one child (also BiAraLI-01) had z-scores at or below –1.25 in both languages. For both comprehension and production, there was a notable overlap between children in the clinical sample and children in the cross-sectional sample.

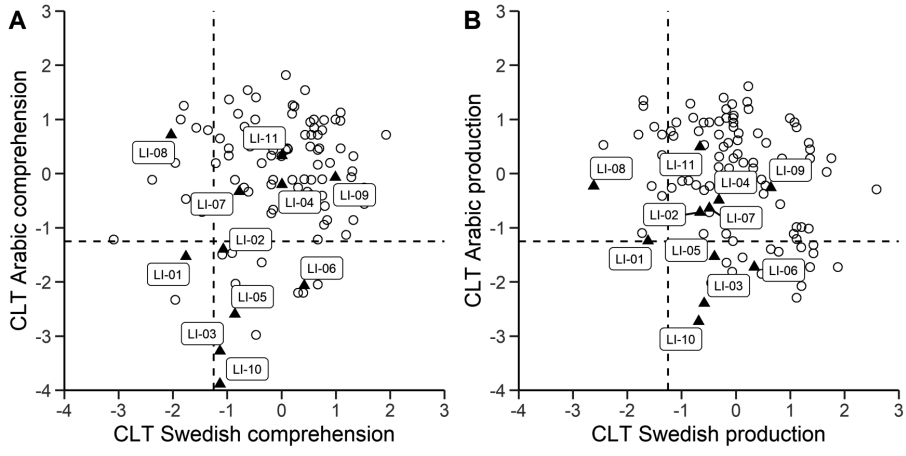

Figure 6.1. Scatterplots showing the z-scores of (A) Arabic and Swedish CLT comprehension, and (B) Arabic and Swedish CLT production of the children in the clinical study (triangles and labels) compared to the children in the cross-sectional study (circles). Dashed lines at –1.25.

6.2.4 Non-word repetition

The non-word repetition skills of the children in the clinical study were assessed using the same tasks as in the cross-sectional study. These tasks were the language-specific Swedish task (LS-Swe; Radeborg et al., 2006), the Swedish (QU-Swe) and the Arabic (QU-Ara) versions of the quasi-universal NWR task (Chiat, 2015), and the Non-word Repetition Task-Lebanese (NWRT-Leb; Abou Melhem et al., 2011). The responses were transcribed and scored in the same way as in the cross-sectional study. Two children (BiAra-LI-04, BiAra-LI-11) had systematic phonological substitutions (e.g. fronting, /k/ > /t/). Scoring was adjusted to account for this.

A majority of the children (6/11) in the clinical sample scored within the range of their age peers on all NWR tasks. Three children scored below the range of their age peers on two tasks (BiAraLI-01: QU-Swe and NWRT-Leb; BiAraLI-02: LS-Swe and NWRT-Leb; BiAraLI-10: QU-Swe and NWRT-Leb), one child scored below the cross-sectional range on three tasks (BiAraLI-06: LS-Swe, QU-Swe and NWRT-Leb), and one child scored below the cross-sectional range on one task (BiAraLI-11: NWRT-Leb).[93]

Next, for comparison, all raw scores were converted into z-scores based on the scores for each respective age group in the cross-sectional study. Z-scores for each child on each NWR task are displayed in Table 6.3. Most children (6/11) in the clinical study had negative z-scores in most tasks, but there were some exceptions. Two children had positive z-scores in two tasks (BiAraLI-

[93] For a detailed comparison of NWR scores of the children in the clinical sample in comparison to the cross-sectional sample (means and ranges), please refer to Table A6.2 in the Appendix.

03: QU-Swe, QU-Ara; BiAraLI-05: QU-Swe, NWRT-Leb), and two children had positive z-scores on one task (BiAraLI-04: QU-Swe; BiAraLI-07: LS-Swe). Finally, BiAraLI-09 had positive z-scores (well above the mean) in three out of four tasks (LS-Swe, NWRT-Leb and QU-Ara), and scored just slightly below the mean on the QU-Swe task. By contrast, six of the children had z-scores that fell below –1.25 in at least two tasks. Three children had z-scores below –1.25 in all four tasks (BiAraLI-01, BiAraLI-02 and BiAraLI-06). One child had z-scores below –1.25 in three tasks (BiAraLI-10: QU-Swe, NWRT-Leb and QU-Ara), and two children in two tasks (BiAraLI-08 and BiAraLI-11: LS-Swe, NWRT-Leb).

Table 6.3. Ages and NWR z-scores of the children in the clinical sample in relation to the cross-sectional means. Z-scores at –1.25 or below are marked with '*' and boldface.

	Age	LS-Swe	QU-Swe	NWRT-Leb	QU-Ara
BiAraLI-01	6;8	**–1.28***	**–2.17***	**–2.46***	**–2.14***
BiAraLI-02	6;1	**–2.30***	**–1.76***	**–3.05***	**–1.69***
BiAraLI-03	5;7	–0.64	0.37	–0.35	0.17
BiAraLI-04	6;0	–0.51	0.73	–0.67	–0.32
BiAraLI-05	7;3	–0.59	0.61	0.02	–0.39
BiAraLI-06	5;4	**–2.07***	**–4.56***	**–3.26***	**–1.79***
BiAraLI-07	6;1	0.26	–0.51	–0.37	–0.32
BiAraLI-08	7;1	**–1.76***	–1.08	**–1.92***	–0.78
BiAraLI-09	6;7	1.02	–0.10	0.82	0.60
BiAraLI-10	5;0	–0.64	**–2.59***	**–3.26***	**–2.28***
BiAraLI-11	6;4	**–2.05***	–0.51	**–3.35***	–0.77

Figure 6.2 shows the z-scores for the four NWR tasks for the children in the clinical sample in relation to the children in the cross-sectional sample. As displayed in the figure, there is a notable overlap in z-scores between children in the clinical sample and children in the cross-sectional sample.

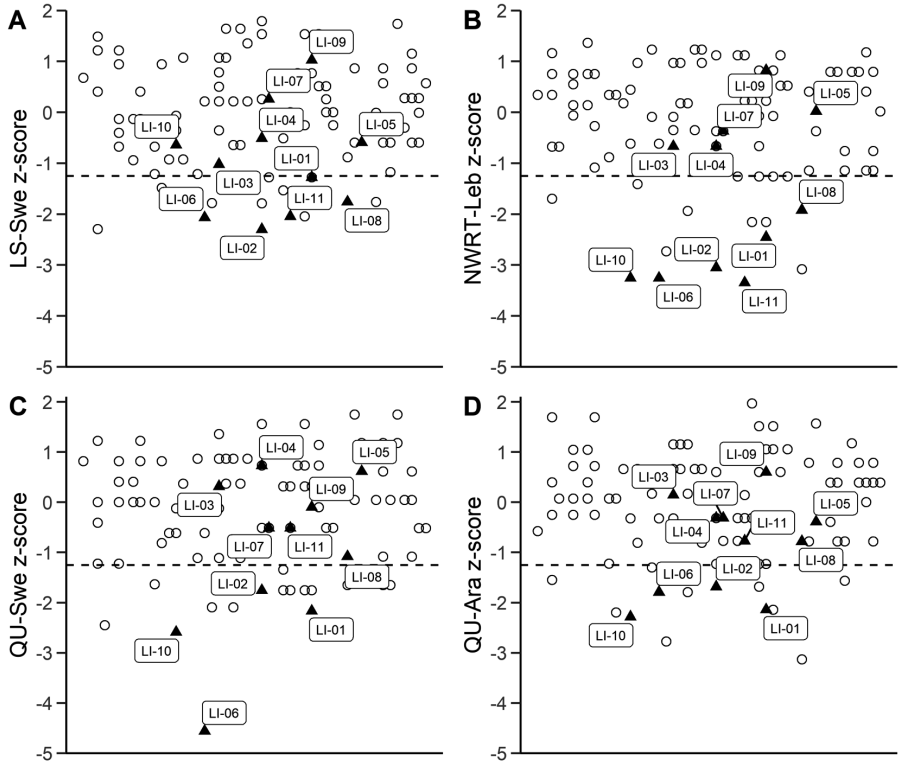

Figure 6.2. Scatterplots showing the z-scores of the (A) LS-Swe, (B) NWRT-Leb, (C) QU-Swe and (D) QU-Ara tasks of the children in the clinical study (triangles and labels) compared to the children in the cross-sectional study (circles). Dashed lines at –1.25.

6.2.4.1 NWR performance in relation to item types

Next, performance on the NWR tasks was analysed for non-word items of different lengths and types. Since the age range of the children in the clinical group 5;0–7;3) was narrower than that of the cross-sectional group (4;0–7;11), the 22 4-year-olds from the cross-sectional study were not included in the comparison. For each task, the accuracy (% correct responses) was compared between the two groups. Comparisons were made separately for items with the same number of syllables in the same task (LS-Swe, QU-Swe and QU-Ara: 2–5 syllables, NWRT-Leb: 1–3 syllables). For LS-Swe and NWRT-Leb, which both had items with and without consonant clusters, the comparison was done separately for these two categories of items. Since there is a large difference in sample size between the two groups ($N_{cross\text{-}sectional} = 77$, $N_{clinical} = 11$), the analyses will be exclusively descriptive and group differences will not be analysed statistically. Because of the uneven sample sizes, a single data point has a larger impact on the percentage in the clinical sample than in the cross-sectional sample. Tables with accuracy levels for each task and item

type (number of syllables and presence vs. absence of clusters) are available in the Appendix (Table A6.3-A6.5).

First, accuracy levels for the LS-Swe task will be analysed. As can be seen in Figure 6.3, there were differences in performance between the TD group and the DLD group. For items containing no clusters, accuracy levels were similar for disyllabic items (TD 80%, DLD 75%), but the DLD children scored lower than the TD group on all items with more than two syllables. The largest difference in accuracy levels between the two groups was for three-syllable items (TD 82% vs. DLD 52%), compared to four-syllable items (TD 84% vs. DLD 64%) and five-syllable items (TD 44% vs. DLD 27%). Interestingly, the accuracy was higher in the DLD group for four-syllable items (64%) than three-syllable items (52%). For items containing clusters, accuracy for disyllabic items was similar and in fact somewhat higher for the DLD group (91%) compared to the TD group (84%). For items with clusters and more than two syllables, the DLD group performed consistently lower than the TD group. For three-syllable items, this difference was fifteen percentage points (TD 45% vs. DLD 30%), and for four-syllable items it was twenty percentage points (TD 26% vs. DLD 6%). For five-syllable items the accuracy of the TD group was low at 16%, but there was a clear floor effect for the DLD group, with 0% correct answers.

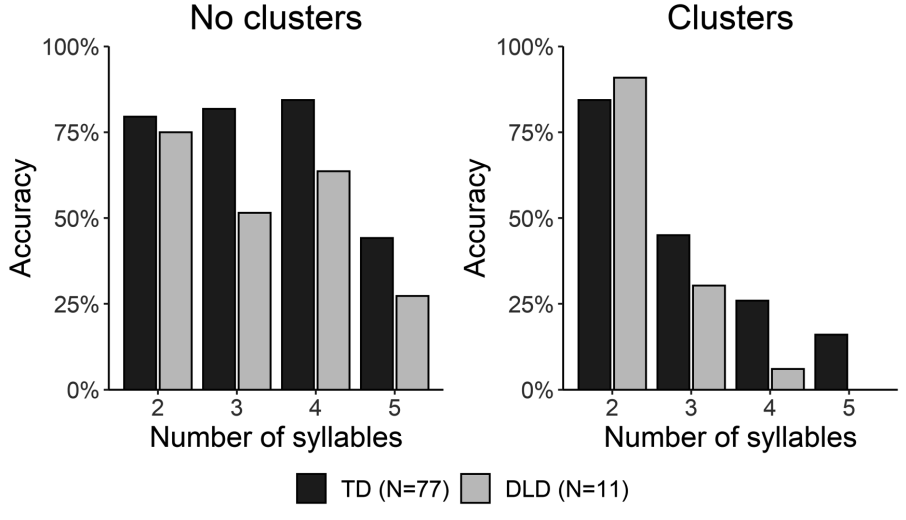

Figure 6.3. Accuracy (% correct answers) for the language-specific Swedish task (LS), by number of syllables (2–5) and clusters (present or not) for children in the cross-sectional study (age groups 5–7) and the clinical study (age 5;0–7;3).

Figure 6.4 shows accuracy levels for the NWRT-Leb task. In general, accuracy was high. As shown in Figure 6.4, accuracy levels are higher for the TD group than in the DLD group in monosyllabic (TD 99% vs. DLD 70%)

and two-syllable (TD 95% vs. DLD 71%) items. However, for three-syllable items, accuracy was the same (80%) for both groups. Thus, while accuracy levels were decreasing as the number of syllables increased in the TD group, the accuracy level for three-syllable items was higher (80%) than disyllabic items (71%) in the DLD group. For items containing clusters, accuracy for monosyllabic items was high in the TD group (92%), and slightly lower (86%) in the DLD group. For two-syllable items, accuracy was still high (88%) in the TD group, but notably lower for the DLD group (58%). The same pattern was evident for three-syllable items, with higher performance in the TD group (74%) than in the DLD group (51%). When comparing overall patterns of accuracy for items of the same syllable length, the TD group had consistently higher accuracy in items containing no clusters than those containing clusters, which was not the case for the DLD group. For monosyllabic items, accuracy was higher for items with clusters than those without clusters (monosyllabic without clusters: 70%, monosyllabic with clusters: 86%).

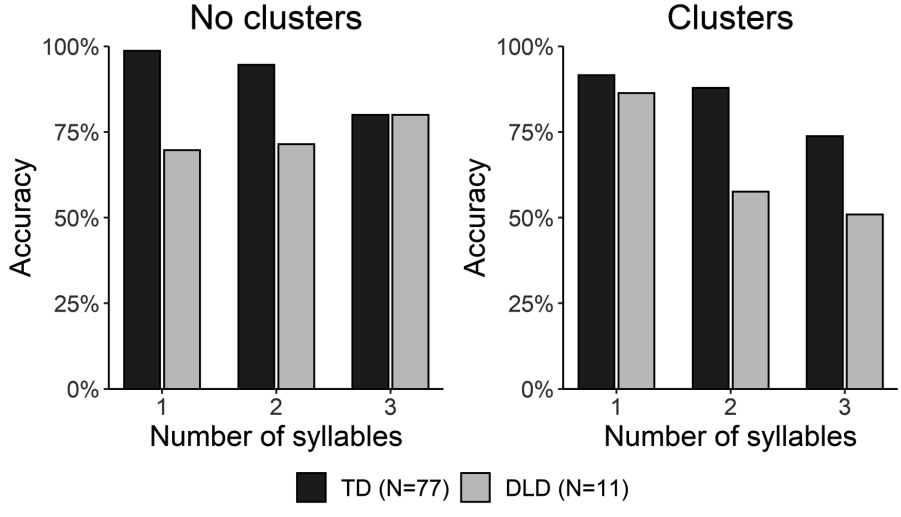

Figure 6.4. Accuracy (% correct answers) for the Non-word Repetition Task Lebanese (NWRT-Leb), by number of syllables (1–3) and clusters (present or not) for children in the cross-sectional study (age groups 5–7) and the clinical study (age 5;0–7;3).

Finally, accuracy levels were analysed for the two quasi-universal tasks. As shown in Figure 6.5, accuracy levels were consistently lower for the DLD group compared to the TD group at all syllable lengths on both tasks, apart from two-syllable items in the QU-Ara task, where the DLD group scored on par with, and in fact slightly higher (93%) than the TD group (90%). In both groups, scores generally decreased with increased number of syllables, apart from the TD group in two-syllable items (90%) compared to three-syllable items (93%).

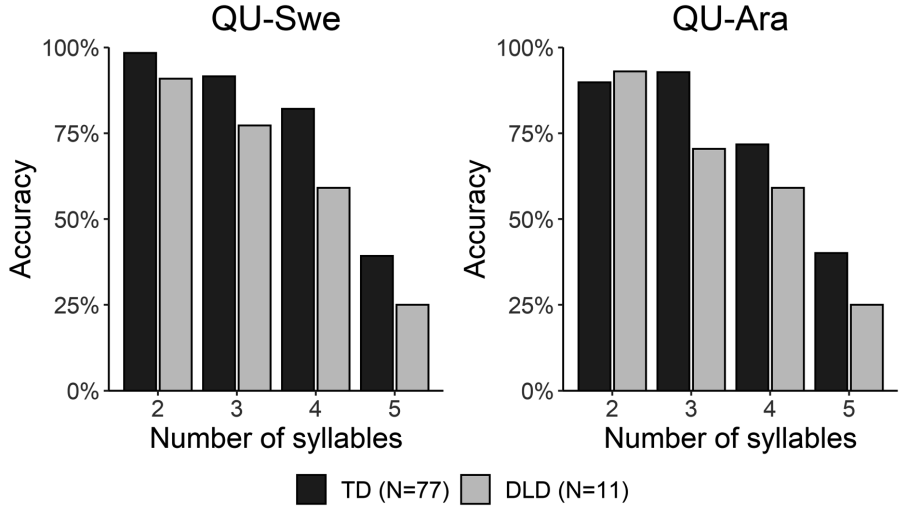

Figure 6.5. Accuracy (% correct answers) for the quasi-universal Swedish task (QU-Swe) and the quasi-universal Arabic task (QU-Ara), by number of syllables (2–5) for children in the cross-sectional study (age groups 5–7) and the clinical study (age 5;0–7;3).

6.2.5 The clinical study: summary

In what follows, a short summary will be provided with respect to the language, communication and social characteristics of the children with a DLD diagnosis, as well as their performance on Arabic and Swedish CLTs and the four NWR tasks (LS-Swe, QU-Swe, QU-Ara and NWRT-Leb).

According to parental reports, 9/11 of children had a delayed language development in the first language (Arabic), and 6/11 had a late development of the second language (Swedish). Six children had an onset of the first word or the first multi-word utterance that fell outside the typical range, and six children had a relative with language or literacy problems. Most children were reported to have problems with both comprehension (e.g. having trouble understanding instructions) and production (such as having deficits in expressive morphosyntax or being difficult to understand). Five children (BiAraLI-01, BiAraLI-02, BiAraLI-06, BiAraLI-08 and BiAraLI-10) were reported to have severe communication difficulties, which lead to frequent misunderstandings and conflicts, while others were described as having milder problems that did not affect social relations. Four children had undergone assessment for suspected neuropsychiatric diagnoses or intellectual disability.

The children in the DLD group generally performed below the mean but within the range for their age group on both vocabulary comprehension and production of Arabic and Swedish. Six children had a z-score below –1.25 for comprehension in one of their languages, but only one child received a z-score

below –1.25 in both languages. The picture was similar for production. Five of the children in the clinical sample received a z-score below –1.25 in one language, but only one child had z-scores at or below –1.25 in both languages. Although the children in the clinical study generally had negative z-scores, there was considerable overlap between the DLD group and the TD group on all tasks.

With respect to performance on the four NWR tasks, six children in the DLD group scored within the range for their age group on all four NWR tasks. Six of the children in the clinical sample had negative z-scores in a majority of the tasks, and the same number of children had z-scores that fell below –1.25 in at least two tasks. One child had good scores in all tasks (at the mean or well above). Accuracy on the NWR tasks was also reported separately for non-word items of different lengths and types. The children in the DLD group generally had lower accuracy rates compared to the children in the cross-sectional group, except for two-syllable items containing clusters in the Swe-LS task, two-syllable items in the QU-Ara task and three-syllable items with no clusters in the NWRT-Leb task.

6.3 Discussion

In this chapter, the following research questions were asked:

- What are the reported early language development, language skills, communicative behaviours and social characteristics of the Arabic-Swedish-speaking children with a DLD diagnosis, as described by parents, SLPs and teachers?
- What are the vocabulary skills in Arabic and Swedish of the children with a DLD diagnosis?
- How do the children with a DLD diagnosis perform on non-word repetition tasks?
- Are there similarities or differences between the children with a DLD diagnosis and the children in the cross-sectional study concerning vocabulary or non-word repetition performance?

In this section, findings are discussed for early language development, current language proficiency and social characteristics (section 6.3.1), vocabulary performance in both languages (section 6.3.2), and performance on the NWR tasks (section 6.3.3). In section 6.3.4, some individual children from the clinical group (section 6.3.4.1) and the cross-sectional group (section 6.3.4.2) are discussed.

6.3.1 Early language development, current language proficiency and social characteristics of the children with a DLD diagnosis

Functional language, communication and social characteristics was explored via parental questionnaires (see section 3.1.2.2 and 3.1.2.3 in the methods chapter), as well as in interviews with parents (section 3.1.2.3), SLPs (section 3.1.2.4) and teachers (section 3.1.2.5). The information was also summarised in section 6.2.2 (this chapter).

According to parental report, all but two children had a late development in the first language (Arabic), but only six children were reported to have a late development in the second language (Swedish). Some parents (BiAraLI-01, BiAraLI-02, BiAraLI-05 and BiAraLI-09) reported that their child had a late language development despite the fact that the first word and the first word-combination appeared within the expected time frame (i.e. by 12 and 24 months respectively). The opposite was also true; one parent reported that their child had normal language development despite the fact that the first word combination appeared well after the expected time (BiAraLI-04). When asked to rate their children's current language proficiency (separately for comprehension and production of Arabic and Swedish), parents generally did not describe their children's proficiency as poor. However, when interviewed more in detail about their children's language skills, the majority of the parents stated that the child had communication problems, particularly in Arabic since many parents said that they could not evaluate their child's Swedish because they had limited proficiency in the language. Thus, parental reports concerning the children's language early language development and current language skills were informative and in line with previous research showing an association between parental reports of poor language proficiency and DLD (Restrepo, 1998). However, it is important to keep in mind that parents may not be able to make a valid estimation of their child's language skills in the majority language if they have limited proficiency themselves in that language.

Although not analysed statistically, the proportion of children who had an onset of the first word or multi-word utterance that fell outside the expected time frame was considerably higher in the clinical sample compared to the cross-sectional sample. This is in line with previous studies, which have found that at group level, the first word and multi-word utterance often appear later in children with DLD compared to their typically developing peers in both monolinguals (Trauner et al., 2000) and bilinguals (Paradis et al., 2010). Furthermore, the proportion of children with reported heredity for language or literacy problems was considerably higher in the DLD group compared to the cross-sectional group. This finding is also in accordance with earlier studies that have found that children with DLD are more likely to have close relatives

with language or literacy problems compared to their typically developing peers (Kalnak et al., 2012; Restrepo, 1998).

In the current study, only comprehension and production of vocabulary and performance on NWR tasks was assessed formally.[94] However, parents, SLPs and teachers were asked to characterise the children's receptive and expressive language skills as well as their vocabulary and communicative behaviour. The parents, SLPs and teachers generally gave similar views about the children's language skills and communicative behaviour, but with some exceptions, notably BiAraLI-04, BiAra-LI-05 and BiAra-LI-09, which will be further discussed in section 6.3.4. Most children were reported to have comprehension as well as production difficulties. All children were also reported to have a weak vocabulary, and five children had word-finding difficulties. When asked to describe the comprehension difficulties, informants often responded that the child had weak language comprehension and difficulties understanding instructions. Language comprehension difficulties are frequently reported in children with DLD (Bishop et al., 2017; Norbury et al., 2016; Skarakis-Doyle, Dempsey, & Lee, 2008; Tomblin et al., 1997). One cause of poor language comprehension is having a weak comprehension of complex morphosyntax, which is commonly found in children with DLD (Friedmann & Novogrodsky, 2004; Marshall et al., 2007). There were several statements about the children's expressive language difficulties which indicated that the child had deficient morphosyntactic skills, such as "the child is not speaking in complete sentences", "sometimes his sentences are funny" and "he has difficulties expressing himself in a grammatically correct manner". Deficient expressive morphosyntax has also been described extensively in the literature on children with DLD, for instance morphological errors (Conti-Ramsden et al., 2001; Hansson et al., 2003), omissions of grammatical morphemes (Hansson & Leonard, 2006; Leonard et al., 1997) and word order errors (Hansson & Nettelbladt, 1990; Lely & Battell, 2003).

Although none of the children in the clinical study had any other diagnoses of neurodevelopmental disorders except for DLD, two children had undergone assessment for suspected autism (BiAraLI-09 and BiAraLI-10), two had been referred for assessment due to suspicions of an unspecified neuro-developmental disorder (BiAraLI-01, BiAraLI-03), and one child (BiAraLI-02) had undergone an assessment for suspected intellectual disability. There were also many children in the DLD group who were described as having attention difficulties or being easily distracted (BiAraLI-01, BiAraLI-02, BiAraLI-03, BiAraLI-05, BiAraLI-06, BiAraLI-07 and BiAraLI-10). It is not unusual that children with DLD also have symptoms associated with other

[94] As previously mentioned in the methods chapter, the children also completed a narrative task, the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012), which is not reported in this study.

neurodevelopmental disorders, such as social, emotional or behavioural problems (Norbury et al., 2016). In fact, there is a high degree of overlap between DLD and conditions such as ADHD and developmental dyslexia (Bishop et al., 2017; Mueller & Tomblin, 2012; Pennington & Bishop, 2009; Ramus et al., 2013).

6.3.2 Vocabulary

Vocabulary knowledge was assessed in both Arabic and Swedish with the Cross-linguistic Lexical Task (Haman et al., 2015). At group level, the children in the clinical sample scored below the cross-sectional mean, but within the range for their age group on comprehension and production of Arabic and Swedish. Two children scored below the cross-sectional range in Arabic comprehension and production (BiAraLI-03 and BiAraLI-10), one child in Arabic comprehension (BiAraLI-05) and one child in Arabic production (BiAraLI-08). There were however large individual differences, and not all children had poor scores in both languages. Five of the children with a DLD diagnosis received z-scores below –1.25 in Arabic comprehension, and one did so for Swedish comprehension. What is more, only one child received a z-score below –1.25 in *both* languages. The picture was similar for production: Four children had a z-score below –1.25 in Arabic, one child had a z-score below –1.25 in Swedish, and only one child had a z-score below –1.25 in *both* languages. It is frequently argued that language difficulties must manifest in both languages in order to qualify for a DLD diagnosis (Kohnert, 2010; Salameh, Nettelbladt, & Gullberg, 2002). However, it is rare to find evidence-based recommendations for how to interpret the performance on language tests in bilingual children, let alone which cut-offs are suitable. Peña et al. (2014) investigated whether cut-offs established for monolingual populations could accurately classify Spanish-English bilinguals with *balanced exposure* to both languages as DLD or TD on the semantics index from the BESA (Bilingual English Spanish Assessment, (Peña et al., 2016)). They found that scoring below the monolingual cut-off in *both* languages correctly classified the children as DLD, whereas taking only one language into account led to overidentification. In the current study, there was a wide variety of age of onset to Swedish, as well as the proportion of relative exposure to each language. Thus, having a z-score below –1.25 in both languages may not be a valid criterion for identifying DLD in this group of children. Notably, most children with very poor vocabulary scores had low scores in Arabic, despite the fact that they had received continuous exposure to the language from birth. For example, BiAraLI-08 had high scores in Arabic comprehension (a z-score of 0.71), but it was reported that she had attended a

bilingual Arabic-Swedish preschool, and that she avoided speaking Swedish.[95] At the same time, both the parents, the SLP and the school reported that she had severe communication difficulties. Thus, her high scores in Arabic vocabulary comprehension may be a reflection of her unbalanced exposure. Another example is BiAraLI-06, who had very poor vocabulary scores in Arabic, but scored above the mean for both comprehension and production in Swedish. BiAraLI-06 had an early age of onset for Swedish (remember that age of onset for Swedish varied immensely in both the cross-sectional and the clinical sample), and he went to a Swedish-speaking preschool. Thus, the relatively good scores in Swedish may be a result of having a larger amount of cumulative exposure to Swedish compared to many of his peers in the cross-sectional study.

The findings in the present study mirror those of Öztekin (2019, Chapter 7), who found that CLT performance varied immensely among Turkish-Swedish speaking children with a DLD diagnosis.[96] On group level, the DLD group generally had negative z-scores, and many had scores below –1.25 in at least one task. However, there were also children who had positive z-scores in one language, just like as in the present study.

To conclude, it cannot be assumed that bilingual children with DLD will score poorly on vocabulary tasks in both languages, even when making comparisons with age-matched peers speaking the same language combination as them. Language proficiency (and vocabulary knowledge in particular), is dependent on exposure for bilinguals with TD and DLD alike. Therefore, it is imperative not only to consider the vocabulary scores in both languages during assessment, but also to evaluate them in light of earlier and current language exposure.

6.3.3 Non-word repetition

Phonological working memory was assessed with four NWR tasks: the language-specific Swedish task (LS-Swe; Radeborg et al., 2006), the Swedish (QU-Swe) and the Arabic (QU-Ara) versions of the quasi-universal NWR task (Chiat, 2015) and the NWRT-Lebanese (NWRT-Leb; Abou Melhem et al., 2011).

At group level, the children in the clinical sample scored below the mean (compared to their age peers in the cross-sectional sample) on most tasks. Six children (BiAraLI-03, BiAraLI-04, BiAraLI-05, BiAraLI-07, BiAraLI-08 and BiAraLI-09) scored within the cross-sectional range for their age group on all tasks, and five children scored below the range on at least one task (BiAraLI-01, BiAraLI-02, BiAraLI-06, BiAraLI-10 and BiAraLI-11). Z-scores were

[95] Remember that BiAraLI-08's Arabic production scores were slightly below the mean, and Swedish comprehension and production scores were both –2 SD from the mean.
[96] The term used by Öztekin was 'LI'.

below –1.25 on at least two tasks for six children, however, five children had positive scores in at least one task. Thus, not all children in the clinical study performed very low. Additionally, there were several children in the cross-sectional sample who scored on par with or below the children in the clinical sample. There was considerable overlap between the groups on all tasks. One explanation for this overlap in performance between the groups may be the heterogeneity in the clinical sample. Some children (BiAraLI-01, BiAraLI-02, BiAraLI-08, BiAraLI-10 and BiAraLI-11) were reported by parents, SLPs and teachers to have severe difficulties, and they typically had very low scores in a majority of the NWR subtasks. Some other children were reported to have milder language problems (BiAraLI-04 and BiAraLI-07), and they typically scored below the age mean, but still within 1 SD from the mean. Yet other children (BiAraLI-03 and BiAraLI-05) had reported language difficulties and poor vocabulary scores but scored slightly below the mean in most NWR tasks and even had positive scores in one task. Moreover, the possible occurrence of overdiagnosis (in the clinical sample) and underdiagnosis (in the cross-sectional sample) of individual cases may add to this overlap of performance between the two groups.

Although poor NWR performance is frequently reported in the literature to be associated with DLD (Dollaghan & Campbell, 1998; Kalnak et al., 2014; Saiegh-Haddad & Ghawi-Dakwar, 2017; Topbaş et al., 2014), there is not a perfect relationship between poor repetition performance and presence of DLD (Ellis Weismer et al., 2000). Additionally, it should be emphasised that there are several reports in the literature of poorer diagnostic accuracy and a higher degree of overlap between TD and DLD groups on NWR tasks in bilingual populations compared to monolingual populations. This pattern has been attested not only for language specific NWR tasks, but also for non-language-like and quasi-universal tasks (Abed Ibrahim & Hamann, 2017; Boerma, Leseman, et al., 2017; dos Santos & Ferré, 2018; Thordardottir & Brandeker, 2013). It should also be acknowledged that DLD manifests in different ways for different children, and NWR performance is not necessarily depressed in all individuals with deficits in their functional language skills.

6.3.4 Individual children

In this section, the children in the clinical study will be characterised in terms of their performance on the vocabulary and NWR tasks. The results will be analysed in light of the background information provided by the parental questionnaires, as well as the interviews conducted with the parents, the SLPs and the teachers.

The second part of the section contains a discussion regarding some of the low performing children from the cross-sectional study, who were in the 'grey area', and were considered as potentially undiagnosed cases of DLD. The same thorough information is not available for these children as for the

children in the clinical study (i.e. there are no in-depth interviews with parents or teachers, and for obvious reasons, no information from an SLP). However, the information available concerning their language skills, language development, exposure patterns and overall background will be discussed.

6.3.4.1 The children with a DLD diagnosis

BiAraLI-03 and BiAraLI-05 had poor vocabulary scores (particularly in Arabic), but received positive z-scores in two NWR tasks and moderately low scores in two others. At the same time, both children are reported to have functional communication difficulties according to the interviews with parents, SLPs and teachers (although the mother of BiAraLI-05 does not think that her child has any difficulties).

Two children (BiAraLI-04 and BiAraLI-07) scored moderately low on both the vocabulary and NWR tasks. In comparison to the other children in the DLD group, these two children had seemingly milder problems. BiAraLI-04 recently had his diagnosis changed from general to expressive language disorder. Although the SLP reported deficits in expressive morphosyntax in both languages, the parents and the preschool staff were of the opinion that he only had problems with speech sounds. BiAraLI-07 was reported to have problems with both comprehension and production according to the parents, the SLP and the preschool teachers. However, it is reported that he plays well with other children and that there are seldom misunderstandings or conflicts.

Five children (BiAraLI-01, BiAraLI-02, BiAraLI-06, BiAraLI-08 and BiAraLI-10) had poor NWR performance in all tasks, as well as having poor vocabulary scores (although some children had good scores in one language (BiAraLI-06 in Swedish) or task (BiAraLI-08 in Arabic comprehension). All five children were described by parents, SLPs and teachers to have severe language difficulties, to the extent that there were often conflicts or other social difficulties, and they also had poor scores overall (in both NWR and vocabulary tasks).

BiAraLI-11 had a large discrepancy between performance in the vocabulary tasks and the NWR tasks. She had good vocabulary scores in Arabic, but also surprisingly good scores in Swedish considering that her age of onset for Swedish was late (age 4;0–5;0). However, NWR performance was poor, especially in the tasks with higher phonological complexity (i.e. clusters). Her comprehension seems to be better than production according to parent, SLP and teacher reports.

Finally, BiAraLI-09 scored high or very high in all NWR tasks. Meanwhile, his vocabulary scores were slightly below the mean in Arabic, and much above the mean in Swedish. The reports from parents, the SLP and the school were inconsistent, as he had received a DLD diagnosis by an SLP, but the parents were of the opinion that he did not have a language disorder, and the school staff perceived the child to be very shy. Considering the fact that BiAraLI-09 had good scores in all NWR tasks, at the same time as having very

good vocabulary scores in Swedish and just below the mean in Arabic compared to the TD group, it could be the case that he was subject to overdiagnosis of DLD. In fact, according to the SLP assessment he had only been assessed in Swedish and his language scores were compared against monolingual Swedish norms.

6.3.4.2 Low performers from the cross-sectional study

This section contains a discussion concerning some of the children who stood out in the cross sectional study, due to poor performance on more than one task, with low performance in both languages. Task performance is related to background information such as reported language development, current proficiency levels and exposure patterns. Since these children participated in the cross-sectional study, no interviews were conducted with teachers or SLPs. Thus, the background information was obtained from parental questionnaires, which form the basis for discussion. All children discussed in this section have been mentioned in earlier chapters with respect to their low scores (BiAra5-06, BiAra6-26, and BiAra6-29). All three children scored low on a majority of the tasks, but not necessarily below –1.25 on all tasks. Furthermore, they had vocabulary scores that were unexpectedly low in one or both languages when considering their exposure patterns (age of onset and current daily exposure). By contrast, for most children in the cross-sectional study, having low vocabulary scores in one language could be explained by their exposure patterns, such as having unbalanced daily exposure levels, or a late age of onset. Table 6.4 and Table 6.5 show the vocabulary and NWR z-scores of the low performing children in the cross-sectional study together with the scores from the children in the clinical study.

Table 6.4. Arabic and Swedish vocabulary z-scores of the children in the clinical study and the three low performing children from the cross-sectional study in relation to the cross-sectional means. Z-scores at –1.25 or below are marked with '*' and boldface.

		Arabic		Swedish	
	Age	**Comp**	**Prod**	**Comp**	**Prod**
BiAraLI-01	6;8	**–1.53***	**–1.25***	**–1.76***	**–1.62***
BiAraLI-02	6;1	**–1.40***	–0.72	–1.07	–0.66
BiAraLI-03	5;7	**–3.28***	**–2.40***	–1.13	–0.58
BiAraLI-04	6;0	–0.20	–0.49	0.01	–0.31
BiAraLI-05	7;3	**–2.60***	**–1.53***	–0.86	–0.40
BiAraLI-06	5;4	**–2.07***	**–1.73***	0.41	0.33
BiAraLI-07	6;1	–0.34	–0.64	–0.78	–0.49
BiAraLI-08	7;1	0.71	–0.23	**–2.03***	**–2.62***
BiAraLI-09	6;7	–0.07	–0.26	0.99	0.64
BiAraLI-10	5;0	**–3.89***	**–2.73***	–1.13	–0.69
BiAraLI-11	6;4	0.33	0.49	0.01	–0.66
BiAra5-06	5;4	**–2.98***	**–1.81***	–0.47	–0.07
BiAra6-26	6;8	**–2.33***	–0.79	**–1.96***	–1.18
BiAra6-29	6;1	–0.47	–0.26	**–1.76***	–1.18

Table 6.5. NWR z-scores of the children in the clinical study and the three low performing children from the cross-sectional study in relation to the cross-sectional means. Z-scores at –1.25 or below are marked with '*' and boldface.

	Age	**LS-Swe**	**QU-Swe**	**NWRT-Leb**	**QU-Ara**
BiAraLI-01	6;8	**–1.28***	**–2.17***	**–2.46***	**–2.14***
BiAraLI-02	6;1	**–2.30***	**–1.76***	**–3.05***	**–1.69***
BiAraLI-03	5;7	–0.64	0.37	–0.35	0.17
BiAraLI-04	6;0	–0.51	0.73	–0.67	–0.32
BiAraLI-05	7;3	–0.59	0.61	0.02	–0.39
BiAraLI-06	5;4	**–2.07***	**–4.56***	**–3.26***	**–1.79***
BiAraLI-07	6;1	0.26	–0.51	–0.37	–0.32
BiAraLI-08	7;1	**–1.76***	–1.08	**–1.92***	–0.78
BiAraLI-09	6;7	1.02	–0.10	0.82	0.60
BiAraLI-10	5;0	–0.64	**–2.59***	**–3.26***	**–2.28***
BiAraLI-11	6;4	**–2.05**	–0.51	**–3.35**	–0.77
BiAra5-06	5;4	–1.21	–1.11	–0.09	**–1.30***
BiAra6-26	6;8	**–1.28***	**–1.76***	**–2.16***	–1.23
BiAra6-29	6;1	**–1.79***	–0.93	–0.67	–1.23

In what follows, the CLT and NWR scores will be discussed for each of the three children in light of the background information reported by the parents.

BiAra5-06 (age 5;4) had very low scores in Arabic vocabulary comprehension and production, z-scores were below –1.25 for both comprehension and production. Her vocabulary scores in Swedish were not as low, but z-scores were still negative for both comprehension and production. NWR scores were substantially below average, with one score being below –1.25. According to the parents, daily exposure was unbalanced with 20% Arabic and 80% Swedish during the day. They also mention that the child attended a bilingual English-Swedish preschool, however, no estimate of the amount of daily input in English was provided. BiAra5-06 started to hear Swedish early, at age one. At home, one parent spoke mostly in Arabic, and the other parent spoke mostly Swedish. The parents reported that they sometimes engaged in book-reading activities (1–2 times a month), but only in Arabic and never in Swedish. However, the child received daily input in both Arabic and Swedish via digital media. With respect to early language development, parents report that the first word and the first multi-word utterance appeared early (in Arabic at 7 and 18 months respectively). However, they also state that they were worried about her language development when she was about two years, because "it was difficult to combine multiple words", and they consulted an SLP about pronunciation difficulties.[97] The fact that the parents of BiAra5-06 reported that she had an unbalanced input pattern, with significantly more daily exposure to Swedish, could be a possible explanation for BiAra5-06's very low scores in Arabic. However, other children in the cross-sectional study with very low scores in one language typically had high scores in the other language, which was not the case here. What is more, exposure patterns cannot explain the fact that she has such low scores on NWR. The fact that the parents had been concerned about their child's language development, and consulted an SLP due to pronunciation difficulties also indicates that this child could have an undetected language disorder.

BiAra6-26 (age 6;8) is the sibling of BiAraLI-01. He had very low vocabulary scores both in Arabic and in Swedish, with comprehension scores being below –1.25 in both languages. NWR scores were substantially below average, with all tasks but one being below –1.25. The parents report that both parents had spoken only Arabic in the home since the child's birth, and that current daily exposure to Arabic and Swedish was fairly balanced, with somewhat less Arabic (40%) than Swedish (60%). The child started being regularly exposed to Swedish relatively late, at age five. The questions concerning how frequently the family engaged in book-reading activities in

[97] Here "it was difficult to combine multiple words" (in Swedish: "vid två års ålder var det svårt att bygga flera ord") is interpreted as the child having had difficulties forming multi-word utterances.

the home were left blank in the questionnaire, which suggests that these activities occur very seldom or not at all. Considering the fact that age of onset to Swedish was relatively recent, the low Swedish vocabulary scores are not surprising. Nonetheless, BiAra6-26 also scored very low in Arabic despite the fact that he had long-term exclusive exposure to Arabic until age five. Also, having uneven exposure patterns cannot explain his low scores in the NWR tasks. The parents did not mention any concern about the language development of BiAra6-26. However, when the teacher of his sibling, BiAraLI-01, was interviewed, she also disclosed some information about BiAra6-26 since he went to the same school. According to the teaching staff, BiAra6-26 has similar but seemingly milder problems as his brother. Considering the poor test scores in combination with the fact that there is heredity for DLD, it is possible that BiAra6-26 could have DLD that has not yet been diagnosed.

BiAra6-29 (age 6;1) had very low scores in Swedish vocabulary comprehension and production, and scores below average in Arabic vocabulary comprehension and production. NWR scores were considerably below the average in all tasks, with one task being below –1.25. According to the parental questionnaire, daily exposure was balanced (50/50) between Arabic and Swedish. The child started receiving regular exposure to Swedish relatively late, at age four. At home, both parents had spoken mostly Arabic since the child's birth. The parents reported that book-reading activities were frequent in the home, carried out every day in both Arabic and Swedish. BiAra6-29 also received input in both Arabic and Swedish via digital media 1–2 times a week. The first word and the first multi-word utterance appeared within the expected time frame (in Arabic at 12 and 18 months respectively). The parents do not voice any concern about the language development, nor do they report heredity for language difficulties. Since BiAra6-29 had not been exposed to Swedish for very long, it is perhaps not surprising that his vocabulary scores in Swedish were not very high. However, just as was the case for BiAra6-26, the vocabulary scores in Arabic were not as high as one would expect, considering that BiAra6-29 had long-term exposure to Arabic prior to starting being exposed to Swedish regularly at age four. Additionally, just as for BiAra5-06 and BiAra6-26, exposure patterns cannot serve as an explanation for the very poor NWR scores. In this case, it is more difficult than for the other two children to speculate whether the poor test scores may reflect an undetected language disorder. In contrast to BiAra5-05 and BiAra6-26, there is no information in the questionnaire (e.g. late language development, heredity or parental concern) that give such indications. This case highlights the importance of collecting additional information from parents, and preferably also from teachers when assessing bilingual children with suspected DLD. Such reports provide information that makes it possible for the clinician to *interpret* the test scores in light of relevant background

information, such as exposure factors (age of onset, input quantity and quality), family history and early language development.

6.3.5 Concluding remarks

In the current study, children who received a DLD diagnosis were compared to children with seemingly typical language development (i.e. who did not have a DLD diagnosis). The children in the clinical group showed very diverse profiles with respect to their performance on the vocabulary and NWR tasks and the information provided by the parents, teachers and SLPs about their background and current language skills. A substantial part of the children in the clinical sample were performing consistently low on (almost) all tasks, and the informants (parents, SLPs and teachers) reported severe functional communication problems, which indicates that the language difficulties are indeed a result of DLD. For other children, the picture was more complex, either because their performance was borderline poor (i.e. in 'the grey area') in all tasks, or because their performance was uneven, e.g. with relatively good vocabulary scores and poor NWR scores. In the following, three possible explanations will be discussed.

First, individuals with DLD may have relative strengths or weaknesses in one modality (comprehension or production) or one or more linguistic domains (phonology, morphosyntax, vocabulary, discourse or pragmatics) (Norbury et al., 2016; Tomblin et al., 1996). The skills assessed in the present study are limited to comprehension and production of isolated words, as well as phonological working memory as measured by NWR tasks.

Second, it should be emphasised that since language abilities exist on a spectrum, there is not a perfect division in performance between TD and DLD groups on any task targeting language (or language processing) abilities.
In this study, a z-score of –1.25 was utilised as a potentially clinically informative cut-off (i.e. identifying the lowest-scoring 10% in the group). As Norbury et al. (2016) point out, cut-offs are arbitrary in the sense that they do not say anything about how a certain score correspond to functional communicative abilities. Thus, it does not necessarily mean that receiving a z-score below –1.25 on a given task is also associated with poor *functional* language skills. Likewise, it does not mean that all individuals who score above (and in some cases well above) –1.25 on a certain task do not experience depressed functional language skills. Thus, no matter which cut-off score is used for a given task, sensitivity and specificity cannot be expected to reach 100%.

Third, both children with TD and children with DLD may score relatively high or low on vocabulary tasks compared to their age-matched peers due to differences in quantity (length of exposure, or relative amount) or quality of input.

In conclusion, scoring relatively well in one or more tasks may not necessarily mean that the child is not experiencing functional language difficulties. Likewise, poor performance is not necessarily associated with impaired functional communicative abilities. Thus, it is imperative to stress once more that the performance on language tasks should be considered in relation to language exposure as well as functional language skills and developmental history.

7 Summary and general discussion

The first part of this chapter provides a summary of the results (section 7.1) of the cross-sectional studies on vocabulary (section 7.1.1) and non-word repetition (section 7.1.2), and of the clinical study (section 7.1.3). The second part of the chapter (section 7.2) contains a general discussion of these results, and discusses limitations of the study and suggestions for further research.

Three main research questions were asked in this thesis:

- **RQ1**: How do vocabulary skills develop with age in both languages of 4–7-year-old Arabic-Swedish-speaking bilinguals, and which external factors influence that development?
- **RQ2**: How do 4–7-year-old Arabic-Swedish-speaking bilinguals perform on NWR tasks, and how is their performance affected by item length and complexity, language-likeness, and language exposure and vocabulary?
- **RQ3**: How do bilingual children with a DLD diagnosis perform on vocabulary and NWR tasks, what are the reported backgrounds, language abilities and communicative behaviours of bilingual children with a DLD diagnosis, and how can these be used to identify DLD in bilinguals?

7.1 Summary of the results

7.1.1 Vocabulary skills of the children in the cross-sectional study

In Chapter 4, the vocabulary (CLT) comprehension and production skills of 99 children were investigated in the minority language (Arabic) and the majority language (Swedish). The following specific research questions were asked:

- Is there a difference between vocabulary scores in the two languages?
- How do vocabulary scores develop with age in Arabic and Swedish?
- What is the relationship between production scores and comprehension scores in each language?
- What is the relationship between length of exposure to Swedish and Swedish vocabulary comprehension and production?

- What is the relationship between amount of daily exposure and vocabulary comprehension and production in Arabic and Swedish?
- What is the relationship between SES and vocabulary comprehension and production in Arabic and Swedish?

Summary of the results:

- As expected, comprehension scores were higher than production scores in both languages.
- There was no difference between Arabic and Swedish CLT total scores for all age groups combined.
- Several children reached the maximum score in Swedish comprehension, but no child did so for Arabic comprehension.
- There were positive correlations between comprehension and production scores in Arabic and between comprehension and production scores in Swedish, but the difference was slightly larger in Swedish.
- There was a wider range in Swedish comprehension scores (with the top score being higher and the lowest score being lower) than in Arabic comprehension scores.
- For production, the top score was the same in both languages, but the lowest score was much lower in Arabic compared to Swedish due to one low-performing child (scoring only 1 p).
- Comprehension and production scores increased with age in both Arabic and Swedish, as there were positive correlations between age in months and scores. For all vocabulary tasks, there was a significant difference between the 4-year-olds and the 7-year-olds.
- Length of exposure to Swedish was a significant predictor of vocabulary scores in Swedish. For comprehension, length of exposure to Swedish had similar predictive strength as chronological age, but for production, it was a stronger predictor than age or percent daily exposure to Swedish.
- There were positive correlations between percent daily exposure to Swedish and Swedish comprehension and production scores. In the multivariate regression models, it was only a significant predictor for production scores.
- There were positive correlations between percent daily exposure to Arabic and Arabic comprehension and production scores. In the multivariate regression models, age was a stronger predictor for comprehension scores than percent daily exposure. For production, age and percent daily exposure had similar predictive strength.
- There was no effect of SES (level of parental education) on children's vocabulary scores in either language.

7.1.2 NWR performance of the children in the cross-sectional study

In Chapter 5, phonological working memory was investigated in 99 Arabic-Swedish-speaking 99 children with four different NWR tasks (LS-Swe, QU-Swe, QU-Ara and NWRT-Leb), with non-word items differing with respect to item length (number of syllables), syllabic complexity (presence or absence of consonant clusters) and whether the items were language-like (i.e. adhering to the phonological rules of Swedish) or non-language-like. The following specific research questions were asked:

- How does performance on the four NWR tasks develop from age 4 to 7?
- Are there differences in performance between tasks?
- What characterises the children performing low on NWR tasks?
- What are the effects of task (LS-Swe vs. QU-Swe), item length (number of syllables), presence of consonant clusters, exposure to Swedish, SES and Swedish vocabulary on the performance on the LS-Swe and the QU-Swe tasks?
- What are the effects of type of task (QU-Swe vs. QU-Ara) and item length (number of syllables) on the performance on the QU-Swe and the QU-Ara tasks?

Summary of the results:

- Scores increased with age on all NWR tasks, as there were positive correlations with age in months, as well as differences in mean scores between the age groups for all tasks. There was a significant difference between the 4-year-olds and the 7-year-olds in all tasks.
- The tasks had different levels of difficulty, which manifest in two ways. First, there was a significant difference in overall performance between all tasks. Second, there was a substantial difference in the proportion of children who scored high or low in each task. This difference was particularly pronounced for the LS-Swe task and the NWRT-Leb. Only one child scored at 90% or higher on the LS-Swe task, whereas 41% of the children did so on the NWRT-Leb.
- Some of the children with poor NWR performance were reported to have certain risk factors associated with DLD (a late language onset, having heredity for language problems, parental concern, etc.), and/or low vocabulary scores, but far from all low-performing children met these criteria.
- Overall performance was lower on the LS-Swe items than the QU-Swe items, and longer items (with more syllables). Swedish vocabulary was a significant predictor of performance on both the (language-like) LS-Swe items and the (non-language-like) QU-Swe items, but children with higher

Swedish vocabulary scores did not show a greater advantage on the LS-Swe items compared to the QU-Swe items. Length of exposure to Swedish was not a significant predictor of performance on the LS-Swe items nor the QU-Swe items.

- Overall performance was lower for LS-Swe items that were longer (with more syllables), and contained consonant clusters. Having a higher Swedish vocabulary score was associated with an overall better performance on the LS-Swe task, but not more so on longer items or items with clusters.
- Overall performance was slightly better on the QU-Swe task than the QU-Ara task at all syllable lengths except for at five syllables. To some extent, item effects could explain the difference in accuracy between the two tasks, where performance was higher for the QU-Swe items compared to the QU-Ara items.

7.1.3 The clinical study

In chapter 6, 11 children with a DLD diagnosis were compared to the 99 children in the cross-sectional sample with respect to their performance on vocabulary comprehension and production (CLT) in the minority language (Arabic) and the majority language (Swedish) and phonological working memory as assessed with four NWR tasks (LS-Swe, QU-Swe, QU-Ara and NWRT-Leb). Furthermore, the children in the clinical sample were characterised in terms of their early language development, language skills, communicative behaviour and social characteristics, as described by their parents, teachers and SLPs. The following specific research questions were asked:

- What are the reported early language development, language skills, communicative behaviours and social characteristics of the Arabic-Swedish-speaking children with a DLD diagnosis, as described by parents, SLPs and teachers?
- What are the vocabulary skills in Arabic and Swedish of the children with a DLD diagnosis?
- How do the children with a DLD diagnosis perform on non-word repetition tasks?
- Are there similarities or differences between the children with a DLD diagnosis and the children in the cross-sectional study concerning vocabulary or non-word repetition performance?

Summary of the results:

- According to parental report, a higher proportion of the children in the DLD group had a late onset of the first words and/or multi-word

utterances and/or a close relative with language or literacy difficulties compared to the TD group.

- The DLD children were often described as having difficulties with language comprehension (such as having difficulties understanding instructions) and language production (such as having deficiencies in expressive morphosyntax), as well as having poor vocabulary skills.
- Some children were described as having severe communication difficulties with frequent misunderstandings and peer conflicts, which had a negative impact on their social relations.
- At group level, the DLD children generally scored within the range but below the mean on both vocabulary (CLT) and NWR tasks compared to their age-matched peers in the cross-sectional study.
- Many children in the DLD group had particularly poor vocabulary z-scores in their first language (Arabic), despite extensive and continuous input from birth.
- While a majority of the children in the DLD group had a vocabulary z-score below –1.25 in one language, only one child scored below –1.25 in *both* languages.
- There was a high degree of overlap in performance between the TD and the DLD group on both the vocabulary and the NWR tasks.
- There was not one NWR task that stood out as being much better than the others for identifying the children in the DLD group.

7.2 General discussion and research contributions

In this section, the results are discussed with respect to previous research, and how they contribute to our understanding of bilingual vocabulary development (section 7.2.1), NWR performance (section 7.2.2), and developmental language disorder in the context of bilingualism (section 7.2.3). Furthermore, limitations of the study and suggestions for future research are discussed in section 7.2.4.

7.2.1 Vocabulary: age effects, input and SES

The present study investigated the relationship between chronological age, estimated daily exposure, SES and vocabulary scores in the minority language Arabic and the majority language Swedish. Additionally, the relationship between length of exposure and vocabulary was investigated in conjunction with the other variables for Swedish.

In the present study, scores increased as a function of age in months for comprehension and production in Arabic as well as in Swedish. The overall results support previous research finding an association between chronological age and performance on the CLTs for monolinguals (Haman et

al., 2017), for monolingual Swedish-speaking 4–6-year-olds (Lindgren, 2018, Chapter 5) and bilingual Turkish-Swedish-speaking bilinguals in their majority language Swedish (Öztekin, 2019, Chapter 4). The results also corroborate earlier findings reported in the literature, namely that there is often a clear development with age for vocabulary in the majority language (in this case, Swedish) (Bialystok et al., 2010; Cobo-Lewis et al., 2002a; Dijkstra et al., 2016; Prevoo et al., 2014). However, there was also a clear development with age in the minority language (in this case, Arabic), a finding that differs from many previous studies that report small or no age effects on vocabulary scores in the minority language (Ganuza & Hedman, 2019; V. C. M. Gathercole & Thomas, 2009; Hoff et al., 2014; Leseman, 2000; Öztekin, 2019, Chapter 4). As children grow older, exposure to the majority language is expected to increase proportionally at the expense of the minority language, in part as an effect of schooling in the majority language (Montanari et al., 2018; Pearson, 2007). As suggested in section 4.4.1, there may be certain characteristics of the Swedish-Arabic-speaking population related to migration patterns and social factors which may boost the children's access to input sources in the majority language. Additionally, many of the children in the sample (N=45) were not born in Sweden (even though they arrived at a young age). Thus, they had likely received extensive and continuous input in the minority language from birth, which may explain the clear age effect in the minority language.

In the current study, vocabulary scores increased as a function of the proportion of daily exposure. The effect was seen in both languages, and it was stronger for comprehension than production. These findings are in line with earlier studies, demonstrating a relationship between percent daily exposure and vocabulary comprehension in the minority language (Prevoo et al., 2014) and the majority language (Unsworth, 2016), as well as for vocabulary production in the majority language (Öztekin, 2019, Chapter 4; Prevoo et al., 2014). They are also in line with Thordardottir's (2011) observation that the effect of relative amount of exposure is stronger for vocabulary production than comprehension.

For the majority language Swedish, length of exposure emerged as the most important predictor of vocabulary scores. For comprehension, it was as strong as chronological age, but for production, it was a stronger predictor than age or percent daily exposure. The amount of cumulative exposure to a language was identified by Thordardottir (2019) as being a more important predictor of vocabulary comprehension scores than age of onset. Early age of onset (and thus, longer exposure) to the second language does not necessarily mean that a child will always have more cumulative exposure to the language. It is however likely that children with longer exposure to Swedish received higher vocabulary scores in Swedish because they had a higher amount of cumulative exposure to the language.

The findings for vocabulary scores with respect to SES in the current study do not match those of many earlier studies, where higher SES is generally associated with better vocabulary scores in the majority language (Buac et al., 2014; Calvo & Bialystok, 2014; Cobo-Lewis et al., 2002a; Prevoo et al., 2014). One reason for the differing result could be methodological. As put forward by Letts, Edwards, Sinka, Schaefer and Gibbons (2013) and Öztekin (2019, Chapter 8), SES is a multidimensional construct that is operationalised in different ways by different researchers, obstructing comparisons between different settings and studies. In the current study, the mean level of education of both parents was used as a proxy for SES. Other ways of operationalising SES is primary caregiver's level of education, maternal or (less commonly), paternal education, parental income or occupation, or characteristics of the community where the child lives (Buac et al., 2014; Calvo & Bialystok, 2014; Engel de Abreu, 2011; Letts et al., 2013).

How might SES be related to vocabulary development? While SES cannot be assumed to be directly influencing language or vocabulary skills, SES can be related to language input in different ways, thus acting as a mediator between SES and language skills. Language input may be affected by influences both in the family and outside the home. Higher SES may co-vary with a higher degree of use of the majority language in the home (Prevoo et al., 2014), or better majority language proficiency of the parents (Buac et al., 2014), which may in turn boost the children's majority language skills if the majority language is spoken in the home. In the present study, almost 80% of the participating families report that both parents spoke to their child only or mainly in the minority language) Arabic (recall Table 3.7 in section 3.1.1.5). Overall, there was very little parental input in Swedish. Thus, parental input in the majority language is not likely to be a relevant explanatory variable influencing the children's majority language skills in this population. SES may also be related to the circumstances in the family. In some countries, economical constraints may affect access to education and opportunities to engage in extracurricular activities. As already mentioned in section 4.4.3, the Swedish setting, which offers affordable childcare and free education to all children no matter their socioeconomic background, may act as a compensating factor for differences in family conditions, providing an environment where all children receive access to high-quality education and input in the majority language. Nevertheless, it should also be acknowledged that Swedish preschools and schools have different conditions that in turn are related to differences in the demographic constitution of different residential areas. As mentioned in section 4.4.3, Andersson et al. (2019) found that 7–8-year-old pupils from low-SES households, speaking Swedish as a second language, as well as attending a school with a high proportion of pupils sharing the same background, ran an increased risk of scoring low on a standardised language test in Swedish. Rydland, Grøver and Lawrence (2014) examined the vocabulary trajectories from age 5 to 10 in the second language

(Norwegian) of 26 Turkish-speaking children in Norway. The Norwegian context is similar to the Swedish one, where childcare is inexpensive and readily available to all children from a young age, regardless of their socioeconomic background. In addition to maternal level of education, Rydland et al. found that (amount and diversity of) teacher-led talk in preschool, peer talk in preschool as well as the concentration of Turkish-speakers in the neighbourhood had an impact on vocabulary comprehension growth curves in the second language of the children. Thus, investigating the nature of input provided by teachers and peers in preschool as well as exploring the effect of neighbourhood demographics could tell us more about L2 development in the Swedish context.

7.2.2 Phonological working memory, NWR and task effects

The present study investigated development with age and task effects for the four NWR tasks (LS-Swe, QU-Swe, QU-Ara and NWRT-Leb). The characteristics of children performing low were outlined. Furthermore, the performance on the (language-like) LS-Swe task was investigated with respect to previous language experience (length of exposure to Swedish), vocabulary (Swedish CLT comprehension), and non-word length (number of syllables) and compared to the (non-language-like) QU-Swe task. For the LS-Swe task, these aspects were also investigated in relation to presence or absence of consonant clusters in the items. The performance on the two quasi-universal tasks (QU-Swe and QU-Ara) was also compared.

There were differences in the overall proportional accuracy between tasks, with the most pronounced difference being between the overall easier NWRT-Leb and the overall more difficult LS-Swe task. While there was a pronounced ceiling effect for the NWRT-Leb (i.e. 41% of the children scored at 90% or higher), only one child scored at 90% or above on the LS-Swe task. Furthermore, the children also performed somewhat better on the QU-Swe task than the QU-Ara task, but this was to some extent driven by item effects. Compared to a Dutch version of the quasi-universal NWR task (Boerma et al., 2015), performance on the QU-Swe and the QU-Ara was much higher (33% and 39% respectively). As discussed in Chapter 5 (section 5.5.1) there may be several reasons for this, e.g. differing assessment, transcription or scoring procedures. These findings support Chiat's (2015, p. 143) suggestion that the QU tasks are likely not completely comparable across language versions. For all tasks, there were significant age effects (although development with age was slightly weaker in the QU-Swe task compared to the other three tasks). These findings are in line with many previous studies, reporting better NWR performance with age (Chiat & Roy, 2007; Kalnak et al., 2014; Radeborg et al., 2006; Topbaş et al., 2014).

NWR performance was explored for the LS-Swe task and the QU-Swe task (as a comparison between language-like and non-language-like items) with

respect to language experience (SES, percent daily exposure to Swedish, length of exposure to Swedish and Swedish vocabulary comprehension). There was no correlation between any of the tasks and SES, mirroring the findings of Chiat and Roy (2007), Engel et al. (2008), and Kalnak et al. (2014). Percent daily exposure to Swedish was not correlated to performance on the Swedish tasks. Furthermore, length of exposure to Swedish was not a significant predictor of performance on the LS-Swe task (nor the QU-Swe task) when chronological age was controlled for. These findings differed from earlier studies. Thordardottir and Brandeker (2013) reported that cumulative amount of exposure to English correlated with performance on the Children's Non-word Repetition test (CNRep; S. E. Gathercole et al., 1994), a task with items that are particularly wordlike, even containing real morphemes in English, a feature known to enhance repetition accuracy (Dollaghan et al., 1993). Similarly, Sorenson Duncan and Paradis (2016) found that length of exposure to English was a significant predictor of the performance on a language-like NWR task (based on the phonotactics of English) for children learning English as a second language. Moreover, English vocabulary scores was a significant predictor of performance on the language-like NWR task. In the current study too, Swedish vocabulary comprehension was a significant predictor of performance on the LS-Swe and the QU-Swe task. This was congruent with several previous studies finding an association between NWR performance and vocabulary scores (Bowey, 1996; Coady & Evans, 2008; Dollaghan et al., 1995; S. E. Gathercole & Adams, 1993, 1994; S. E. Gathercole & Baddeley, 1989). In contrast to what was expected though, the effect of Swedish vocabulary size on repetition accuracy was not stronger for the (language-like) LS-Swe items than for the (non-language-like) QU-Swe items. Szewczyk et al. (2018) suggested that vocabulary size may serve as a proxy for the range and depth of sublexical representations that are stored in long-term memory. In the framework of the multicomponent model of working memory (recall section 5.1.1 and Figure 5.1), having a larger vocabulary also means that the range of sublexical representations (i.e. patterns of phoneme combinations occurring at different frequencies in a given language) stored in long-term memory are larger and more salient. This in turn facilitates the process of perceiving and processing a non-word when accessed through the phonological loop. Although the items in the LS-Swe task were constructed to adhere to the phonotactics of Swedish (e.g. with consonant clusters present in both onset and coda), they were not controlled for phonotactic probability. Additionally, even though the QU-Swe items (which had only simple CV syllable structures, with no clusters or coda) were not specifically modelled to adhere to the phonological rules of Swedish, it should be recognised that CV syllables are common in Swedish lexical phonology. Therefore, the phonological structures of the items in the QU-Swe task were by no means illegal according to Swedish lexical phonology. Consequently, it is not known to which extent the items in the two tasks

actually differ with respect to their phonotactic probability in Swedish. This may provide an explanation for why Swedish vocabulary size was not a stronger predictor of accuracy for the LS-Swe items compared to the QU-Swe items.

NWR performance was further explored with respect to the characteristics of the items in the LS-Swe task and the QU tasks: non-word length and syllabic complexity (presence or absence of consonant clusters). For all tasks, repetition accuracy generally decreased as a function of number of syllables, mirroring several previous studies finding a length effect (Boerma et al., 2015; Chiat & Roy, 2007; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Radeborg et al., 2006; Thordardottir & Brandeker, 2013; Topbaş et al., 2014). In the LS-Swe task, items containing clusters were more difficult to repeat than those without clusters, mirroring previous findings (Abed Ibrahim & Hamann, 2017; dos Santos & Ferré, 2018). In the LS-Swe task, the presence of consonant clusters seemed to contribute more to item difficulty than number of syllables. Although the presence of consonant clusters contributed more to item difficulty in the LS-Swe task, overall accuracy rates were by far the highest for the NWRT-Leb (the only other task containing consonant clusters). Thus, mere presence of consonant clusters cannot be assumed to contribute more to item difficulty than number of syllables across tasks. The LS-Swe task had items with a wide phoneme inventory that was Swedish-like, whilst the NWRT-Leb had a restricted phoneme inventory that was not specific for a particular language. Furthermore, the NWRT-Leb items were shorter with respect to number of syllables (1–3) compared to the LS-Swe task (2–5). Thus, there may be an interplay between phoneme inventory, cluster types and item length (number of syllables) affecting item and overall task difficulty. In order to investigate the effect of syllabic complexity, future studies could make use of a NWR task with items that are carefully controlled for the number of clusters in onset and coda position at every syllable length.

Even though poor NWR performance is frequently reported to be associated with DLD (Dollaghan & Campbell, 1998; Kalnak et al., 2014; Saiegh-Haddad & Ghawi-Dakwar, 2017; Topbaş et al., 2017), several of the low-performing children in the cross-sectional study were not reported to have been a late talker or having heredity for language or literacy difficulties. Moreover, as demonstrated in Chapter 6 (recall section 6.2.4 and Figure 6.2), there was substantial overlap in performance on all NWR tasks between the TD group and the DLD group. Because of this, NWR should be used with caution as a diagnostic tool for identifying DLD.

In sum, these findings underscore the importance of interpreting the performance on NWR tasks in combination with performance on other language measures, as well as reports of early language development, heredity and hearing problems. Length of exposure to Swedish and current amount of exposure to Swedish did not affect performance on the LS-Swe nor the QU-Swe task, and children with a lower Swedish vocabulary score were equally

disadvantaged on the LS-Swe task and the QU-Swe task compared to children with higher Swedish vocabulary scores. Consequently, there was no support in this study to conclude that language-specific NWR tasks are necessarily less suitable for bilinguals than quasi-universal tasks (as suggested by e.g. Boerma et al. (2015)), if their performance is compared to other children who share the same language background (i.e. bilinguals who speak the same language combination).

7.2.3 Developmental Language Disorder in the context of bilingualism

In order to explore how DLD manifests in Arabic-Swedish-speaking bilinguals, 11 children with a DLD diagnosis were assessed with respect to their vocabulary skills and non-word repetition performance, and they were compared to age matched peers on these measures. In addition, detailed information was gathered about their early language development, language proficiency, communicative behaviours and social characteristics as described by parents, SLPs and teachers.

Since there is a lack of research on the language development of typically developing Arabic-Swedish-speaking bilinguals, the major focus of the thesis was on investigating vocabulary skills and NWR performance in children with typical language development. This setup was necessary in order to establish reference data for different age groups concerning the performance on the vocabulary and NWR tasks, and it made it possible to explore how factors such as age, language input and SES were related to performance on the studied tasks.

The children with a DLD diagnosis were compared to the TD children in several ways. There was a higher proportion of children in the clinical study who had a reported late language development and/or heredity of language or literacy problems. These findings are in accordance with earlier research showing that delayed language development and heredity for language and/or literacy difficulties is proportionally more common in children with DLD compared to their typically developing peers (Kalnak et al., 2012; Paradis et al., 2010; Restrepo, 1998; Trauner et al., 2000). Parents, SLPs and teachers were asked to characterise the children's language skills in terms of comprehension and production. A majority of the children in the clinical study were characterised as having deficits in their functional language skills. Descriptions of weak language comprehension were common (e.g. having difficulties understanding instructions), which is also frequently reported in the literature about children with DLD (Bishop et al., 2017; Friedmann & Novogrodsky, 2004; Marshall et al., 2007; Norbury et al., 2016; Skarakis-Doyle et al., 2008; Tomblin et al., 1997). Additionally, there were several reports about poor expressive abilities, for instance having deficits in

expressive morphosyntax, which is also a common feature among children with DLD (Conti-Ramsden et al., 2001; Hansson & Leonard, 2006; Hansson & Nettelbladt, 1990; Hansson et al., 2003; Lely & Battell, 2003; Leonard et al., 1997).

At group level, the children in the clinical study scored below the mean but within the range for their age group on vocabulary comprehension and production in Arabic and Swedish, as well as on most NWR tasks. There was however large individual variation in task performance. Possible explanations for this individual variation have been discussed in sections 6.3.4 and 6.3.5, for instance differences in exposure patterns (for vocabulary scores), individual differences in the language profiles (strengths and weaknesses in certain language domains or modalities) of the children in the clinical study (Norbury et al., 2016; Tomblin et al., 1996), or misdiagnosis, i.e. over-diagnosis in the clinical sample and underdiagnosis in the cross-sectional sample; Grimm & Schulz, 2014.

Although it is argued in the literature that language difficulties should manifest in both languages in order for a child to qualify for a DLD diagnosis (Kohnert, 2010; Salameh, Nettelbladt, & Gullberg, 2002), it is rare to find evidence-based recommendations on how to interpret the performance of bilingual children on standardised language tests, let alone which cut-offs are suitable. Peña, Bedore and Kester (2016) found that scoring below –1 SD in *both languages* compared to monolingual norms correctly identified Spanish-English-speaking *balanced bilinguals* as TD or DLD. Thordardottir (2015) proposes that adjusting the cut-off scores for monolingual norms with respect to the proportion of relative exposure for each language may be suitable for *simultaneous bilinguals*. In the present study, the raw scores of the children in the clinical study were converted into z-scores in order to compare their performance to their age-matched peers in the cross-sectional study (i.e. not monolingual norms). A cut-off of –1.25 was used as a potentially clinically informative threshold (i.e. identifying the lowest-scoring 10% of the children). In the current study, only one child in the clinical sample had a z-score below –1.25 on vocabulary tasks in *both languages*. Since there was large variation in both the cross-sectional group and the clinical sample with respect to the age of onset of bilingualism and estimated daily exposure to each language, differences in input patterns likely had a major impact on the overlap in vocabulary scores between the TD and the DLD group. Worthy of note though was that many of the children in the clinical study scored very low in their first language Arabic (with z-scores sometimes well below –1.25), *despite* the fact that they had received extensive and continuous exposure in the home language from birth. Öztekin (2019, Chapter 7), who investigated comprehension and production of vocabulary skills and narrative abilities in Turkish-Swedish-speaking bilinguals aged 4–7, also found that scoring low in one of the languages despite ample input in that language could be an indication of DLD. Thus, the findings presented in the clinical study indicate

that interpreting the performance on standardised language measures in relation to previous and current input patterns may be informative when assessing bilinguals with suspected DLD. Future studies aiming at collecting reference data for different language measures for bilinguals should therefore collect detailed information about the history of language exposure in order to be useful for clinical assessment.

Although the children in the clinical sample generally scored below the cross-sectional mean on the NWR tasks, there was considerable overlap in performance between the groups on all tasks. Some possible reasons for this overlap have been discussed in Chapter 6 (sections 6.3.3 and 6.3.5), namely heterogeneity in the clinical sample (i.e. having different language profiles), or misdiagnosis in both the clinical sample and the cross-sectional sample. These findings are in line with literature reporting a higher degree of overlap of NWR performance in bilingual populations compared to monolingual populations (Abed Ibrahim & Hamann, 2017; Boerma et al., 2015; dos Santos & Ferré, 2018; Thordardottir & Brandeker, 2013). Whether this higher degree of overlap actually reflects poorer diagnostic accuracy for NWR tasks in bilingual populations or whether it is a reflection of a higher risk of misdiagnosis in bilingual samples remains to be established. Importantly, no NWR task stood out as being better than any of the other tasks for identifying the children with a DLD diagnosis, as there was similar overlap in performance between the TD and the DLD group on all tasks, and most children in the DLD group had similar performance in all tasks.

7.2.4 Limitations and future studies

The current study has a number of limitations that will be discussed in the first part of this section. The second part outlines avenues for future research.

One limitation of the present study is that there were different Arabic varieties that were represented in the samples of both the cross-sectional and the clinical studies (in order of proportion in the cross-sectional sample: Syrian, Palestinian, Iraqi, Lebanese and Egyptian). The sample was intended to reflect the Arabic-speaking population in Sweden of today, but at the same time, it opened up for some potential shortcomings related to the data collection procedure and comparability. Assessment protocols and procedures were adapted to the different varieties, and the variety spoken by the child was matched to that of the experimenter for the Arabic session, in order not to disadvantage any child. During the (CLT) vocabulary comprehension task, children were prompted again (using a synonym) if they seemed not to understand the first prompt immediately. It cannot be ruled out that this procedure may have boosted the scores for some individual children in the vocabulary comprehension task. The fact that the children spoke different varieties may also be problematic for the comparability between children. For instance, certain words may be less frequent, or acquired later in some

varieties compared to others. The group sizes differed, as did the proportion of speakers of the different varieties across age groups, which did not allow a statistical analysis. Impressionistically though, no dialect group stood out as being particularly disadvantaged on the vocabulary comprehension task.

The second limitation of the study has to do with the sampling of the cross-sectional and the clinical studies. The sheer amount of children in the cross-sectional study (N=99) did not allow for making more thorough investigations with respect to their developmental and family history, for instance through detailed parental interviews. Thus, it is possible that some children in the cross-sectional study could have undetected language disorders. Moreover, although all children in the clinical study had undergone language assessment and received a DLD diagnosis by a licensed SLP, there could also be cases of misclassification in the clinical sample due to an increased risk of overidentification of DLD in bilinguals.

One final limitation that will be discussed here has to do with the assessment procedure and the location of testing. The intent was to create a monolingual setting in a quiet and calm environment during data collection. Ideally, all children should be assessed in each language in a setting where it was natural for the child to speak that language (e.g. Arabic in the home, and Swedish at (pre)school). However, since this setup was not always convenient for the parents or the (pre)school, some children were assessed in Swedish at home and in Arabic at (pre)school. Additionally, the environments where the children were assessed (i.e. at (pre)school, in the home, or at a cultural or religious centre) were not always as calm or quiet as we preferred them to be. In some cases, it may have distracted the children and affected the quality of the data.

Findings from the current study, as well as the extensive amount of data that has been collected in conjunction with the overall project, opens up avenues for future research. In the cross-sectional study, age effects were seen with increases in performance with age for both vocabulary and NWR scores. However, there was also great individual variation, particularly in vocabulary comprehension and production. In order to better understand development with age, and the factors that influence the development of vocabulary skills and NWR performance over time, the tendencies observed in the cross-sectional studies in this thesis should be further investigated with longitudinal studies of individual children. In the BiLI-TAS-project, longitudinal data from a sub-group of the youngest children in this study is currently being collected.

Vocabulary and NWR encompass a very narrow part of children's language skills and language processing abilities. Also, the linguistic manifestations of DLD differ between individuals as well as between languages. Therefore, future studies investigating other types of language processing skills (e.g. sentence repetition) or other language domains (e.g. comprehension and production of morphosyntax, discourse and narratives) could bring valuable additions to the existing body of research concerning the language skills of

bilingual Swedish-speaking children with and without DLD in general, and Arabic-Swedish-speaking children in particular. As mentioned in the methods chapter, the test battery also included a narrative task, the MAIN (Gagarina et al., 2012), targeting comprehension and production of narrative macrostructure. In addition to investigations of macrostructure, the narrative production data also allows for exploring narrative microstructure, for instance morphosyntactic complexity, referentiality, cohesion and lexical diversity in both Arabic and Swedish of the children participating in the current study. The MAIN data will be explored in a forthcoming PhD thesis by Rima Haddad (in preparation). The extensive background information provided by the parents in the questionnaires could also be informative in terms of how different qualitative aspects of language input (that were not analysed in the current study) are related to language proficiency in both languages. Examples of such qualitative aspects of language input are literacy activities (such as storytelling and book reading), media consumption, and extracurricular activities.

Comparing the results of the Arabic-Swedish-speaking bilinguals in this study to bilinguals speaking other language combinations, such as Turkish-Swedish-speaking bilinguals (Öztekin, 2019) in the BiLI-TAS-project as well as German-Swedish-speaking bilinguals and Swedish-speaking monolinguals (Lindgren, 2018) may be further illuminating. For instance, comparative studies may increase our understanding of how language proficiency is affected in different domains (e.g. vocabulary and narratives) by linguistic aspects such as typological differences and the presence of cognates, as well as by environmental aspects such as input patterns and SES. Two such comparative studies, with data included in the current study, are in the process of being published. Bohnacker et al. (2020/in press) investigated age development in vocabulary comprehension and production in the majority language Swedish of monolingual Swedish-speaking and bilingual Arabic-Swedish-speaking and Turkish-Swedish-speaking children aged 4–6. Bohnacker et al. (2021/in press) investigated the receptive and expressive vocabulary skills in the minority language (Arabic or Turkish) and the majority language (Swedish) of 4–7-year-old bilinguals, and explored how vocabulary is affected by age, age of onset, SES, and minority language exposure via parents and mother tongue instruction.

Finally, there is still much research needed in order to better understand how DLD manifests in different language domains of bilinguals speaking different language combinations. In particular, future research should focus on investigating how to interpret the performance of bilinguals on standardised language tests, and which cut-offs may be suitable depending on exposure history. Furthermore, future studies should look deeper into how performance on language tasks are related to functional language skills, and to what extent reports and observations concerning functional language skills may be useful in assessment of suspected DLD in bilingual populations.

8 Sammanfattning på svenska

Många barn som växer upp i Sverige idag är flerspråkiga, och en stor andel av dem har arabiska som modersmål. Samtidigt råder det osäkerhet bland personal inom skolväsende och hälso- och sjukvård kring hur normal flerspråkighetsutveckling ser ut, och hur man ska bedöma språkfärdigheter hos flerspråkiga barn på ett tillförlitligt sätt. Flerspråkiga barn med misstänkt språkstörning rapporteras t ex bli remitterade senare för språklig bedömning hos logoped än enspråkiga barn. Flerspråkighet har också omnämnts som en försvårande omständighet vid logopedisk utredning, som gör det svårt att bedöma om ett barn har språkstörning. En orsak till förvirringen kring vad som kännetecknar förväntad respektive icke-förväntad språkutveckling hos flerspråkiga är att det finns en viss överlappning i de språkliga drag som är utmärkande för typisk andraspråksinlärning och sådana drag som är förknippade med språkstörning hos enspråkiga. En annan orsak är att det råder brist på lämpliga bedömningsmetoder och referensdata för språkfärdigheter i olika flerspråkiga grupper. En tredje orsak är att språkfärdigheterna hos ett flerspråkigt barn kan skilja sig åt mellan språken, eftersom språkfärdigheter påverkas av språkliga exponeringsmönster. Sammantaget leder detta till en risk för såväl överdiagnostik som underdiagnostik av språkstörning hos flerspråkiga. För att komma tillrätta med problemet krävs att man tar hänsyn både till kunskap om flerspråkighet och kunskap om språkstörning.

Olika språkliga domäner påverkas i olika utsträckning av den språkliga exponering barnet utsätts för. Ordförrådet är kanske det område som påverkas allra mest av exponering, och flerspråkiga barn kan ha olika ordkunskaper i sina olika språk. Att enbart bedöma ordförrådet på det ena språket kan därför grovt underskatta flerspråkiga barns ordkunskap, särskilt om flerspråkiga barns prestationer jämförs med enspråkiga barn. Samtidigt har barn med språkstörning ofta svårt att lära sig nya ord, och ett mindre omfattande ordförråd jämfört med jämnåriga med typisk språkutveckling. Att ha ett väl utvecklat ordförråd är en viktig faktor för framtida skolframgång. Därför är det viktigt att ta reda på mer om vilka faktorer som påverkar ordförrådsutvecklingen hos flerspråkiga barn.

Eftersom språkkunskaper (och särskilt ordförråd) är beroende av språkexponering, kan det vara lämpligt att bedöma språklig bearbetningsförmåga vid utredning av misstänkt språkstörning hos flerspråkiga. Ett exempel på en uppgift som mäter språklig bearbetningsförmåga är nonordsrepetition, då barnet får lyssna till påhittade ord för att sedan upprepa dem så likt de kan. I

flera studier har man sett att det är svårt för barn med språkstörning att repetera nonord, medan barn med typisk språkutveckling ofta klarar uppgiften utan större svårigheter. Nonordsrepetition har hävdats vara ett relativt rent mått på fonologiskt arbetsminne, eftersom man inte har hjälp av tidigare språkkunskap när det språkliga materialet är okänt. Senare studier har dock visat att det finns ett samband mellan att ha ett stort ordförråd och god prestation på nonordsrepetition. Sambandet är särskilt starkt i de fall nonorden i hög utsträckning liknar riktiga ord. Det har också visat sig att flerspråkiga barn kan vara missgynnade i förhållande till enspråkiga om nonorden är mycket språkliknande (lika ord i ett visst språk), och att sådana uppgifter tenderar att i lägre utsträckning identifiera de barn som har språkstörning bland flerspråkiga än bland enspråkiga. Vissa forskare har därför föreslagit att det istället kan vara lämpligt att använda uppgifter där nonorden inte är lika något särskilt språk. Förutom effekter av språklikhet och ordförrådskunskaper har tidigare studier visat att svårighetsgraden varierar beroende på hur testorden är uppbyggda. Till exempel ökar svårighetsgraden med fler stavelser och ökad fonologisk komplexitet (förekomst av konsonantkluster och koda).

Utöver att kartlägga barns språkliga färdigheter, så vet vi från internationella studier att information från föräldrar och skolpersonal också kan utgöra viktiga pusselbitar i processen att avgöra om ett barn har språkstörning eller inte. Barnens föräldrar kan berätta om barnets tidiga språkutveckling, nuvarande språkfärdigheter samt huruvida det finns ärftlighet för språksvårigheter i familjen. Barnens lärare kan bidra med värdefull information om barnets språkfärdigheter i förhållande till jämnåriga, samt hur språk- och kommunikationsförmågan fungerar i samtal och inlärningssituationer. Till skillnad från vad många tror så leder flerspråkighet inte till en försenad språkutveckling. Tvärtom kan man förvänta sig att flerspråkiga barn når de tidiga milstolparna i språkutvecklingen – det vill säga när de börjar säga sina första ord och bygga enkla meningar – vid samma tidpunkt som enspråkiga. Däremot är det inte säkert att det sker samtidigt på båda språken. Språksvårigheter finns ofta i familjen; det är vanligare bland barn med språkstörning att ha en släkting med liknande svårigheter än bland barn som inte har språkstörning.

Föreliggande avhandling har tre huvudsakliga syften. Det första syftet är att undersöka ordförrådskunskaper hos 4–7-åriga arabisk-svensktalande barn med typisk språkutveckling. Ordförståelse och ordproduktion undersöks i minoritetsspråket (arabiska) och majoritetsspråket (svenska). Ordförrådsutveckling studeras över tid genom att jämföra barn i olika åldrar. Vidare utforskas hur externa faktorer relaterade till språkexponering (procentuell daglig exponering och exponeringslängd) och familjebakgrund (socioekonomisk status, SES) påverkar barnens ordförråd i minoritetsspråket och majoritetsspråket. Det andra syftet är att förstå hur prestationen på nonordsrepetitionsuppgifter utvecklas med åldern hos arabisk-svensktalande 4–7-åringar med typisk språkutveckling. En del av detta andra syfte är att

undersöka hur prestationen påverkas av egenskaper hos nonorden (antal stavelser, stavelsekomplexitet och språklikhet), samt av barnens tidigare språkerfarenheter (språkexponering och ordförrådsstorlek). Det tredje syftet är att undersöka huruvida flerspråkiga barn med språkstörning kan särskiljas från flerspråkiga barn med typisk språkutveckling, baserat på en bedömning av språkkunskaper (ordförråd) och fonologiskt arbetsminne (nonordsrepetition), samt information om deras tidiga språkutveckling, nuvarande språkfärdigheter, kommunikativa beteende samt deras språkexponering och språkanvändning. Följande övergripande forskningsfrågor ställdes:

- Hur utvecklas ordförrådskunskaper med åldern i båda språken hos 4–7-åriga arabisk-svensktalande barn, och vilka externa faktorer påverkar denna utveckling?
- Hur presterar 4–7-åriga arabisk-svensktalande barn på nonordsrepetition, och hur påverkas deras prestation av nonordens längd, komplexitet och språklikhet, samt av språkexponering och ordförråd?
- Hur presterar flerspråkiga barn med språkstörning på ordförråds- och nonordsrepetitionsuppgifter, vad rapporteras gällande barnens bakgrund, språkfärdigheter och kommunikativa beteende, och hur kan denna information användas för att identifiera språkstörning hos flerspråkiga?

För att besvara dessa frågeställningar genomfördes tre delstudier. Totalt deltog 110 arabisk-svensktalande barn mellan fyra och sju år (4;0–7;11), varav 99 hade typisk språkutveckling och 11 hade språkstörning. För att undersöka ordförståelse och ordproduktion i minoritetsspråket (arabiska) och majoritetsspråket (svenska) användes en svensk och en arabisk version av Cross-linguistic Lexical Task (CLT; Haman et al. 2015). CLT har tagits fram för att undersöka ordförrådskunskaper med jämförbara bedömningsmaterial på båda språken hos flerspråkiga barn. För att undersöka hur barnen presterade på nonordsrepetition användes fyra olika uppgifter. Den första var ett svenskt språkspecifikt test, med 2–5-staviga testord som efterliknade den stavelsestruktur som finns i svenska (LS-Swe; Radeborg et al. 2006). Den andra och tredje uppgiften var en svensk och en arabisk version av ett s.k. kvasiuniversellt test, med 2–5-staviga testord med enkel stavelsestruktur som inte liknar något specifikt språk (QU-Swe och QU-Ara; Chiat 2015). Den fjärde uppgiften var en arabisk (libanesisk) version av ett test med 1–3-staviga testord, som var konstruerat för att undersöka hur stavelsekomplexitet påverkar prestationen på nonordsrepetition (NWRT-Leb; Abou Melhem et al. 2011). Bakgrundsfaktorer som tidig språkutveckling, språklig exponering och SES (föräldrarnas utbildningsnivå) kartlades via ett frågeformulär som barnens föräldrar fyllde i. För barnen med språkstörning genomfördes även intervjuer med barnens föräldrar, logoped samt lärare.

I kapitel 4 undersöktes ordförrådskunskaperna hos de 99 barnen med typisk språkutveckling. Som förväntat hade barnen högre poäng på ordförståelse än

på ordproduktion. Det fanns ingen skillnad i totalpoäng mellan arabiska och svenska. Flera barn presterade maxpoäng på svensk ordförståelse, men inget barn gjorde det på arabisk ordförståelse. Det var en större spridning i resultat på ordförståelse på svenska än på arabiska. För ordproduktion var det stor spridning både på svenska och arabiska. Ordförståelse och ordproduktion ökade med stigande ålder för både arabiska och svenska, då det fanns positiva samband mellan ålder och antalet poäng, och skillnader i medelpoäng mellan fyraåringarna och sjuåringarna på alla deltest. Procentuell daglig exponering (hur stor andel av tiden som de blev exponerade för respektive språk) spelade roll för barnens prestation på ordförrådstesten, men förklarade resultaten i högre utsträckning för ordproduktion än för ordförståelse. Hur länge barnen hade blivit exponerade för svenska var en signifikant prediktor för deras svenska ordförråd. Exponeringslängden var lika viktig som kronologisk ålder för att förklara deras prestation på ordförståelse, medan den var viktigare än kronologisk ålder och för att förklara deras prestation på ordproduktion. Barnens SES (föräldrarnas utbildningsnivå) hade ingen inverkan på barnens prestation på ordförrådstesten varken på (minoritetsspråket) arabiska eller på (majoritetsspråket) svenska.

I kapitel 5 undersöktes fonologiskt arbetsminne med de fyra nonords-repetitionsuppgifterna hos de 99 barnen med typisk språkutveckling. Prestationen ökade med stigande ålder; det fanns positiva samband mellan ålder i månader och antal poäng, samt skillnader i medelpoäng mellan fyraåringarna och sjuåringarna på alla uppgifter. För vissa barn med låga resultat på nonordsrepetition rapporterade föräldrarna att de hade någon riskfaktor förknippad med språkstörning (t ex sen språk-utveckling, ärftlighet för språksvårigheter, föräldraoro, etc.), men det stämde inte in på alla lågpresterande barn. De fyra uppgifterna hade olika svårighetsgrad, och det var en betydande skillnad i hur stor andel av barnen som presterade högt eller lågt. Av de fyra uppgifterna var NWRT-Leb lättast (41 % av barnen presterade 90 % eller bättre) och LS-Swe svårast (endast ett barn presterade 90 % eller bättre). Barnen presterade något bättre på den svenska versionen (QU-Swe) än den arabiska versionen (QU-Ara) av det kvasiuniversella testet. Nonordslängd och stavelsekomplexitet påverkade barnens prestation; nonord som hade fler stavelser eller som innehöll konsonantkluster repeterades korrekt i lägre utsträckning än kortare nonord med lägre grad av komplexitet. En jämförelse gjordes mellan det språklika testet (LS-Swe) och det icke-språklika kvasiuniversella testet (QU-Swe) för att undersöka om språkexponering och ordförrådets storlek hade någon inverkan på prestationen på det språklika testet. Mängden daglig exponering för svenska och exponeringslängd spelade ingen roll för barnens prestation på något av de två testerna. Barn som hade ett större ordförråd på svenska hade bättre poäng på det språklika testet (LS-Swe), men effekten av ordförrådskunskaper på svenska var lika stor för den svenska versionen av det kvasiuniversella testet (QU-Swe).

I kapitel 6 jämfördes de 11 barnen med språkstörning med de 99 barnen med typisk språkutveckling avseende deras prestation på ordförståelse och ordproduktion (CLT) i (minoritetsspråket) arabiska och (majoritetsspråket) svenska, samt på nonordsrepetition (LS-Swe, QU-Swe, QU-Ara och NWRT-Leb). Barnen med språkstörning beskrevs också avseende deras tidiga språkutveckling, språkfärdigheter och kommunikativa beteende, utifrån den information som framkom ur föräldraenkäter och de intervjuer som gjordes med föräldrar, logopeder och lärare. En högre andel av barnen med språkstörning hade en sen språkutveckling och/eller ärftlighet för språksvårigheter jämfört med barnen med typisk språkutveckling. Barnen med språkstörning beskrevs ofta ha svårt med såväl språkförståelse (t ex svårigheter att förstå instruktioner) som språklig uttrycksförmåga (t ex bristande grammatik), samt att de hade ett svagt ordförråd. Några av barnen beskrevs ha stora kommunikationssvårigheter, med många missförstånd och konflikter med jämnåriga, vilket hade en negativ inverkan på deras sociala relationer. På gruppnivå presterade barnen med språkstörning under medelvärdet men inom spannet för sin ålder både på ordförrådstesten och nonordsrepetition. En stor andel av barnen med språkstörning hade särskilt låga ordförrådspoäng på sitt förstaspråk (arabiska), trots omfattande och kontinuerlig exponering sedan födseln. De flesta barnen med språkstörning hade mycket låga resultat på ordförrådstestet i det ena språket, men endast *ett* barn hade mycket låga resultat på *båda* språken. Det fanns en stor överlappning i prestation både på ordförråd och nonordsrepetition mellan barnen med språkstörning och barnen med typisk språkutveckling. Inget av nonordsrepetitionsuppgifterna framstod som tydligt bättre än de andra för att identifiera barn med språkstörning.

Sammantaget visar resultaten att det är viktigt att kartlägga bakgrundsfaktorer som tidig språkutveckling och språkexponering vid utredning av flerspråkiga barn med misstänkt språkstörning. Vidare understryks vikten av att sätta resultat på tester av språkfärdigheter, exempelvis ordförrådstester, i relation till språklig exponering. Flerspråkiga barn med språkstörning behöver inte nödvändigtvis ha väldigt låga ordförrådspoäng på båda språken, men låga resultat på förstaspråket trots omfattande och kontinuerlig exponering kan vara ett varningstecken. Nonordsrepetition bör användas med försiktighet som diagnosverktyg, eftersom det finns en överlappning i prestation mellan barn med språkstörning och barn med typisk språkutveckling. Slutligen behövs mycket mer forskning för att bättre förstå hur språkstörning yttrar sig i olika språkliga domäner hos flerspråkiga som talar olika språkkombinationer. Framtida forskning bör särskilt fokusera på hur resultat på standardiserade språktester bör tolkas i relation till tidigare språkexponering. Framtida studier bör även titta närmare på hur prestation på språktester förhåller sig till funktionell språkförmåga, och i vilken utsträckning redogörelser (från t ex föräldrar och lärare) och observation av funktionell språkförmåga kan vara användbart vid utredning av misstänkt språkstörning i olika flerspråkiga grupper.

Appendix

Appendix 3: Methods

Table A3.1. Items in the LS-Swe task (adapted from Radeborg et al., 2006). Phonemic transcription, syllable structure, number of syllables, number of clusters, and number of codas.

Phonemic transcription	Syllable structure	No of syllables	No of clusters	No of codas
/glʏˈvoː/	ccv.cv	2	1	0
/aˈpɛt/	v.cvc	2	0	1
/ɪˈfʉːm/	v.cvc	2	0	1
/ˈɧɔɾjɛ/	cvc.cv	2	0	1
/naˈkiːt/	cv.cvc	2	0	1
/ˈspʉːmɛ/	ccv.cv	2	1	0
/lɛbʊˈsʉːf/	cv.cv.cvc	3	0	1
/mɵstɾɛˈfalj/	cv.cccv.cvcc	3	2	1
/glɛŋɛˈsɵlp/	ccvc.v.cvcc	3	1	2
/salʊˈtɑːn/	cv.cv.cvc	3	0	1
/hœntˈpʉːlɛ/	cvcc.cv.cv	3	1	1
/nɛsʊˈloː/	cv.cv.cv	3	0	0
/ˈmaŋɛʂˌblɛgɛ/	cvc.vc.ccv.cv	4	1	2
/ɛlʊˈmɔkɪ/	v.cv.cv.cv	4	0	0
/ɔlɪˈtʉːkɛ/	v.cv.cv.cv	4	0	0
/spɵrɪfɾaˈgoːl/	ccv.cv.ccv.cvc	4	2	1
/tɪbɛˈfiːmɛ/	cv.cv.cv.cv	4	0	0
/lɵtʊspɛˈlʉːn/	cv.cv.ccv.cvc	4	1	1
/tœlɪpaˈleːɾʊ/	cv.cv.cv.cv.cv	5	0	0
/ɛɵlɛˌkɾɔmpaˈmiːd/	cv.cv.ccvc.cv.cvc	5	1	2
/fɪmɪglaˈnɛftɪ/	cv.cv.ccv.cvc.cv	5	1	1
/hɪlʊteɾaˈpʉːd/	cv.cv.cv.cv.cvc	5	0	1
/flɛtɛˌmɪŋɛˈɾoːf/	ccv.cv.cvc.v.cvc	5	1	2
/dalabɛlˈhiːmɛ/	cv.cv.cvc.cv.cv	5	0	1

Table A3.2. Items in the QU-Swe task, adapted from Chiat (2015). Phonemic transcription, syllable structure, number of syllables, number of clusters, and number of codas.

Phonemic transcription	Syllable structure	No of syllables	No of clusters	No of codas
/sɪbʊ/	cv.cv	2	0	0
/dʊla/	cv.cv	2	0	0
/nagɪ/	cv.cv	2	0	0
/lʊnɪ/	cv.cv	2	0	0
/sɪpʊla/	cv.cv.cv	3	0	0
/banʊdɪ/	cv.cv.cv	3	0	0
/malɪtʊ/	cv.cv.cv	3	0	0
/lɪmɪka/	cv.cv.cv	3	0	0
/sɪbalɪta/	cv.cv.cv.cv	4	0	0
/mʊkɪdala/	cv.cv.cv.cv	4	0	0
/gasʊlʊmɪ/	cv.cv.cv.cv	4	0	0
/lɪdɪsakʊ/	cv.cv.cv.cv	4	0	0
/sɪpʊnakɪla/	cv.cv.cv.cv.cv	5	0	0
/tʊlɪgasʊmʊ/	cv.cv.cv.cv.cv	5	0	0
/malʊsɪgʊba/	cv.cv.cv.cv.cv	5	0	0
/lɪdapɪmʊtɪ/	cv.cv.cv.cv.cv	5	0	0

Table A3.3. Items in the QU-Ara task, adapted from Chiat (2015). Phonemic transcription, syllable structure, number of syllables, number of clusters, and number of codas.

Phonemic transcription	Syllable structure	No of syllables	No of clusters	No of codas
/zibu/	cv.cv	2	0	0
/lita/	cv.cv	2	0	0
/naki/	cv.cv	2	0	0
/muli/	cv.cv	2	0	0
/sibula/	cv.cv.cv	3	0	0
/banudi/	cv.cv.cv	3	0	0
/nalitu/	cv.cv.cv	3	0	0
/limika/	cv.cv.cv	3	0	0
/sibalita/	cv.cv.cv.cv	4	0	0
/mukidala/	cv.cv.cv.cv	4	0	0
/kasulumi/	cv.cv.cv.cv	4	0	0
/lidizaku/	cv.cv.cv.cv	4	0	0
/sibunakila/	cv.cv.cv.cv.cv	5	0	0
/dulikasumu/	cv.cv.cv.cv.cv	5	0	0
/maluzikuba/	cv.cv.cv.cv.cv	5	0	0
/lidabimudi/	cv.cv.cv.cv.cv	5	0	0

Table A3.4. Items in the NWRT-Leb (Abou Melhem et al., 2011), adapted from the NWR-FRENCH task, (both now part of the LITMUS assessment battery). Phonemic transcription, syllable structure, number of syllables, number of clusters, and number of codas.

Phonemic transcription	Syllable structure	No of syllables	No of clusters	No of codas
/fla/	ccv	1	1	0
/fuk/	cvc	1	0	1
/kib/	cvc	1	0	1
/baf/	cvc	1	0	1
/klu/	ccv	1	1	0
/bli/	ccv	1	1	0
/lafi/	cv.cv	2	0	0
/bukli/	cvc.cv	2	0	1
/kafib/	cv.cvc	2	0	1
/blaklu/	ccv.ccv	2	2	0
/flablu/	ccv.ccv	2	2	0
/flukif/	ccv.cvc	2	1	1
/faku/	cv.cv	2	0	0
/kabi/	cv.cv	2	0	0
/bilu/	cv.cv	2	0	0
/blufa/	ccv.cv	2	1	0
/klifak/	ccv.cvc	2	1	1
/fliku/	ccv.cv	2	1	0
/fablu/	cv.ccv	2	1	0
/bukif/	cv.cvc	2	0	1
/bufaki/	cv.cv.cv	3	0	0
/biklafu/	cv.ccv.cv	3	1	0
/fikubla/	cv.cv.ccv	3	1	0
/kabufik/	cv.cv.cvc	3	0	1
/kifabu/	cv.cv.cv	3	0	0
/kubafli/	cv.cv.ccv	3	1	0
/flibuka/	ccv.cv.cv	3	1	0
/kuflabi/	cvc.cv.cv	3	0	1
/bifakub/	cv.cv.cvc	3	0	1
/klibafu/	ccv.cv.cv	3	1	0

Appendix 5: Phonological working memory

Table A5.1. Z-scores for all NWR tasks and CLT tasks for children who received z-scores below –1.25 on at least one NWR task. Z-scores at –1.25 or lower are marked with '*'. Missing data is marked with '---'.

Child	Age	LS-Swe z-score	QU-Swe z-score	QU-Ara z-score	NWRT-Leb z-score	Ara Comp z-score	Ara Prod z-score	Swe Comp z-score	Swe Prod z-score
BiAra4-02	4;10	–1.21	**–1.63***	–1.22	0.34	**–1.64***	**–2.02***	–0.37	–0.46
BiAra4-10	4;02	0.40	0.00	**–1.55***	–0.68	0.45	0.86	0.12	–0.04
BiAra4-11	4;01	**–2.29***	–1.22	**–1.55***	**–1.70***	–0.33	–0.21	–0.61	–0.74
BiAra4-14	4;03	---	**–2.45***	0.07	–0.68	–0.86	**–1.36***	0.84	1.20
BiAra4-24	4;10	**–1.48***	–0.82	**–2.20***	**–2.93***	1.24	1.19	0.24	0.23
BiAra5-01	5:05	**–1.78***	**–2.10***	**–2.78***	**–2.73***	–0.26	**–1.64***	0.30	–0.17
BiAra5-03	5;09	**–1.50***	**–2.10***	**–1.79***	**–1.94***	–0.26	–0.14	–0.69	–0.99
BiAra5-06	5;04	–1.21	–1.11	**–1.30***	–0.09	**–2.98***	**–1.81***	–0.47	–0.07
BiAra5-10	5;02	–0.93	0.37	–0.81	**–1.41***	0.50	0.36	0.53	0.03
BiAra6-03	6;11	–0.26	0.31	–0.32	**–1.26***	–0.60	**–1.55***	0.30	0.12
BiAra6-05	6;08	0.51	–0.10	**–2.14***	**–1.26***	1.00	0.57	0.99	0.55
BiAra6-06	6;06	**–2.05***	**–1.76***	–1.23	**–1.26***	0.46	0.34	–0.98	**–1.36***
BiAra6-08	6;04	–1.02	**–1.76***	0.14	0.52	–1.13	**–1.47***	1.19	1.42
BiAra6-10	6;07	–0.26	0.31	**–1.69***	–0.67	**–2.20***	**–1.40***	0.40	0.64
BiAra6-16	6;04	0.51	**–1.34***	–0.77	–0.07	–0.20	0.72	–0.19	–0.31
BiAra6-19	6;05	–0.77	0.73	–0.32	**–2.16***	–0.47	**–1.25***	0.50	–0.05
BiAra6-22	6;02	**–1.28***	–0.51	–0.77	–0.37	0.73	0.49	–0.19	–0.23

BiAra6-23	6;03	**–1.53***	**–1.76***	–0.32	**–1.26***	1.00	–0.72	–0.48	–0.49
BiAra6-26	6;08	**–1.28***	**–1.76***	–1.23	**–2.16***	**–2.33***	–0.79	**–1.96***	–1.18
BiAra6-29	6;01	**–1.79***	–0.93	–1.23	–0.67	–0.47	–0.26	**–1.76***	–1.18
BiAra7-05	7;05	**–1.76***	0.05	0.39	0.40	–1.22	0.53	**–3.09***	**–2.44***
BiAra7-16	7;06	–1.18	**–1.65***	**–1.57***	–0.76	–0.12	–0.23	**–2.39***	**–1.55***
BiAra7-19	7;01	–0.89	**–1.65***	**–3.13***	**–3.09***	0.44	0.09	0.08	0.31
BiAra7-21	7;05	**–2.05***	**–1.65***	–0.78	0.79	–0.67	0.09	–0.16	–0.75

Note: Language specific Swedish task (LS-Swe), Quasi-universal Swedish task (QU-Swe), Quasi-universal Arabic task (QU-Ara), and the Non-word repetition task-Lebanese (NWRT-Leb).

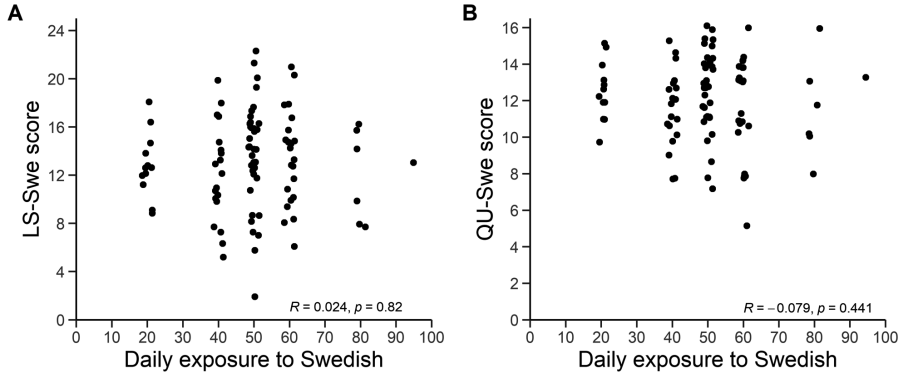

Figure A5.1. Scatterplots of scores on the LS-Swe task (A) and the QU-Swe task (B) and daily exposure to Swedish. The correlation coefficient (Pearson) and the p-value are shown in the bottom right corner.

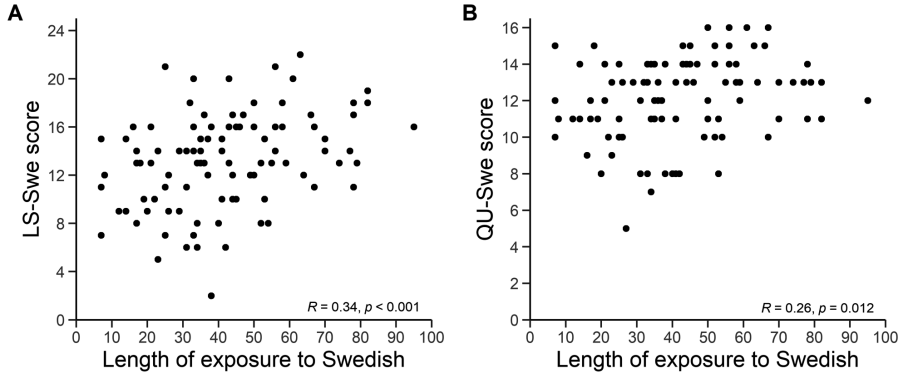

Figure A5.2. Scatterplots of scores on the LS-Swe task (A) and the QU-Swe task (B) and length of exposure to Swedish. The correlation coefficient (Pearson) and the p-value are shown in the bottom right corner.

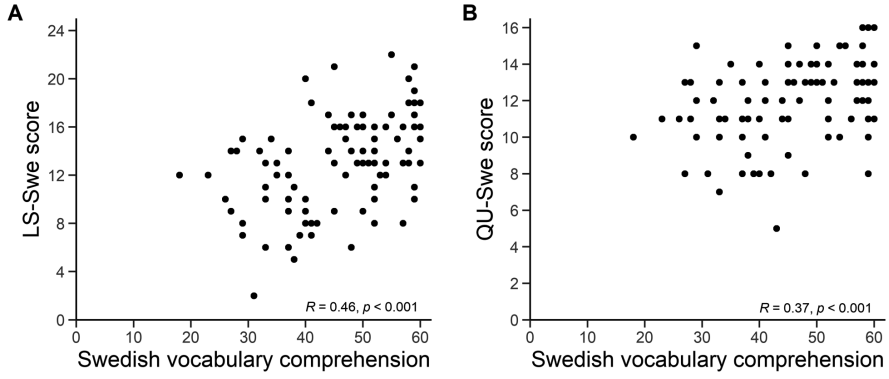

Figure A5.3. Scatterplots of scores on the LS-Swe task (A) and the QU-Swe task (B) and scores on the Swedish CLT comprehension. The correlation coefficient (Pearson) and the p-value are shown in the bottom right corner.

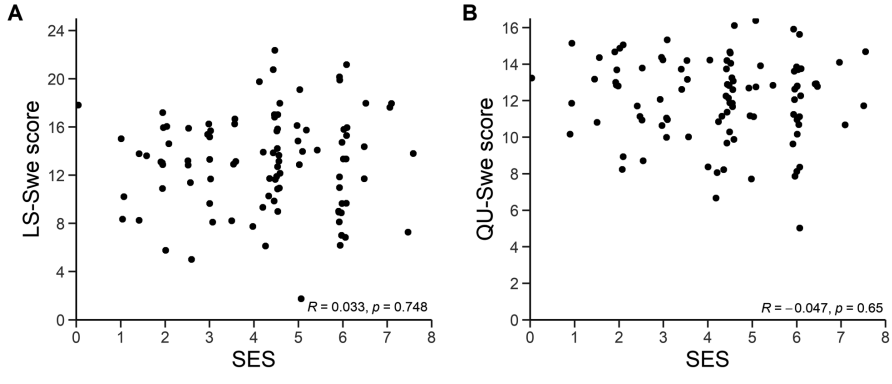

Figure A5.4. Scatterplots of scores on the LS-Swe task (A) and the QU-Swe task (B) and SES (parental education). The correlation coefficient (Pearson) and the p-value are shown in the bottom right corner.

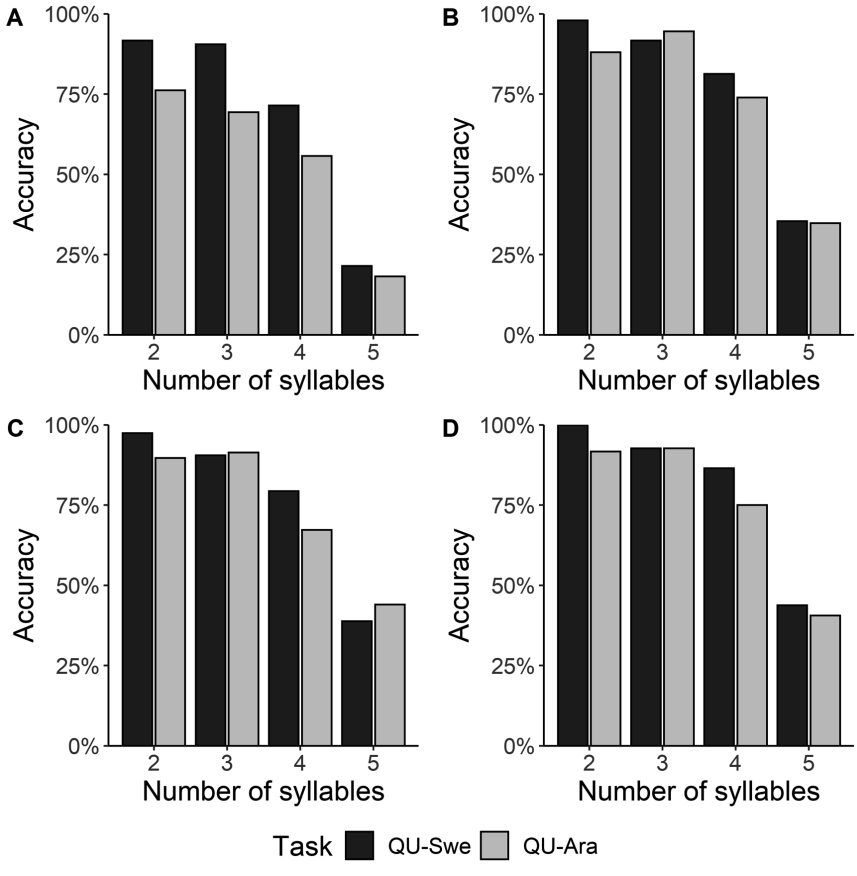

Figure A5.5. Overall accuracy (% correct answers) for items in QU-Swe (Swedish Quasi-universal) and QU-Ara (Arabic Quasi-universal), by age group: 4-year-olds (A), 5-year-olds (B), 6-year-olds (C), and 7-year-olds (D).

Table A5.2. Accuracy, quasi-universal NWR tasks (QU-Swe and QU-Ara): 4-year-olds vs. the three older age groups (5-year-olds, 6-year-olds and 7-year-olds)

Model summary				
Random effects	***(s^2)***	***SD***	***corr.***	
Participant (intercept)	.83	.91		
Task QU-Ara (slope)	.20	.44	.20	
Item (intercept)	.64	.80		
Fixed effects	***Coef.***	***SE***	***z***	***p***
Intercept	.21	.31	.67	.51
Age (4 vs. 5–7)	1.31	.28	4.81	<.001***
Task (QU-Ara vs. QU-Swe)	1.21	.37	3.26	.001**
Length	–1.24	.22	–5.77	<.001***
Task (QU-Ara) x Age	–.49	.27	–1.81	0.07
Task (QU-Ara) x Length	–.62	.32	–1.95	.05
Model evaluation	***Marginal R^2***		***Conditional R^2***	
	.37		.58	

Note. *$p < .05$, **$p < .01$, ***$p < .001$. Logit mixed effects model with random effects: random intercepts for participant and test item, and by-participant random slopes for task. Model fit with maximum likelihood (Laplace approximation). The reference level for the categorical variable is the first category. The values have been rounded off to two decimals. QU-Swe = quasi-universal Swedish task, QU-Ara = quasi-universal Arabic task, Length = item length (number of syllables).

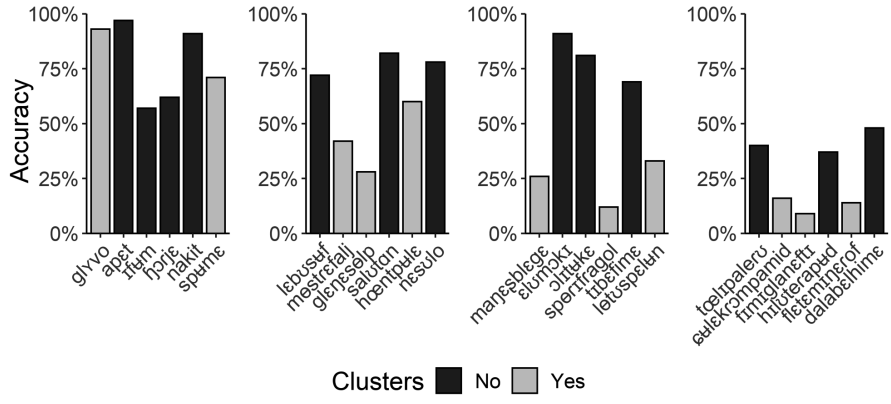

Figure A5.6. Accuracy (% correct answers) for items in the LS-Swe task (Radeborg et al., 2006), by number of syllables (2–5) and clusters (present or not).

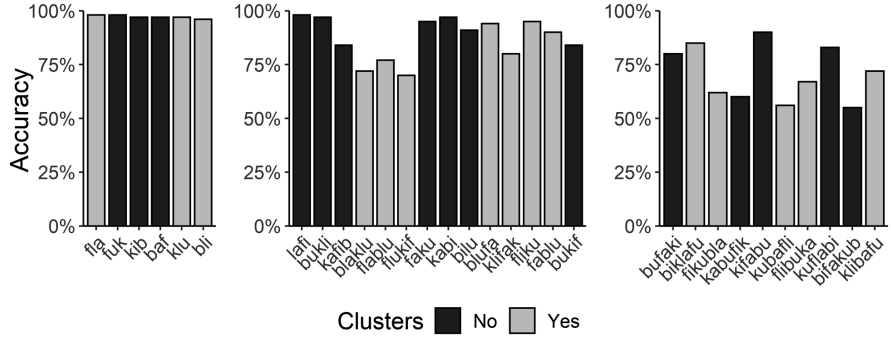

Figure A5.7. Accuracy (% correct answers) for items in the NWRT-Leb (Abou Melhem et al., 2011), by number of syllables (1–3) and clusters (present or not).

Appendix 6: The clinical study

Table A6.1. Ages and Arabic and Swedish vocabulary (CLT) scores of the children in the clinical sample. Mean scores and ranges of the same-age group children in the cross-sectional study in parentheses, max=60 points.

		Arabic		Swedish	
	Age	**Comp**	**Prod**	**Comp**	**Prod**
BiAraLI-01	6;8	37	18	29	13
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)
BiAraLI-02	6;1	38	25	36	24
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)
BiAraLI-03	5;7	25	4	35	24
Cross-sec mean		(46.8)	(33.5)	(45.1)	(29.8)
Cross-sec range		(27–56)	(11–48)	(29–59)	(15–48)
BiAraLI-04	6;0	47	28	47	28
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)
BiAraLI-05	7;3	43	23	46	35
Cross-sec mean		(52.4)	(37.1)	(53.3)	(39.5)
Cross-sec range		(45–59)	(16–52)	(27–60)	(12–53)
BiAraLI-06	5;4	33	12	49	33
Cross-sec mean		(46.8)	(33.5)	(45.1)	(29.8)
Cross-sec range		(27–56)	(11–48)	(29–59)	(15–48)
BiAraLI-07	6;1	46	26	39	26
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)
BiAraLI-08	7;1	55	35	36	10
Cross-sec mean		(52.4)	(37.1)	(53.3)	(39.5)
Cross-sec range		(45–59)	(16–52)	(27–60)	(12–53)
BiAraLI-09	6;7	48	31	57	39
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)
BiAraLI-10	5;0	21	0	35	23
Cross-sec mean		(46.8)	(33.5)	(45.1)	(29.8)
Cross-sec range		(27–56)	(11–48)	(29–59)	(15–48)
BiAraLI-11	6;4	51	41	47	24
Cross-sec mean		(48.5)	(34.5)	(46.9)	(31.6)
Cross-sec range		(31–58)	(10–53)	(27–60)	(11–48)

Table A6.2. Ages and Arabic and Swedish NWR scores of the children in the clinical sample. Mean scores and ranges of the same-age group children in the cross-sectional study in parentheses. Max LS-Swe=24, max QU-Swe=16, max QU-Ara=16, max NWRT-Leb=30.

	Age	LS-Swe	QU-Swe	NWRT-Leb	QU-Ara
BiAraLI-01	6;8	9	7	18	7
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)
BiAraLI-02	6;1	5	8	16	8
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)
BiAraLI-03	5;7	10	13	24	12
Cross-sec mean		(12.3)	(12.3)	(25.3)	(11.7)
Cross-sec range		(6–18)	(8–15)	(15–30)	(6–14)
BiAraLI-04	6;0	12	14	24	11
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)
BiAraLI-05	7;3	13	14	27	11
Cross-sec mean		(15.0)	(12.9)	(27.0)	(12.0)
Cross-sec range		(8–22)	(10–16)	(19–30)	(4–16)
BiAraLI-06	5;4	5	3	13	8
Cross-sec mean		(12.3)	(12.3)	(25.3)	(11.7)
Cross-sec range		(6–18)	(8–15)	(15–30)	(6–14)
BiAraLI-07	6;1	15	11	25	11
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)
BiAraLI-08	7;1	9	11	22	10
Cross-sec mean		(15.0)	(12.9)	(27.0)	(12.0)
Cross-sec range		(8–22)	(10–16)	(19–30)	(4–16)
BiAraLI-09	6;7	18	12	29	13
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)
BiAraLI-10	5;0	10	7	13	7
Cross-sec mean		(12.3)	(12.3)	(25.3)	(11.7)
Cross-sec range		(6–18)	(8–15)	(15–30)	(6–14)
BiAraLI-11	6;4	6	11	15	10
Cross-sec mean		(14.0)	(12.2)	(26.2)	(11.7)
Cross-sec range		(6–21)	(8–16)	(19–30)	(7–16)

Table A6.3. Accuracy (% correct answers) for items in the LS-Swe task by number of syllables and presence of clusters for the 5–7-year olds in the cross-sectional study (N=77) and the children in the clinical study (N=11, age=5;0–7;3).

	TD	**DLD**
Without clusters		
2 syllables	80%	75%
3 syllables	82%	52%
4 syllables	84%	64%
5 syllables	44%	27%
With clusters		
2 syllables	84%	91%
3 syllables	45%	30%
4 syllables	26%	6%
5 syllables	16%	0%

Table A6.4. Accuracy (% correct answers) for items in the NWRT-Leb by number of syllables and presence of clusters for the 5–7-year olds in the cross-sectional study (N=77) and the children in the clinical study (N=11, age=5;0–7;3).

	TD	**DLD**
Without clusters		
1 syllable	99%	70%
2 syllables	95%	71%
3 syllables	80%	80%
With clusters		
1 syllable	92%	86%
2 syllables	88%	58%
3 syllables	74%	51%

Table A6.5. Accuracy (% correct answers) for items in the QU-Swe task and the QU-Ara task by number of syllables for the 5–7-year olds in the cross-sectional study (N=77) and the children in the clinical study (N=11, age=5;0–7;3).

	TD	**DLD**
QU-Swe		
2 syllables	98%	91%
3 syllables	92%	77%
4 syllables	82%	59%
5 syllables	39%	25%
QU-Ara		
2 syllables	90%	93%
3 syllables	93%	70%
4 syllables	72%	59%
5 syllables	40%	25%

References

Abed Ibrahim, L., & Hamann, C. (2017). Bilingual Arabic-German and Turkish-German children with and without Specific Language Impairment: Comparing performance in sentence and nonword repetition tasks. In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 1–17). Somerville, MA: Cascadilla Press.

Abou Melhem, N. (2017, October). *Phonological assessment: the use of Non-Word Repetition Task (LITMUS)*. Presented at the Arabic & Multilingualism Workshop held in Uppsala, 9–10 October.

Abou Melhem, N., Kouba Hreich, E., & dos Santos, C. (2011). *The Non-Word Repetition Task-Lebanese (adapted from the NWR-FRENCH task, now part of the LITMUS battery).* (Unpublished material).

Ahlström, L., & Ljungman, H. (2011). *Åldersreferenser för Peabody Picture Vocabulary Test IV på svenska för flerspråkiga barn i skolår 4.* [Reference data for the Peabody Picture Vocabulary Test IV in Swedish for bilingual children in grade 4]. (Unpublished MA thesis). Karolinska Institute, Stockholm.

Alkass Yousef, S., & Bergström, H. (2011). *Ordförråd på modersmål hos barn i förskoleklass i Sverige: En jämförelse mellan tvåspråkiga och enspråkiga barn.* [Vocabulary in the mother tongue in Swedish children in preparatory class: A comparison between bilingual and monolingual children]. (Unpublished BA thesis). Linköping University, Linköping.

American Psychiatric Association. (1995). *Diagnostic and statistical manual of mental disorders (DSM-4)* (fourth edition). Washington DC: American Psychiatric Press.

American Speech-Language-Hearing Association. (2004). *Preferred practice patterns for the profession of Speech-Language Pathology* (No. PP2004-0019).

Andersson, K., Hansson, K., Rosqvist, I., Lyberg Åhlander, V., Sahlén, B., & Sandgren, O. (2019). The contribution of bilingualism, parental education, and school characteristics to performance on the Clinical Evaluation of Language Fundamentals: fourth edition, Swedish. *Frontiers in Psychology* 10, article 1586.

Anthoni, H., Sucheston, L. E., Lewis, B. A., Tapia-Páez, I., Fan, X., Zucchelli, M., Taipale, M., Stein, C. M., Hokkanen, M-E., Castrén, E., Pennington, B. F., Smith, S. D., Olson, R. K., Tomblin, B. J., Schulte-Körne, G., Nöthen, M., Schumacher, J., Müller-Myhsok, B., Hoffmann, P., Gilger, J. W., Hynd, G. W., Nopola-Hemmi, J., Leppanen, P. H. T., Lyytinen, H., Schoumans, J., Nordenskjöld, M., Spencer, J., Stanic, D., Boon, W. C., Simpson, E., Mäkelä, S., Gustafsson, J-Å., Peyrard-Janvid, M., Iyengar, S. & Kere, J. (2012). The aromatase gene CYP19A1: several genetic and functional lines of evidence supporting a role in reading, speech and language. *Behavior Genetics*, *42*(4), 509–527.

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning*, *62*(3), 769–812.

Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – the nature of the problem. *Bilingualism: Language and Cognition*, *15*(1), 1–4.
Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment.* Bristol: Multilingual Matters.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.
Baddeley, A. D. (2002). Is working memory still working? *American Psychologist*, *56*(11), 851–864.
Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.
Baddeley, A. D., & Hitch, Graham. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory.* (Vols 1–8, pp. 47–89). New York: Academic Press.
Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals–fourth edition, Spanish. *Language, Speech & Hearing Services in Schools*, *49*(2), 292–305.
Barthelom, E., & Åkesson, M. (1995). *Konstruktion, testning och utvärdering av nonord* (Unpublished MA thesis). Lund University, Lund.
Bartoń, K. (2020). *MuMIn: Multi-model inference. R package version 1.43.17.* Retrieved from https://CRAN.R-project.org/package=MuMIn
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech & Hearing Services in Schools*, *36*(3), 188–200.
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.
Bishop, D. V. M. (1997). *Uncommon Understanding. Development and Disorders of Language Comprehension in Children.* Hove, East Sussex: 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 impariments in children: Causes, characteristics, intervention and outcome* (pp. 99–113). Hove, East Sussex: Psychology Press.
Bishop, D. V. M. (2003). *Test for reception of grammar* (second edition). London: Pearson Assessment.
Bishop, D. V. M. (2009a). Genes, cognition, and communication: Insights from neurodevelopmental disorders. *Annals of the New York Academy of Sciences*, *1156*(1), 1–18.
Bishop, D. V. M. (2009b). *Test for reception of grammar – Swedish version* (second edition). Stockholm: Pearson Assessment.
Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems: Terminology for children with language problems. *International Journal of Language & Communication Disorders*, *49*(4), 381–415.

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Developmental Medicine and Child Neurology*, *37*(1), 56–71.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language & Hearing Research*, *58*(6), 1747–1760.

Boerma, T., Leseman, P., Wijnen, F., & Blom, E. (2017). Language proficiency and sustained attention in monolingual and bilingual children with and without language impairment. *Frontiers in Psychology* 8, article 1241.

Boerma, T., Wijnen, F., Leseman, P., & Blom, E. (2017). Grammatical morphology in monolingual and bilingual children with and without language impairment: the case of Dutch plurals and past participles. *Journal of Speech, Language and Hearing Research*, *60*(7), 2064–2080.

Bohnacker, U. (2013). *Språkstörning eller typisk språkutveckling? Utveckling av metoder för språklig bedömning av flerspråkiga barn i Sverige.* [Language Impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden.] Grant proposal to the Swedish research council (Vetenskapsrådet, VR).

Bohnacker, U. (2017). *Sveriges arabisktalande befolkning. Intern rapport.* [The Arabic-speaking population of Sweden. Internal report]. Department of Linguistics and Philology, Uppsala University.

Bohnacker, U. in collaboration with Dillström, S., Lindgren, J. & Öztekin, B. (2018). *Guidelines for scoring CLT (version March 2018).* [Guidelines developed for the BiLI-TAS project]. Department of Linguistics and Philology, Uppsala University.

Bohnacker, U., Haddad, R., Lindgren, J., Öberg, L., & Öztekin, B. (2020/in press). Ordförrådsutveckling på svenska: En jämförelse mellan enspråkiga och flerspråkiga förskolebarn [Vocabulary development in Swedish: A comparison between monolingual and bilingual preschool children]. Forthcoming in S. Haapamäki, L. Forsman, & L. Huldén (Eds.), *Svenskans beskrivning 37: förhandlingar vid trettiosjunde sammankomsten. Åbo 8-9 maj 2019.* Åbo: Åbo Akademi.

Bohnacker, U., Haddad, R., Lindgren, J., Öberg, L., & Öztekin, B. (2021/in press). Ordförrådsutveckling hos flerspråkiga arabisk-svensktalande och turkisk-svensktalande barn i förskoleåldern och vid skolstart. [Vocabulary development in bilingual Arabic-Swedish-speaking and Turkish-Swedish-speaking children in preschool age and at school start]. *Språk och stil*, *31*(1).

Bohnacker, U., Lindgren, J., & Öztekin, B. (2016). Turkish-and German-speaking bilingual 4-to-6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research*, *1*(1), 17–41.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

Botting, N., & Conti-Ramsden, G. (1999). Pragmatic language impairment without autism: the children in question. *Autism*, *3*(4), 371–396.

Botting, N., Conti-Ramsden, G., & Crutchley, A. (1997). Concordance between teacher/therapist opinion and formal language assessment scores in children with language impairment. *International Journal of Language & Communication Disorders*, *32*(3), 317–327.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in 5-year-olds. *Journal of Experimental Child Psychology*, *63*(1), 44–78.

Brusewitz, K., & Tallberg, I.-M. (2010). The Boston Naming Test and Swedish children: normative data and response analysis. *European Journal of Developmental Psychology*, *7*(2), 265–280.

Buac, M., Gross, M., & Kaushanskaya, M. (2014). The role of primary caregiver vocabulary knowledge in the development of bilingual children's vocabulary skills. *Journal of Speech Language and Hearing Research*, *57*(5), 1804–1816.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288.

Carlsson, S. (2009). *Barn och ungdomars prestationer och svarsmönster på ordflödestest.* [Children's and adolescents' performances on verbal fluency tasks]. (Unpublished MA thesis). Karolinska institute, Stockholm.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, *49*(6), 649–671.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1142–1157.

Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 125–147). Bristol: Multilingual Matters.

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1179–1189.

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 429–443.

Choudhury, N., & Benasich, A. A. (2003). A family aggregation study: The influence of family history and other risk factors on language development. *Journal of Speech, Language, and Hearing Research*, *46*(2), 261–272.

Coady, J. A. (2013). Rapid naming by children with and without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *56*(2), 604–617.

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

Cobo-Lewis, A., Pearson, B., Eilers, R., & Umbel, V. (2002a). Effects of bilingualism and bilingual education on oral and written English skills: a multifactor study of standardized test outcomes. In D. K. Oller & R. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 64–97). Clevedon: Multilingual Matters.

Cobo-Lewis, A., Pearson, B., Eilers, R., & Umbel, V. (2002b). Effects of bilingualism and bilingual education on oral and written Spanish skills: a multifactor study of standardized test outcomes. In D. K. Oller & R. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 98–117). Clevedon: Multilingual Matters.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*(3), 544–560.

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533.

DeThorne, L. S., Hart, S. A., Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., & Davison, M. D. (2006). Children's history of speech-language difficulties: genetic influences and associations with reading-related measures. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1280–1293.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168.

Dijkstra, J., Kuiken, F., Jorna, R. J., & Klinkenberg, E. L. (2016). The role of majority and minority language input in the early development of a bilingual vocabulary. *Bilingualism: Language and Cognition*, *19*(1), 191–205.

Dollaghan, C. A., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech & Hearing Research*, *36*(5), 1051–1054.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, *16*(2), 211.

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

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: a meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1077–1088.

dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, *25*(1), 58–71.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (PPVT-IV)* (fourth edition). Bloomington MN: Pearson Assessment.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test (PPVT-III)* (third edition). Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale (BPVS-II)* (second edition). London: NFER-Nelson.

Dunn, L. M., Theriault-Whalen, C. M., & Dunn, L. M. (1993). *The Échelle de vocabulaire en images de Peabody: Adaptation française du Peabody Picture Vocabulary Test*. Toronto, ON: PsyCan.

Edwards, S., Letts, C., & Sinka, I. (2011). *The New Reynell Developmental Language Scales*. London: GL-Assessment.

Eklund, H. (1996). *Ordracet – bedömningsmaterial för snabb automatisk benämningsförmåga*. [Ordracet – a material for assessing rapid automatic naming] Linköping: Eget förlag.

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with Specific Language Impairment: A language endowment spectrum? In R. Paul (Ed.), *Language disorders from a developmental perspecitve* (pp. 83–101). Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, M., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865–878.

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*(5), 529–537.

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research*, *51*(6), 1580–1587.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. J. & Reilly, J. S. (1993). *The MacArthur Comminicatve Development Inventories: User's guide and technical manual.* San Diego, CA: Singular Publishing Group.

Fichman, S., Altman, C., Voloskovich, A., Armon-Lotem, S., & Walters, J. (2017). Story grammar elements and causal relations in the narratives of Russian-Hebrew bilingual children with SLI and typical language development. *Journal of Communication Disorders*, *69*, 72–93.

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

Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., Golcher, F. & Walters, J. (2014). Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian-Hebrew preschool children. In R. K. Silberstein & P. F. Titzmann (Eds.), *The Challenges of Diaspora Migration: Interdisciplinary Perspectives on Israel and Germany* (pp. 63–82). Farnham, Surrey: Ashgate.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2012). *Multilingual Assessment Instrument for Narratives (MAIN)*. Berlin: ZAS.

Ganuza, N., & Hedman, C. (2019). The impact of mother tongue instruction on the development of biliteracy: Evidence from Somali–Swedish bilinguals. *Applied Linguistics*, *40*(1), 108–131.

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

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543.
Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology*, *29*(4), 770–778.
Gathercole, S. E., & Adams, A.-M. (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., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*(2), 200–213.
Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360.
Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*(4), 439–454.
Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127.
Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*(3), 349–367.
Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, *12*(2), 213–237.
Gathercole, V. C. M., Thomas, E. M., Roberts, E. J., Hughes, C. O., & Hughes, E. K. (2013). Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. In V. C. M. Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 20–55). Bristol: Multilingual Matters.
Gibson, T. A., Summers, C., Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. M. (2015). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*, *18*(3), 551–560.
Gillam, R., Peña, E., Bedore, L., Bohman, T., & Mendez-Perez, A. (2013). Identification of Specific Language Impairment in bilingual children: I. assessment in English. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1813–1823.
Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology Section A*, *56*(6), 929–954.
Goral, M., & Conner, P. S. (2013). Language disorders in multilingual and multicultural populations. *Annual Review of Applied Linguistics*, *33*, 128–161.
Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195.
Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech & Hearing Services in Schools*, *30*(2), 196–206.
Grimm, A., & Schulz, P. (2014). Specific Language Impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research*, *7*(4), 821–841.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-language Pathology*, *23*(4), 574–586.

Haddad, R. (in preparation). *Language development in bilingual Arabic-Swedish-speaking children in Sweden* (Doctoral dissertation in preparation). Uppsala University, Uppsala.

Haddad, R. (2017). *Cross-Linguistic Lexical Tasks: Selected Arabic dialects. Adapted from the Lebanese version (CLT-ARA).* (Unpublished material).

Håkansson, G. (2001). Tense morphology and verb-second in Swedish L1 children, L2 children and children with SLI. *Bilingualism: Language and Cognition*, *4*(1), 85–99.

Håkansson, G., & Nettelbladt, U. (1993). Developmental sequences in L1 (normal and impaired) and L2 acquisition of Swedish syntax. *International Journal of Applied Linguistics*, *3*(2), 131–155.

Håkansson, G., Salameh, E.-K., & Nettelbladt, U. (2003). Measuring language development in bilingual children: Swedish-Arabic children with and without language impairment. *Linguistics*, *41*(2), 255–288.

Haman, E., Łuniewska, M., Hansen, P., Simonsen, H.G., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskienė, I., Engel de Abreu, P., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, E., Kapalková, S., Kunnari, S., Levorato, C., Lindgren, J., Mieszkowska, K., Montes Salarich, L., Potgieter, A., Ribu, I., Ringblom, N., Rinker, T., Roch, M., Slančová, D., Southwood, F., Tedeschi, R., Tuncer, A.M., Ünal-Logacev, Ö., Vuksanović, J., & Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, *31*(11–12), 818–843.

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: disentangling bilingualism from language impairment* (pp. 196–239). Bristol: Multilingual Matters.

Hansson, K., & Leonard, L. B. (2006). The use and productivity of verb morphology in specific language impairment: an examination of Swedish. *Linguistics*, *41*(2), 351–379.

Hansson, K., & Nettelbladt, U. (1990). The verbal interaction of Swedish language-disordered pre-school children. *Clinical Linguistics & Phonetics*, *4*(1), 39–48.

Hansson, K., & Nettelbladt, U. (2010). *Grammatiktest för barn (GRAMBA)*. [Grammar test for children]. Lund: Studentlitteratur.

Hansson, K., Nettelbladt, U., & Leonard, L. B. (2003). Indefinite articles and definite forms in Swedish children with Specific Language Impairment. *First Language*, *23*(3), 343–362.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul Brookes.

Hellqvist, B. (1991). *Fonemtest.* [The Phoneme test]. (first edition). Lund: Studentlitteratur.

Hellqvist, B. (2011). *Språkligt impressivt test för barn (SIT).* [Language comprehension test for children]. Lund: Studentlitteratur.

Hellqvist, B. (2013). *Fonemtest.* [The Phoneme test]. (second edition). Lund: Studentlitteratur.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: a systematic review. *Campbell Systematic Reviews*, *14*, 1–156.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433–444.

Holmberg, E., & Sahlén, B. (2000). *Nya NELLI (Neurolingvistisk undersökningsmodell för språkstörda barn)*. [Nya Nelli: A neurolinguistic assessment model for children with language impairment]. (second edition). Malmö: Pedagogisk Design.

Holmberg, E., & Stenkvist, H. (1983). *Nya Lundamaterialet. Kartläggning och bedömning av barns språkliga förmåga.* [The Lund test of Phonology and Grammar. Description and assessment of children's linguistic abilities]. Malmö: Utbildningsproduktion AB.

Holmström, K. (2015). *Lexikal organisation hos en- och flerspråkiga skolbarn med språkstörning.* [Lexical organisation in mono- and bilingual school children with language impariment]. (Unpublished doctoral dissertation). Lund University, Lund.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*, 1–13.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language and Hearing Research*, *53*(6), 1642–1655.

Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Sahlén, B. (2014). Nonword repetition – A clinical marker for specific language impairment in Swedish associated with parents' language-related problems. *PLOS ONE*, *9*(2), article e89544.

Kalnak, N., Peyrard-Janvid, M., Sahlén, B., & Forssberg, H. (2012). Family history interview of a broad phenotype in specific language impairment and matched controls. *Genes, Brain and Behavior*, *11*(8), 921–927.

Kapalková, S., & Slančová, D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics & Phonetics*, *31*(11–12), 893–909.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (second edition). Philadelphia: Lea & Febiger.

Karner, M., & Mattsin, P. (2017). *Anpassning av ett ordförrådstest : En reviderad svensk översättning av Peabody Picture Vocabulary Test - IV*. [Adaptation of a vocabulary test: A revised Swedish translation of the Peabody Picture Vocabulary Test - IV]. (Unpublished MA thesis). Uppsala University, Uppsala.

Kent, H. G., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Insanity*, *67*(2), 37–96.

Ketelaars, M. P., Cuperus, J., Jansonius, K., & Verhoeven, L. (2010). Pragmatic language impairment and associated behavioural problems. *International Journal of Language & Communication Disorders*, *45*(2), 204–214.

Khoury Aouad Saliby, C., dos Santos, C., Kouba Hreich, E., & Messarra, C. (2017). Assessing Lebanese bilingual children: The use of Cross-linguistic Lexical Tasks in Lebanese Arabic. *Clinical Linguistics & Phonetics*, *31*(11–12), 874–892.

Khoury Aouad Saliby, C., Kouba Hreich, E., & Messarra, C. (2017). *Cross-Linguistic Lexical Tasks: Lebanese version (CLT-ARA)*. (Unpublished material).

Kohnert, K. (2010). Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, *43*(6), 456–473.

Krüger-Wahlqvist, M. (2012). *Snabb benämningsförmåga enligt Ordracet: normer samt felsvarsanalys för 6–10-åringar*. [Rapid naming ability in Ordracet: norms and error analysis for 6–10-year-olds]. (Unpublished MA thesis). Uppsala University, Uppsala.

Kupisch, T. (2018). Introduction: Recent developments in early bilingualism. *Bilingualism: Language and Cognition*, *21*(4), 653–655.

Lagergren, C., & Larsson, A. (1992). *Språklig medvetenhet hos förskolebarn – ett prövningsmaterial*. [Linguistic awareness in preschool children - An assessment material]. Göteborg: SIH Läromedel.

Lely, H. K. J. van der, & Battell, J. (2003). Wh-Movement in children with grammatical SLI: a test of the RDDR hypothesis. *Language*, *79*(1), 153–181.

Leonard, L. B. (2014). *Children with Specific Language Impairment* (second edition). Cambridge, MA: MIT Press.

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (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.

Leseman, P. P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Multilingual and Multicultural Development*, *21*(2), 93–112.

Letts, C. (2013). What are the building blocks for language acquisition? Underlying principles of assessment for language impairment in the bilingual context. In V. M. C. Gathercole (Ed.), *Solutions for the assessment of bilinguals* (pp. 36–56). Bristol: Multilingual Matters.

Letts, C., Edwards, S., Sinka, I., Schaefer, B., & Gibbons, W. (2013). Socio-economic status and language acquisition: Children's performance on the new Reynell Developmental Language Scales. *International Journal of Language & Communication Disorders*, *48*(2), 131–143.

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6*. (Doctoral dissertation, Uppsala University). Uppsala: Acta Universitatis Upsaliensis.

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely-related languages. *Linguistic Approaches to Bilingualism*, *10*(5), 587–622.

Lindström, L. (2016, October 8). Modersmålsundervisning dras in. [Mother tongue instruction is cancelled]. *Upsala Nya Tidning*. Retrieved from https://unt.se/uppland/uppsala/modersmalsundervisning-dras-in-4398795.aspx

Loong, T.-W. (2003). Understanding sensitivity and specificity with the right side of the brain. *The British Medical Journal*, *327*(7417), 716–719.

Lukács, Á., Kas, B., & Leonard, L. B. (2013). Case marking in Hungarian children with specific language impairment. *First Language*, *33*(4), 331–353.

Lundeborg Hammarström, I., Kjellmer, L., & Hansson, K. (2016). *The New Reynell Developmental Language Scales – Swedish version*. Stockholm: Hogrefe.

Magnusson, E., & Nauclér, K. (1993). *Bedömning av språklig medvetenhet hos förskolebarn och skolbarn*. [Assessment of linguistic awareness in preschool children and school children]. Löddeköpinge: Pedagogisk Design.
Marshall, C., Marinis, T., & van der Lely, H. (2007). Passive verb morphology: The effect of phonotactics on passive comprehension in typically developing and Grammatical-SLI children. *Lingua*, *117*(8), 1434–1447.
McGee, S. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine*, *17*(8), 647–650.
Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology* 8, article 1442.
Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*(1), 5–34.
Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352.
Michas, I. C., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology*, *12*(2), 147–163.
Mikoczy, T., & Nyman, A. (2008). *Lexikal förmåga hos flerspråkiga svensk-arabiska barn i skolår 4. En undersökning med PPVT-III och Kent & Rosanoffs associationstest.* [Lexical abilities in bilingual Swedish-Arabic children in grade 4. An assessment using PPVT-III and the Kent & Rosanoff word association task]. (Unpublished MA thesis). Lund University, Lund.
Miranda, A. E., McCabe, A., & Bliss, L. S. (1998). Jumping around and leaving things out: A profile of the narrative abilities of children with Specific Language Impairment. *Applied Psycholinguistics*, *19*(4), 647–667.
Montanari, E. G., Abel, R., Graßer, B., & Tschudinovski, L. (2018). Do bilinguals create two different sets of vocabulary for two domains? Vocabulary development and overlap in the first years of schooling. *Linguistic Approaches to Bilingualism*, *8*(4), 502–522.
Mueller, K. L., & Tomblin, J. B. (2012). Examining the comorbidity of Language Impairment and Attention-Deficit/Hyperactivity Disorder. *Topics in Language Disorders*, *32*(3), 228–246.
Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, *48*(1), 61–78.
Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1033–1047.
Namei, S. (2004). Bilingual lexical development: a Persian–Swedish word association study. *International Journal of Applied Linguistics*, *14*(3), 363–388.
Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, *48*(2), 439–458.
Nayeb, L., Wallby, T., Westerlund, M., Salameh, E.-K., & Sarkadi, A. (2015). Child healthcare nurses believe that bilingual children show slower language development, simplify screening procedures and delay referrals. *Acta Paediatrica*, *104*(2), 198–205.

Nettelbladt, U. (2007). Lexikal utveckling. [Lexical development]. In U. Nettelbladt & E.-K. Salameh (Eds.), *Språkutveckling och språkstörning hos barn* (pp. 199–230). Lund: Studentlitteratur.

Newbury, D. F., & Monaco, A. P. (2010). Genetic advances in the study of speech and language disorders. *Neuron*, *68*(2), 309–320.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with Language Impairments: Examining syntactic development. *American Journal of Speech-Language Pathology*, *17*(4), 356–366.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with a history of language impairments: a follow-up investigation. *American Journal of Speech- Language Pathology*, *18*(3), 241–251.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, A., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257.

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist–2. *International Journal of Language & Communication Disorders*, *39*(3), 345–364.

Oberauer, K. (2009). Design for a working memory. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 45–100).

Öberg, L., & Lindgren, J. (2019). Referensdata för två nonordsrepetitionstest för svensktalande vuxna: effekter av nonordstyp, nonordslängd och förekomst av konsonantkluster. *Logopeden*, (4), 12–17.

Öztekin, B. (2019). *Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7*. (Doctoral dissertation, Uppsala University). Uppsala: Acta Universitatis Upsaliensis.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: implications of similarities with specific language impairment. *Language, Speech & Hearing Services in Schools*, *36*(3), 172–187.

Paradis, J. (2007). Early bilingual and multilingual acquisition. In P. Auer & L. Wei (Eds.), *Handbook of Multilingualism and Multilingual Communication* (pp. 15–44). Berlin: De Gruyter Mouton.

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.

Paradis, J., & Crago, M. (2000). Tense and temporality: A comparison between children learning a second language and children with SLI. *Journal of Speech, Language, and Hearing Research*, *43*(4), 834–847.

Paradis, J., Emmerzael, K., & Sorenson 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., & Grüter, T. (2014). Introduction to 'Input and experience in bilingual development'. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 1–14). Amsterdam: John Benjamins.

Parkvall, M. (2015). *Sveriges språk i siffror: vilka språk talas och av hur många?* [Sweden's languages in numbers: which languages are spoken and by how many speakers?]. Stockholm: Språkrådet & Morfem.
Patterson, J. L. (2000). Observed and reported expressive vocabulary and word combinations in bilingual toddlers. *Journal of Speech, Language, and Hearing Research*, *43*(1), 121–128.
Paul, R., & Hernandez, R. (1996). Narrative development in late talkers: Early school age. *Journal of Speech & Hearing Research*, *39*(6), 1295–1303.
Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(3), 399–410.
Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*, *43*(1), 93–120.
Peña, E. D., 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.
Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B., & Bedore, L. (2014). *Bilingual English Spanish Assessment*. San Rafael, CA: A-R clinical publishing.
Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, *60*(1), 283–306.
Petersen, F., & Wail Yassin, M. (2012). *Lexikal förmåga och användning hos arabisk-svenska barn med och utan språkstörning.* [Lexical ability and use in Arabic-Swedish children with and without language impairment]. (Unpublished MA thesis). Lund University, Lund.
Prevoo, M. J. L., Malda, M., Mesman, J., Emmen, R. A. G., Yeniad, N., Van Ijzendoorn, M. H., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: different pathways for host and ethnic language. *Journal of Child Language*, *41*(05), 963–984.
Purse, K., & Gardner, H. (2013). Does formal assessment of comprehension by SLT agree with teachers' perceptions of functional comprehension skills in the classroom? *Child Language Teaching and Therapy*, *29*(3), 343–357.
R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* Retrieved from https://www.R-project.org/
Radeborg, K., Barthelom, E., Sjöberg, M., & Sahlén, B. (2006). A Swedish non-word repetition test for preschool children. *Scandinavian Journal of Psychology*, *47*(3), 187–192.
Ramus, F., Marshall, C., Rosen, S., & van der Lely, H. (2013). Phonological deficits in Specific Language Impairment and developmental dyslexia: towards a multidimensional model. *Brain*, *136*(2), 630–645.
Renfrew, C. (1969). *Bus Story Test: A test of narrative speech*. Oxon: Winslow Press.
Rescorla, L., & Alley, A. (2001). Validation of the language development survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research*, *44*(2), 434–445.
Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1398–1411.
Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without Specific Language Impairment: a longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language and Hearing Research*, *58*(2), 345–359.

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research*, *51*(2), 394–407.

Ringblom, N., Håkansson, G., & Lindgren, J. (2014). *Cross-Linguistic Lexical Tasks: Swedish version (CLT-SWE)*. (Unpublished material).

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *55*(3), 683–694.

Rispens, J., Baker, A., & Duinmeijer, I. (2015). Word recognition and nonword repetition in children with language disorders: The effects of neighborhood density, lexical frequency, and phonotactic probability. *Journal of Speech, Language and Hearing Research*, *58*(1), 78–92.

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development*, *70*(1), 92–106.

Rom, A., & Leonard, L. B. (1990). Interpreting deficits in grammatical morphology in specifically language-impaired children: Preliminary evidence from Hebrew. *Clinical Linguistics & Phonetics*, *4*(2), 93–105.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Ryder, N., & Leinonen, E. (2014). Pragmatic language development in language impaired and typically developing children: incorrect answers in context. *Journal of Psycholinguistic Research*, *43*(1), 45–58.

Rydland, V., Grøver, V., & Lawrence, J. (2014). The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: the role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood. *Journal of Child Language*, *41*(2), 352–381.

Sahlén, B., Reuterskiöld-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Non-word repetition in children with language impairment – pitfalls and possibilities. *International Journal of Language & Communication Disorders*, *34*(3), 337–352.

Saiegh-Haddad, E., & Ghawi-Dakwar, O. (2017). Impact of diglossia on word and non-word repetition among language impaired and typically developing Arabic native speaking children. *Frontiers in Psychology* 8, article 2010.

Salameh, E.-K. (2003). *Language impairment in Swedish bilingual children – epidemiological and linguistic studies.* (Doctoral dissertation, Lund University). Lund: Studies in Logopedics and Phoniatrics.

Salameh, E.-K. (2011). Lexikal utveckling på svenska och arabiska vid tvåspråkig undervisning. [Lexical development in Swedish and Arabic in bilingual schooling]. *EDUCARE 3: Tvåspråkig undervisning på svenska och arabiska i mångkulturella storstadsskolor*, 205–225.

Salameh, E.-K. (2015). *Grammatisk undersökning av Svenska som andraspråk (GrUS)*. [Grammatical assessment of Swedish as a second language]. Lund: Nypon förlag.

Salameh, E.-K., Håkansson, G., & Nettelbladt, U. (2004). Developmental perspectives on bilingual Swedish-Arabic children with and without language impairment: a longitudinal study. *International Journal of Language & Communication Disorders*, *39*(1), 65–90.

Salameh, E.-K., Nettelbladt, U., & Gullberg, B. (2002). Risk factors for language impairment in Swedish bilingual and monolingual children relative to severity. *Acta Paediatrica*, *91*(12), 1379–1384.

Salameh, E.-K., Nettelbladt, U., Håkansson, G., & Gullberg, B. (2002). Language impairment in Swedish bilingual children: a comparison between bilingual and monolingual children in Malmö. *Acta Paediatrica*, *91*(2), 229–234.

Salameh, E.-K., Nettelbladt, U., & Norlin, K. (2003). Assessing phonologies in bilingual Swedish-Arabic children with and without language impairment. *Child Language Teaching and Therapy*, *19*(3), 338–364.

Samson, I., & Rasmussen, P. (2003). *Ordracet – normdata för 6-åringar.* [Ordracet – normative data for 6-year-olds]. (Unpublished MA thesis). Karolinska institute, Stockholm.

Schneider, P., & Hayward, D. (2010). Who does what to whom: introduction of referents in children's storytelling from pictures. *Language, Speech & Hearing Services in Schools*, *41*(4), 459–473.

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology*, *30*(4), 224–238.

Semel, E., Wiig, E. H., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals (CELF-IV), Swedish version* (fourth edition). Stockholm: Pearson Assessment.

SFS 2010:800, chapter 10, §7. *Skollagen.* [The Education Act]. Stockholm: Utbildningsdepartementet.

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), 1704–1719.

Sices, L., Taylor, H. G., Freebairn, L., Hansen, A., & Lewis, B. (2007). Relationship between speech-sound disorders and early literacy skills in preschool-age children: impact of comorbid language impairment. *Journal of Developmental & Behavioral Pediatrics*, *28*(6), 438–447.

Skarakis-Doyle, E., Dempsey, L., & Lee, C. (2008). Identifying language comprehension impairment in preschool children. *Language, Speech & Hearing Services in Schools*, *39*(1), 54–65.

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, *57*(12), 1360–1369.

Sorenson Duncan, T., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language and Hearing Research*, *59*(1), 39–48.

SOU 2016:46. *Samordning, ansvar och kommunikation: vägen till ökad kvalitet i utbildningen för elever med vissa funktionsnedsättningar: slutbetänkande*. [Coordination, responsibility and communication: The path to increase the quality of education for pupils with certain disabilities: final report]. Retrieved from https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2016/06/sou-201646/

Spaulding, T. J., Hosmer, S., & Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *International Journal of Speech-Language Pathology*, *15*(5), 453–462.

Spitz, R. V., Tallal, P., Flax, J., & Benasich, A. A. (1997). Look who's talking: A prospective study of familial transmission of language impairments. *Journal of Speech, Language, and Hearing Research*, *40*(5), 990–1001.

Stark, R., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, *46*(2), 114–122.

Statistics Sweden. (2017a). Utrikesfödda efter födelseland och invandringsår. [Inhabitants born abroad by country of birth and immigration year]. Retrieved 24 September 2017, from https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/#TabelleriStatistikdatabasen

Statistics Sweden. (2017b). Utrikesfödda efter län, kommun och födelseland. [Inhabitants born abroad by county, municipality and country of birth]. Retrieved 4 May 2017, from https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/#TabelleriStatistikdatabasen

Statistics Sweden. (2020). Statistikdatabasen. [The statistics database]. Retrieved 2 June 2020, from https://www.statistikdatabasen.scb.se/

Stromswold, K. (1998). Genetics of spoken language disorders. *Human Biology*, *70*(2), 297–324.

Sundström, S., Samuelsson, C., & Lyxell, B. (2014). Repetition of words and non-words in typically developing children: The role of prosody. *First Language*, *34*(5), 428–449.

Svensson, Y., & Tuominen-Eriksson, A.-M. (2000). *Bussagan*. [The Bus Story Test, Swedish version]. Göteborg: Specialpedagogiska institutet.

Swedish Migration Agency. (2017). Beviljade uppehållstillstånd efter 30 största medborgarskapsländerna per kategori rangordnat efter senaste året 2009-2016. [Granted residence permits by the 30 largest countries of origin per category ordered by the latest year 2009-2016]. Retrieved 4 May 2017, from https://www.migrationsverket.se/Om-Migrationsverket/Statistik/Oversikter-och-statistik-fran-tidigare-ar.html

Swedish National Agency for Education. (2017). Antal elever i grundskolan som det finns information om att de är berättigade till och deltar i modersmåls-undervisning i arabiska. [Number of pupils in grade 1–9 eligible for and participating in mother tongue instruction in Arabic]. Retrieved 4 May 2017, from https://www.skolverket.se/polopolyfs/1.259231!/GrundEleverRiksTab8B2017webb.xls

Swedish National Agency for Education. (2019). Elever med modersmåls-undervisning. [Pupils with Mother Tongue Instruction]. Retrieved 2 June 2020, from https://www.skolverket.se/skolutveckling/statistik/sok-statistik-om-forskola-skola-och-vuxenutbildning

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36.

Tallberg, I.-M. (2005). The Boston Naming Test in Swedish: Normative data. *Brain and Language*, *94*(1), 19–31.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445.

Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 331–358). Bristol: Multilingual Matters.

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, *23*(1), 236–255.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16.

Thordardottir, E., & Ellis Weismer, S. (2002). Content mazes and filled pauses in narrative language samples of children with specific language impairment. *Brain and Cognition*, *48*(2-3), 587–592.

Tomblin, J. B., & Buckwalter, P. R. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research*, *41*(1), 188–199.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245–1260.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech & Hearing Research*, *39*(6), 1284–1294.

Topbaş, S., Güven, S., Aydin Uysal, A., & Kazanoglu, D. (2017). Language impairment in Turkish-speaking children. In B. Haznedar & F. N. Ketrez (Eds.), *The acquisition of Turkish in childhood* (pp. 295–324). Amsterdam: John Benjamins.

Topbaş, S., Kaçar-Kütükçü, D., & Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical Linguistics & Phonetics*, *28*(7/8), 602–616.

Trauner, D., Wulfeck, B., Tallal, P., & Hesselink, J. (2000). Neurological and MRI profiles of children with developmental language impairment. *Developmental Medicine & Child Neurology*, *42*(7), 470–475.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment.* (pp. 301–330). Bristol: Multilingual Matters.

UNESCO Institute for Statistics. (2012). *International Standard Classification of Education: ISCED 2011*. Retrieved from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

Unsworth, S. (2013a). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*(1), 86–110.

Unsworth, S. (2013b). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, *33*, 21–50.

Unsworth, S. (2016). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of Child Language*, *43*(3), 608–634.

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Development of phonological processing skills in children with Specific Language Impairment with and without literacy delay: a 3-year longitudinal study. *Journal of Speech, Language, and Hearing Research*, *55*(4), 1053–1067.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212.

Wehbe, P. (2020). *Introduktion av karaktärer vid berättande hos flerspråkiga svensk-arabisktalande barn med respektive utan språkstörning.* [Introduction of characters in the narratives of bilingual Arabic-Swedish-speaking children with and without Lanugage Impairment]. (Unpublished MA thesis). Uppsala University, Uppsala.

Westlin, M., & Ytterdal, O. (2007). *Strategier vid bildbenämning: En explorativ studie av 3–4-åringars ordförråd.* [Strategies in picture naming: An explorative study of vocabulary in 3–4-year-olds]. (Unpublished MA thesis). Karolinska Institute, Stockholm.

Widaman, K. F. (2006). Missing data: What to do with or without them. In K. McCartney, M. R. Burchinal, & K. L. Bub (Eds.), *Best practices in quantitative methods for developmentalists.* (Vol. 3, pp. 42–64). Boston, MA: Blackwell

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool* (second edition). Toronto: The Psychological Corporation.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York, NY: Routledge.

World Health Organization. (1992). *ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: an epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1562–1592.

ACTA UNIVERSITATIS UPSALIENSIS

Studia Linguistica Upsaliensia

Editors: Michael Dunn and Joakim Nivre

1. *Jörg Tiedemann*, Recycling translations. Extraction of lexical data from parallel corpora and their application in natural language processing. 2003.
2. *Agnes Edling*, Abstraction and authority in textbooks. The textual paths towards specialized language. 2006.
3. *Åsa af Geijerstam*, Att skriva i naturorienterande ämnen i skolan. 2006.
4. *Gustav Öquist,* Evaluating Readability on Mobile Devices. 2006.
5. *Jenny Wiksten Folkeryd,* Writing with an Attitude. Appraisal and student texts in the school subject of Swedish. 2006.
6. *Ingrid Björk,* Relativizing linguistic relativity. Investigating underlying assumptions about language in the neo-Whorfian literature. 2008.
7. *Joakim Nivre, Mats Dahllöf and Beáta Megyesi,* Resourceful Language Technology. Festschrift in Honor of Anna Sågvall Hein. 2008.
8. *Anju Saxena & Åke Viberg*, Multilingualism. Proceedings of the 23rd Scandinavian Conference of Linguistics. 2009.
9. *Markus Saers*, Translation as Linear Transduction. Models and Algorithms for Efficient Learning in Statistical Machine Translation. 2011.
10. *Ulrika Serrander*, Bilingual lexical processing in single word production. Swedish learners of Spanish and the effects of L2 immersion. 2011.
11. *Mattias Nilsson*, Computational Models of Eye Movements in Reading : A Data-Driven Approach to the Eye-Mind Link. 2012.
12. *Luying Wang,* Second Language Acquisition of Mandarin Aspect Markers by Native Swedish Adults. 2012.
13. *Farideh Okati*, The Vowel Systems of Five Iranian Balochi Dialects. 2012.
14. *Oscar Täckström,* Predicting Linguistic Structure with Incomplete and Cross-Lingual Supervision. 2013.
15. *Christian Hardmeier,* Discourse in Statistical Machine Translation. 2014.
16. *Mojgan Seraji,* Morphosyntactic Corpora and Tools for Persian. 2015.
17. *Eva Pettersson,* Spelling Normalisation and Linguistic Analysis of Historical Text for Information Extraction. 2016.
18. *Marie Dubremetz,* Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora. 2017.
19. *Josefin Lindgren,* Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6. 2018.
20. *Vera Wilhelmsen,* A Linguistic Description of Mbugwe with Focus on Tone and Verbal Morphology. 2018.
21. *Yan Shao,* Segmenting and Tagging Text with Neural Networks. 2018.
22. *Ali Basirat,* Principal Word Vectors. 2018.
23. *Marc Tang*, A typology of classifiers and gender. 2018.
24. *Miryam de Lhoneux*, Linguistically Informed Neural Dependency Parsing for Typologically Diverse Languages. 2019.
25. *Buket Öztekin*, Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7. 2019.

26. *Gongbo Tang,* Understanding Neural Machine Translation: Investigating Linguistic Phenomena and Attention Mechanisms. 2020.
27. *Linnéa Öberg.* Words and non-words: Vocabulary and phonological working memory in Arabic-Swedish-speaking 4–7-year-olds with and without a diagnosis of Developmental Language Disorder. 2020.