

# Dissertation

## The role of language input in typical bilingual language development.

Implications for the assessment of atypical bilingual language development.

Theresa Sophie Bloder, M.Sc.  
Wastiangasse 4/8, 8010 Graz, Österreich

Catholic University Eichstätt-Ingolstadt  
Sprach- und Literaturwissenschaftliche Fakultät (SLF)

First Advisor Prof. Dr. Tanja Rinker  
Second Advisor Prof. Dr. Valerie Shafer

## Acknowledgements

I would like to thank my advisors Prof. Dr. Tanja Rinker and Prof. Dr. Valerie Shafer for their input and guidance throughout this PhD while at the same time giving me the liberty to develop and follow my own ideas. Tanja, your words of advice and encouragement made me not lose track of the big picture and taught me that perseverance always pays off not only, but especially amidst a pandemic. Valerie, thank you so much for welcoming me to your lab – twice – and for introducing me to the world of EEG. My deepest gratitude goes to the two of you for always allocating time in your busy schedules to meet with me and discuss any questions, queries or concerns I had not only with regard to purely academic but also personal perspectives. Further, I would like to thank Dr. Maria Luisa Lorusso and Maren Eikerling for letting me join in on parts of their project. Thank you, Maria Luisa, for all the valuable lessons on statistics and academic writing. They will accompany me throughout the course of my academic career. To Prof. Dr. Suzanne van der Feest, thank you for teaching me how to measure Voice Onset Time and for letting me join your lab meetings all the way across the Atlantic Ocean. Last but not least, special thanks to Prof. Dr. Hia Datta for spontaneously agreeing to help me record the stimuli for my EEG experiment.

I am further grateful for my fellow MultiMind and New York colleagues who soon became friends. I would especially like to thank Maren for being the best travel companion, writing buddy, co-conference presenter, and PhD confidant I could have wished for. Your work ethic and ideas have inspired and pushed me. Without you, this PhD experience would not have been the same and I am already looking forward to all our collaborations and joint ventures in the future!

Finally, a project like this would not have been possible without the necessary financial support. I would, thus, like to thank the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 765556 for funding this PhD and the Deutscher Akademischer-Austauschdienst (DAAD) for allowing me to finish this project at Prof. Dr. Valerie Shafer's lab at the Graduate Center at City University of New York.

## **Kurzzusammenfassung**

Der Kontakt zu mehr als einer Sprache beeinflusst die sprachliche Entwicklung von Kindern (Kehoe & Kannathasan, 2021). Es wurde gezeigt, dass insbesondere die relative Menge an Sprachinput bei zweisprachigen Kindern den Fortschritt in ihrer Sprachentwicklung beeinflusst. Zweisprachige Kinder können nie die gleiche Zeit für den Erwerb einer ihrer Sprachen aufwenden wie einsprachige Gleichaltrige, da ihr alltägliches sprachliches Umfeld auf zwei Sprachen aufgeteilt ist. Dies kann letztendlich dazu führen, dass typisch entwickelte zweisprachige Kinder mit reduziertem Kontakt zu einer ihrer Sprachen ähnliche sprachliche Merkmale wie einsprachiger Kinder mit einer Sprachentwicklungsstörung (SES) zeigen (Paradis, 2010). Aus diesem Grund ist es für Kliniker\*innen (z.B. Logopäd\*innen, klinische Linguist\*innen, Patholinguist\*innen etc.) besonders schwierig, zwischen einer tatsächlichen Sprachentwicklungsstörung und typischen Abweichungen von der (einsprachigen) Norm aufgrund begrenzter Sprachexposition zu unterscheiden. In Deutschland hat fast die Hälfte aller Kinder in logopädischer Behandlung einen zwei- oder mehrsprachigen Sprachintergrund (Lüke & Ritterfeld, 2011). Trotz des hohen Anteils an bilingualen Kindern in der Logopädie ist die Beurteilung und Behandlung von SES in dieser Population überwiegend monokulturell und monolingual geprägt (Scharff Rethfeldt, 2016). Frühere Studien haben Aufgaben zum Nachsprechen von Nichtwörtern („Nonword repetition tasks“; NWRs) als vielversprechende Instrumente für die klinische Differenzierung zwischen bilingualen typisch entwickelten und SES-Kindern identifiziert, da sie relativ frei vom Einfluss der kindlichen Spracherfahrung sind (Schwob et al., 2021). Einige Wissenschaftler\*innen haben jedoch die Frage aufgeworfen, ob zweisprachige Kinder durch sprachspezifische Charakteristika von Nichtwörtern benachteiligt werden könnten (z.B. Chiat, 2015).

In drei separaten Studien soll in dieser Dissertation (1) der Einfluss bilingualer Sprachexposition auf die Sprachlautverarbeitung, ein wesentlicher Baustein der sprachlichen Entwicklung, aufgezeigt werden, (2) evaluiert werden, wie bilinguale Kinder in Deutschland im klinischen Kontext diagnostiziert werden und welche Faktoren zu positiven Einstellungen von Logopäd\*innen der Mehrsprachigkeit gegenüber beitragen, und (3) untersucht werden, ob die Nachsprechleistung von sprachspezifischen versus unspezifischen (d.h. sprachneutraleren) Nichtwörtern bei bilingualen Kindern dem Einfluss ihrer individuellen Sprachexposition unterliegt.

In der ersten Studie (siehe Kapitel 2) wurde die Wahrnehmung und Produktion von Voice Onset Time (VOT) bei bilingualen italienisch-deutschen, monolingualen deutschen und monolingualen italienischen Kindern mithilfe einer Kombination von neurophysiologischen und behavioralen Messmethoden untersucht. Vierzig fünfjährige Kinder (16 monolingual deutsch-, 24 bilingual italienisch-deutschsprachig) nahmen an der elektroenzephalographischen (EEG) Studie teil, in der die automatische Sprachlautverarbeitung von italienisch- bzw. deutschspezifischen VOT-Kontrasten untersucht wurde. Um die sprachspezifische VOT Produktion der Teilnehmer\*innen zu vergleichen, wurde zusätzlich zu den 16 monolingual deutschen und 24 bilingual italienisch-deutschen Kindern eine Kontrollgruppe von fünf monolingual italienisch-sprachigen Kindern mit einer Bildbenennungsaufgabe getestet, in der die Produktion sprachspezifischer, wortinitialer VOT anhand von je zehn Wörtern erfasst wurde. Akustische Messungen wurden in Praat (Boersma & Weenink, 2013) durchgeführt. Der sprachliche Hintergrund der Kinder wurde anhand eines Elternfragebogens ermittelt. In der zweiten Studie (siehe Kapitel 3) nahmen 66 Logopäd\*innen aus Deutschland an einer Online-Umfrage teil, in der ihre Einstellungen zu und ihr Umgang mit Zweisprachigkeit in ihrer täglichen klinischen Praxis untersucht wurden. Schließlich wurde in der dritten Studie (siehe Kapitel 4) ein neuer NWRT entwickelt, der sowohl italienisch- und deutschspezifische als auch sprachunspezifische Nichtwörter enthält. Der NWRT wurde in einer Online-Plattform implementiert und mit einer Gruppe von 26 zweisprachigen italienisch-deutschen TD-Kindern durchgeführt. Die Nachsprechleistung der Kinder in den verschiedenen Nichtwort-Unterkategorien wurde im Kontext ihres sprachlichen Hintergrunds analysiert (erfasst anhand desselben Fragebogens wie in Studie 1).

Studie 1 zeigte, dass sich italienisch-deutsch bilinguale fünfjährige Kinder und die monolingual deutsche Kontrollgruppe bei der Verarbeitung sprachspezifischer VOT-Stimuli ähnlich verhalten. Auf die Sprachexposition zurückzuführende Unterschiede zwischen den beiden Gruppen zeigten sich allerdings bei der Verarbeitung eines deutschen „Long Lag“-Stimulus. Überraschenderweise hatte die bilinguale Gruppe keinen Vorteil bei der Verarbeitung der italienischen „Voicing Lead“-Stimuli. Das deutet darauf hin, dass die zweisprachigen Fünfjährigen die italienischen Kontraste (noch) nicht automatisch verarbeiten (Yu et al. 2019). Die akustische Messung der produktiven VOT für italienische und deutsche Wörter stimmte mit dieser Interpretation überein. Insbesondere zeigten die bilingualen Kinder die Beherrschung des deutschen, aber Variabilität hinsichtlich des italienischen VOT-Kontrasts. Es wurden Hinweise für *crosslinguistic influence* (vgl. Paradis & Genesee, 1996) beobachtet. Studie 2

zeigte, dass es ein Ungleichgewicht zwischen dem Wissen der Logopäd\*innen über die spezifischen Anforderungen für die Erbringung ihrer Leistungen für zweisprachige Kinder und ihrer tatsächlichen Anwendung in der klinischen Praxis zu geben scheint, was auf ausreichendes Bewusstsein, aber immer noch eine relativ hohe Präsenz einsprachiger Ansätze im Kontext der Logopädie hindeutet. Vermutlich ist dies auf einen Mangel an verfügbaren Materialien und/oder Ressourcen zurückzuführen. Abschließend zeigte Studie 3, dass italienisch-deutsche zweisprachige Kinder unabhängig von ihrer individuellen Sprachexposition und ihren Sprachdominanzmustern bei italienisch- und deutschspezifischen sowie sprachunspezifischen Nichtwörtern gleich gut abschnitten, was die Auffassung unterstützt, dass sogar sprachspezifische NWRTs zur klinischen Beurteilung der sprachlichen Leistungen zweisprachiger Kinder mit relativ wenig Kontakt zu einer ihrer beiden Sprachen verwendet werden können. Dieses Ergebnis muss jedoch noch mit einer Gruppe von Kindern mit SES repliziert und somit bestätigt werden.

Zusammengefasst verdeutlichen diese drei Studien die Kluft zwischen wissenschaftlicher Forschung und klinischer Praxis und belegen die Relevanz der Beteiligung von Kliniker\*innen an Forschungsprojekten, um (1) den Wissenstransfer zwischen den beiden Bereichen zu verbessern und (2) evidenzbasierte Praktiken im Kontext der Logopädie fördern.

## Table of Contents

<b><u>1. INTRODUCTION</u></b> .....	<b>9</b>
<b>1.1 BILINGUALISM AND DEVELOPMENTAL LANGUAGE DISORDER</b> .....	<b>11</b>
<b>1.2 MEASURING BILINGUAL CHILDREN’S LANGUAGE EXPERIENCE</b> .....	<b>12</b>
<b>1.3 RESEARCH GOALS AND DISSERTATION OUTLINE</b> .....	<b>15</b>
<b>1.4. TERMINOLOGY USED IN THIS DISSERTATION</b> .....	<b>17</b>
<b><u>2. BILINGUAL LANGUAGE EXPERIENCE AND SPEECH PERCEPTION AND PRODUCTION</u></b> .....	<b>19</b>
<b>2.1 THE DEVELOPMENT OF LANGUAGE ATTUNEMENT</b> .....	<b>23</b>
<b>2.2 NEUROPHYSIOLOGICAL CORRELATES OF LANGUAGE ATTUNEMENT</b> .....	<b>26</b>
2.2.1 THE MISMATCH NEGATIVITY .....	26
2.2.2 AUTOMATICITY IN SPEECH SOUND PERCEPTION .....	28
2.2.3 ELECTROPHYSIOLOGY OF SPEECH SOUND DISCRIMINATION IN INFANCY AND EARLY CHILDHOOD.....	28
2.2.4 VOICE ONSET TIME .....	31
<b>2.3 THE PRESENT STUDY</b> .....	<b>33</b>
<b>2.4 RESEARCH QUESTIONS/HYPOTHESES</b> .....	<b>33</b>
<b>2.5 METHODOLOGY</b> .....	<b>35</b>
2.4.1 PARTICIPANTS .....	35
2.4.2 BEHAVIORAL MEASURES.....	36
2.4.3 MEASURES OF QUANTITY AND QUALITY OF CHILDREN’S LANGUAGE INPUT.....	39
2.4.4 ELECTROPHYSIOLOGICAL DATA .....	42
2.4.5 PROCEDURE .....	47
2.4.6 DATA ANALYSIS.....	47
<b>2.6 RESULTS</b> .....	<b>49</b>
2.5.1 DESCRIPTION OF THE SAMPLE .....	49
2.5.2 VOT PERCEPTION .....	50
2.5.3 VOT PRODUCTION .....	58
<b>2.7 DISCUSSION</b> .....	<b>63</b>
2.7.1 VOT PERCEPTION .....	63
2.7.2 VOT PRODUCTION .....	66
2.7.3 VOT PERCEPTION X PRODUCTION.....	67
2.7.4 LIMITATIONS AND FUTURE DIRECTIONS.....	68
<b>2.8 CONCLUSION</b> .....	<b>68</b>
<b><u>3. BILINGUAL LANGUAGE EXPERIENCE IN THE CONTEXT OF SPEECH AND LANGUAGE PATHOLOGY</u></b> .....	<b>70</b>
<b>3.1. SPEECH AND LANGUAGE PATHOLOGY IN BILINGUAL CONTEXTS</b> .....	<b>71</b>
3.1.1. LACK OF LANGUAGE PROFICIENCY BEYOND THE SOCIETAL LANGUAGE.....	73
3.1.2. LACK OF SATISFACTORY DIAGNOSTIC TOOLS AVAILABLE FOR THE BILINGUAL POPULATION.....	74
3.1.3. SLPs’ PREPAREDNESS AND CONFIDENCE WHEN ASSESSING AND TREATING BILINGUAL CHILDREN .....	76
<b>3.2. THE PRESENT STUDY</b> .....	<b>77</b>

<b>3.3 RESEARCH QUESTIONS/HYPOTHESES</b> .....	<b>77</b>
<b>3.4. METHODOLOGY</b> .....	<b>78</b>
3.4.1. PARTICIPANTS .....	78
3.4.2. QUESTIONNAIRE DESIGN .....	78
3.4.3. PROCEDURE .....	78
3.4.4. DATA ANALYSIS.....	79
<b>3.5 RESULTS</b> .....	<b>79</b>
3.5.1 SLPs’ CHARACTERISTICS .....	79
3.5.2. SLPs’ ATTITUDES AND BELIEFS ABOUT THE ASSESSMENT OF BILINGUAL CHILDREN ...	81
3.5.3. SLPs’ CLINICAL PRACTICES IN THE LANGUAGE ASSESSMENT OF BILINGUAL CHILDREN	84
3.5.4 EFFECTS OF SLPs’ PRIOR EXPERIENCE.....	87
3.5.5 EFFECTS OF SLPs’ PERSONAL LANGUAGE BACKGROUND.....	88
3.5.6 EFFECTS OF SLPs’ ATTITUDES TOWARDS CHILDREN’S HERITAGE LANGUAGE .....	89
<b>3.6 DISCUSSION</b> .....	<b>91</b>
3.6.1. EXPERIENCE RELATED EFFECTS .....	92
3.6.3. SLPs’ LANGUAGE BACKGROUND RELATED EFFECTS .....	93
3.6.4. DISCREPANCIES BETWEEN KNOWLEDGE AND PRACTICE .....	93
3.6.5. LIMITATIONS AND FUTURE PERSPECTIVES .....	93
<b>3.7 CONCLUSIONS</b> .....	<b>94</b>
<b><u>4. BILINGUAL LANGUAGE EXPERIENCE AND NONWORD REPETITION</u></b> .....	<b><u>95</u></b>
<b>4.1. NONWORD CHARACTERISTICS AND CHILDREN’S REPETITION PERFORMANCE</b> .....	<b>96</b>
<b>4.2 LANGUAGE-SPECIFIC NONWORD REPETITION TASKS</b> .....	<b>99</b>
<b>4.3 NONWORD REPETITION IN TWO LANGUAGES</b> .....	<b>101</b>
<b>4.4 LANGUAGE-UNSPECIFIC NONWORD REPETITION TASKS</b> .....	<b>105</b>
<b>4.5 RESEARCH QUESTIONS/HYPOTHESES</b> .....	<b>107</b>
<b>4.6 METHODOLOGY</b> .....	<b>109</b>
4.6.1 PARTICIPANTS .....	109
4.6.2 ASSESSMENT TOOLS .....	109
4.6.3 PROCEDURE .....	116
4.6.4 DATA ANALYSIS.....	116
<b>4.8 RESULTS</b> .....	<b>117</b>
4.8.1 DESCRIPTION OF THE SAMPLE .....	117
4.8.2 CHILDREN’S NONWORD REPETITION PERFORMANCE .....	119
<b>4.9 DISCUSSION</b> .....	<b>120</b>
4.9.1 NONWORD CHARACTERISTICS AND CHILDREN’S REPETITION ACCURACY .....	120
4.9.2 LANGUAGE EXPERIENCE AND CHILDREN’S REPETITION ACCURACY.....	121
4.9.1 LIMITATIONS AND FURTHER DIRECTIONS.....	122
<b>4.10 CONCLUSIONS</b> .....	<b>123</b>
<b><u>5.GENERAL DISCUSSION</u></b> .....	<b><u>124</u></b>
<b>5.1 BILINGUAL CHILDREN’S PRODUCTION AND PERCEPTION OF VOT</b> .....	<b>126</b>
<b>5.2 ASSESSING BILINGUAL LANGUAGE DEVELOPMENT IN CLINICAL PRACTICE</b> .....	<b>129</b>
<b>5.3 APPROACHING THE DIAGNOSTIC DILEMMA</b> .....	<b>130</b>
<b>5.4 PRACTICAL IMPLICATIONS</b> .....	<b>131</b>
<b><u>REFERENCES</u></b> .....	<b><u>133</u></b>

**APPENDIX A ..... 155**

**APPENDIX B..... 157**



# 1. Introduction

Bilingualism describes the circumstance in which a person is confronted with more than one language in their everyday environment (Grosjean, 1982). A heritage language (also referred to as minority, family, home, or first language; L1) is the language that a person uses or a child acquires at home, while the societal language (also referred to as majority, community, or second language; L2) is the language spoken by the majority of the society they live in. While the typical course of monolingual language development has been described in great detail, bilingual language development has, until rather recently, received far less attention. Since the world has been increasingly recognizing bilingualism as a common phenomenon in recent years (Grosjean, 2015) and it is highly likely that today, in the 21<sup>st</sup> century, more children grow up being exposed to two (or even more) languages than children who acquire only a single language, the topic of bilingual language development is of great importance. Bilingual language development in comparison to monolingual language acquisition is far more heterogeneous as each bilingual child acquires their languages under diverse conditions, with highly individual language contact patterns, and, thus, with highly variable expressions of their language proficiency (Carroll, 2017). Therefore, the situations, settings, and pathways that lead children to becoming bilingual are characterized by a high degree of inter- and intra-individual variety. Generally, a global distinction can be made between children who grow up with two (or more) languages from birth (i.e., simultaneous bilinguals who acquire their two languages in parallel; this scenario has also been termed bilingual first language acquisition; De Houwer, 1990) and children who start off being exposed to only a single language at birth, but later on (before the age of six) are confronted with a second language (i.e., sequential bilinguals, this scenario has also been termed early second language acquisition; De Houwer, 1990). However, this distinction between the two groups that is based on children's age at the onset of their acquisition of the societal language is artificial in nature and there is no clear-cut boundary between them. Both simultaneous and sequential bilinguals acquire two languages in early childhood (defined as the period between birth and six years; Serratrice, 2019). However, their respective acquisition scenarios may differ to a certain extent. Simultaneous bilinguals acquire two first languages in parallel and are usually exposed to both of them in their homes. Overall, the same developmental mechanisms that apply to monolingual children's first language acquisition apply to them as well. Even so, simultaneous bilingual children differ from their monolingual peers in terms of language production and comprehension (Serratrice, 2013). In contrast, sequential bilinguals, are exposed to an L2 once they have already acquired their L1 to a certain extent and they are very likely hear and use the two languages in quite different

contexts. While young sequential bilingual children's L1 is usually the language used in their homes (i.e. heritage language; see Section 1.4 for more details on terminology), they typically first start being immersed in their L2 (usually the societal language) once they enter some form of formal childcare settings.

In both simultaneous and sequential bilingual children, uneven development has been commonly observed (De Houwer, 2018). Uneven development refers to unequal performance and/or progress in the development of children's two languages, which can be at a similar level (i.e., balanced), but more often linguistic abilities in one of the languages may be better developed than the other (cf. language dominance) (Crescentini, Marini, & Fabbro, 2012). However, the concept of language dominance is not static, and it is influenced by various factors throughout the course of a child's language development. Initially one language may develop faster than the other but then the other language may take over. In some cases, children's abilities in one of the two languages (commonly the heritage language) may wane completely. Whether and to what extent this so-called language attrition of children's skills in the heritage language will occur is dependent on how families deal with their two languages at home (MacLeod, Fabiano-Smith, Boegner-Pagé, & Fontollet, 2013). In particular, the amount of exposure to the heritage language inside the home has been found to be linked to children's overall proficiency in the heritage language (Dijkstra, Kuiken, Jorna, & Klinkenberg, 2016; Gutiérrez-Clellen & Kreiter, 2003).

Bilingual children's language performance in either language alone should not be expected to be comparable to that of their monolingual peers. Many aspects of bilingualism influence language performance, such as (1) the onset, duration, quantity, and quality of exposure, (2) cross-linguistic influence specific to each language combination, and (3) dynamic and context-specific language use. This fluid and complex interplay of factors eventually results in great variability in language performance among bilingual children. Independent of the particular pattern of language exposure, children who acquire two languages are not able to dedicate the same amount of time to each of them compared to monolinguals because their waking hours and thus their opportunities to experience one or the other language are split between their heritage and the societal language (Oller, Pearson, & Cobo-Lewis, 2007), which is particularly important in the context of the identification of Developmental Language Disorder (DLD) in the bilingual population.

## 1.1 Bilingualism and Developmental Language Disorder

DLD is a condition in which, according to the definition of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR text rev.; American Psychiatric Association, 2022), a child's language performance significantly differs from the performance of the majority of their peers without any underlying biomedical explanation. Different linguistic areas like phonology, syntax, lexicon, and semantics can be affected by DLD to varying degrees (Bishop, Snowling, Thompson, & Greenhalgh, 2017). As opposed to previous, narrower, definitions of DLD as a disorder that is reflective of a child's intrinsic difficulty to acquire language (e.g., Leonard, 2000), the CATALISE Consensus group (Bishop et al., 2017) recently agreed that the presence of neurobiological and/or environmental risk factors does not rule out a diagnosis of DLD. The potential risk factors that are most frequently associated with DLD include, among others, a family history of language impairment, male gender, and a low level of parental education and/or socioeconomic status (SES; Arrhenius, et al., 2018; Boivin, Kakooza, Warf, Davidson, & Grigorenko, 2015; Sansavini et al., 2021).

Studies comparing the language abilities of bilingual children with DLD to those of monolingual children with DLD have found that simultaneous bilingual children with DLD exhibit error patterns that are similar to those made by monolingual DLD children of the respective language (Armon-Lotem, de Jong, & Meir, 2015). However, typically-developing (TD) bilingual children with limited exposure (especially sequential bilinguals) also often resemble the linguistic characteristics of monolingual children with DLD (Paradis, 2010). In the past, these disparities between monolingual and bilingual language acquisition had led to the assumption that bilingualism poses a disadvantage on language development and may play a role in the genesis of DLD (Genesee, 2009). Research, however, indicates that in general, DLD affects monolingual and bilingual children alike (prevalence of 7%; Tomblin et al., 1997; Kohnert, Ebert, & Pham, 2020). Even so, the great variability and inherent complexity within and across bilingual individuals with respect to their patterns of language exposure, use, and proficiency present a challenge for clinicians (i.e., Speech and Language Pathologists; SLPs). It is often very difficult to ascertain whether a bilingual child, whose development is delayed compared to a monolingual peer, has DLD or whether the delay results from insufficient exposure. An additional challenge is that assessment tools are available for only a small number of languages, even for monolingual assessment. Moreover, these existing tools are not necessarily comparable across languages and few of them have been normed for the bilingual population. Until typical bilingual language development is better understood, along with the many factors that impact it, language assessments of bilingual children will have poor validity

and will often lead to erroneous clinical and educational decisions (Thordardottir, 2011). Both, underidentification, (i.e., false negative for DLD) and overidentification (i.e., false positive for DLD) of language impairment in bilingual children increases children's risk for social disadvantages or stigmatization, respectively (Scharff Rethfeldt, 2019). For instance, children who have are underidentified (i.e., false negatives) will not receive services to remediate their language deficits, whereas those who are over-identified may not be offered certain educational and advanced classes in school. In order to approach this diagnostic dilemma and reduce inaccuracies in the diagnosis of DLD in bilinguals, many efforts have been made to measure and, thus, quantify bilingual children's language experience.

## 1.2 Measuring Bilingual Children's Language Experience

Children acquire language through exposure, and a large amount of this exposure is usually provided by the parents (Hart & Risley, 1995). Even in monolinguals, the quantity of language input can be fairly variable (Hart & Risley, 1995; Hoff, 2006). The input of bilingual children may, moreover, differ from monolingual children's input in qualitative aspects. Operationalizing and documenting the language experience of bilingual children in terms of both quantity and quality is difficult due to the diverse and multifaceted ways in which they can become bilingual. Nevertheless, it is of theoretical importance in order to promote our understanding of the general mechanisms involved in (bilingual) language development. Additionally, it is valuable to the identification of DLD in the bilingual population. Measuring a bilingual child's particular language exposure and input can help us understand whether any indication of delayed language acquisition can be traced back to limited exposure or whether it is, in fact, indicative of DLD (Armon-Lotem & Meir, 2019).

Several measures have been used to quantify bilingual children's language experience (see Kaščelan et al., 2021 for an overview). Two rather basic measures of bilinguals' language experience are (1) Age of Onset (AoO) of exposure and (2) Length of Exposure (LoE). In sequential bilinguals, AoO of a second language marks the end of monolingual language acquisition. Simultaneous bilinguals, however, do not experience a period of monolingual language acquisition as they are exposed to both languages from birth. Their AoO is thus identical for both of their languages. Additionally, LoE measures the time since children's first exposure to a language. LoE is calculated by subtracting children's AoO from their chronological age (e.g., Paradis, 2011). The two measures therefore correlate with each other but neither AoO nor LoE take into account the distribution of children's exposure to the two

languages in their everyday lives nor do they express any variation in the amount of exposure to either of the two languages throughout the course of children's development.

In addition to AoO and LoE, measures of relative and absolute exposure can provide more precise insight into children's individual patterns of bilingual language experience. Measures of relative language exposure provide an approximation of children's language experience with one language relative to the other. A relative measure is a strictly quantitative measure (usually expressed in percentages) that estimates the time children hear (i.e., language input) and/or use (i.e., language output) a certain language. Children's relative exposure is typically assessed through parental questionnaires in which parents (and sometimes also teachers e.g., Gutiérrez-Clellen & Kreiter, 2003) are asked to provide details about their children's daily routines and the language(s) used for each of their activities. Relative exposure is often operationalized as "frequency of exposure", typically assessed in questionnaires through questions asking about the overall exposure to a particular language or more specifically concerning language exposure with respect to a particular type of language experience, such as reading (Kaščelan et al., 2021). Measures of relative exposure have been found to be good predictors of children's lexical and phonological knowledge. For instance, when examining the relative exposure to English and Spanish in five-year-old bilingual Spanish-English children, Ruiz-Felter, Cooperson, Bedore, and Peña, (2016) found that children's current proportion of use of each language predicted their phonological production accuracy of both vowels and consonants in the respective language and similar effects of relative exposure were found on bilingual children's vocabulary development (Dijkstra et al., 2016).

Further components of relative language exposure have been proposed. For instance, Unsworth (2013) argued that even within one and the same child the amount of language exposure may vary considerably across time and proposed the notion of cumulative length of exposure. This composite measure includes a child's current and past exposure summed over time. Unsworth tested 136 simultaneous bilingual English-Dutch children in the Netherlands. For each of her participants chronological age, AoO, and LoE for English and Dutch were therefore identical. Unsworth (2013) found that current exposure and cumulative length of exposure were good predictors of children's gender markings on Dutch determiners. Furthermore, the differences between the groups in a gender marking task that were observed when participants were matched solely based on their chronological age, disappeared when bilingual English-Dutch child participants and monolingual Dutch participants were matched based on bilinguals' cumulative experience with Dutch.

As previously stated, the measures of relative exposure described above usually rely on parental reports that have been argued to not necessarily align with children's actual everyday language experience. For example, Nakamura (2015) found that, while a mother of Thai descent reported solely speaking Japanese to her child, observations of interactions between this mother and her child revealed that, in fact, 13.5% of all the utterances she addressed to the child were in Thai. Measures of relative exposure do not provide insights into the particular linguistic content children receive, whereas measures of absolute exposure that are assessed through language samples of natural interactions between a child and a communication partner can quantify the specific characteristics of this child's language input in one language regardless of the other or in both languages. Such characteristics can include the presence and frequency of particular phonemes, lexical items or morphosyntactic structures among others. Marchman, Martínez, Hurtado, Grüter, and Fernald (2017) compared the effect of relative and absolute language exposure on bilingual Spanish-English children's language use. During interviews with the caregivers, most families reported a predominant Spanish exposure (80%). This rather global indirect measure was not representative of the variability of language use within these families and the actual number of words the children on average heard in each language within one hour. These latter measures were obtained via recordings of naturalistic interactions between each child and their caregivers. The two measures of language exposure were moderately correlated. However, the observed number of words directed at a child (i.e., an absolute measure) was more predictive of their performance on a standardized language test than measures obtained through parental report (i.e., a relative measure).

Despite absolute measures of language exposure being considered most objective when quantifying children's bilingual language experience, they do also have their downfalls. First, due to limited resources and research capacities it is not always possible to collect the appropriate data. Second, depending on how and where the data are collected, recordings of everyday interactions and conversations within the family home may be considered intrusive on the privacy of the participants and some families might be hesitant to give their consent to such procedures. Lastly, recordings obtained within a lab context may not necessarily be representative of children's everyday language experience and parents might feel pressured to behave in a certain way during the recordings. The most valid procedure to estimate bilingual children's language exposure is therefore still a matter of debate and relative measures are still considered useful to obtain.

In addition to the quantification of bilingual language input, some authors have further proposed that qualitative aspects of bilingual children's language experience characterize and

influence their language development. Place and Hoff (2016) found bilingual Spanish-English children's language experience to be approximately equally divided between Spanish and English using a relative measure. The authors noted, however, that 94% of children's exposure to the heritage language were provided by native speakers of Spanish, whereas only 40% of exposure to the societal language were provided by native speakers of English as assessed via self-reports through a parental background questionnaire. Therefore, despite equal exposure to both languages, the degree of native-ness of exposure might not be the same for both languages. While Place and Hoff (2016) suggested that nonnative input in the societal language may be less supportive to its development, Nakamura (2015) found that the occasional morphosyntactic errors in Japanese produced by a Thai descent mother were not reproduced by her child.

Even more general aspects of bilingualism, that go beyond the quantification or the qualitative description of language exposure (e.g. children's ethnolinguistic identities and the attitudes towards their heritage and the societal language) may be influential in bilingual language acquisition (e.g., Armon-Lotem, Joffe, Abutbul-Oz, Altman, & Walters, 2014). Armon-Lotem and colleagues compared the L2 proficiency (vocabulary, listening comprehension, sentence repetition, sentence production, narratives, and pronunciation) of four-to-six-year-old sequential Russian-Hebrew and English-Hebrew children in Israel, where Hebrew is the societal language. They found that overall the group of Russian-Hebrew children, who lived in communities where there was a lot of mixed use of the societal and the heritage language, showed better performance in the societal language. By contrast, in the communities of the English-Hebrew participants, motivation for integration in the host society and thus the use of the societal language was lower. The authors pointed out that despite Russian being well supported within Russian-Hebrew communities, its reputation in the Israeli society is relatively low compared to English that is considered highly important for international communication. Armon-Lotem and colleagues, thus, argued that the sense of prestige associated with a certain language can mediate the effect of relative exposure.

### 1.3 Research Goals and Dissertation Outline

The present dissertation is concerned with the influence of language experience (measures of language input and output) on simultaneous bilingual TD children's language knowledge and performance in their two languages and how bilingual experience relates to the clinical practice in Speech and Language Pathology services where this information is particularly relevant for distinguishing between the typical variance in language proficiency as a function of reduced

exposure (when compared to monolingual children) and DLD. The three experimental chapters that constitute this dissertation and their main research questions are summarized below.

**Chapter 2, *Bilingual Language Experience and Speech Perception and Production*** provides the basis of this dissertation and reports the impact that specific language experience poses on children's language development. The chapter is concerned with children's pre-attentive perception and production of language-specific phonemic cues that phonetically differ between their two languages. More specifically, I tested whether and how bilingual Italian-German children living in Germany perceive and produce voicing contrasts in bilabial stop consonants. In this electroencephalography (EEG) study, I explored how children's relative amount of current exposure to their heritage and the societal language influenced neurophysiological correlates of language processing (indexing automatic speech sound discrimination) and production. Furthermore, I posed the question of whether bilingual children differ from their monolingual peers in their development of selective perceptual routines and, thus, their automaticity in speech sound processing.

**Chapter 3, *Bilingual Language Experience in the Context of Speech and Language Pathology*** is concerned with Speech and Language Pathologists' attitudes and approaches towards bilingualism. While this study is primarily concerned with clinicians' general beliefs about bilingualism, several questions target how Speech and Language Pathologists assess bilingual children, and whether and how they take language exposure into account. In particular, the study is embedded in the German context, where some (standardized) assessment tools are already available for the identification of DLD in the bilingual population. However, their applicability and usefulness as rated by the study participants is only moderately satisfactory.

**Chapter 4, *Bilingual Language Experience and Nonword Repetition*** is concerned with the creation, implementation and validation of a novel Nonword Repetition Task, a screening method that has been proposed to be less language experience-dependent, and thus particularly suitable for the bilingual population. The precise impact of language experience on children's nonword repetition performance is still a matter of debate (Chiat, 2015). The task used in this study constitutes language-specific (Italian and German) and language-unspecific (i.e., language neutral) stimuli. I examined whether the nonword repetition performance of bilingual Italian-German TD children differed as a function of their current relative language exposure (i.e., their language input and output in each of their two languages) and whether there was a difference between their performance on language-specific compared to the more neutral nonwords.



**Chapter 5**, the *Discussion*, connects the findings of the chapters 2 to 4, summarizes and discusses the main findings of this dissertation and their theoretical and practical relevance to the field of bilingualism research.

**Chapter 6**, the *Conclusion*, closes this dissertation and provides an overview of possible avenues for future research to further advance the field of bilingual language acquisition in early childhood. Furthermore, practical implications are presented that are suggested to improve clinical processes and policies when dealing with bilingual children.

#### 1.4. Terminology used in this dissertation

**Heritage language (speaker).** Simultaneous bilinguals differ in terms of the language status of their socio-cultural environment. In some communities (countries or cities; e.g. Barcelona, Montreal, Luxembourg), children are not only exposed to two languages in their home but also in their broader social environment. Other simultaneous bilinguals grow up in a monolingual socio-cultural environment where the language of the society does not match the language provided by one or both of their parents. To capture the differences between those who are immersed in a bilingual socio-cultural environment and those who acquire a minority language in a majority language context, the term heritage language (HL) speaker has been introduced (e.g., Valdés, 1995; Kupisch & Rothman, 2018). All studies that constitute this dissertation have been conducted in Germany, an exclusively monolingual socio-cultural environment, where German is the one official and majority language. Therefore, for the entirety of this dissertation, the term heritage language (speakers) will be used to refer to bilingual Italian-German children.

**Language input quantity.** In the context of this thesis, the concept of input quantity is operationalized as bilinguals' proportional amount of language exposure to each of their two languages. For two of the three studies contributing to this dissertation (discussed in chapter 2 and chapter 4), children's current relative language exposure was assessed and measured by means of a parental questionnaire to gain an objective estimate of the average proportion of the time participants heard (input) and spoke (output) their heritage language (Italian) compared to the societal language (German) during a typical week of their everyday lives.

**Language dominance versus language balance.** The term language dominance has been used to refer to a person's degree of bilingualism, that is, their proficiency in one language relative to the other (Hernández-Chávez, Burt, & Dulay, 1978). A person who exhibits equal abilities

on their two languages is considered a balanced bilingual. Along the same lines, the terms dominant versus weak language will be used to distinguish between children's languages. It is common in bilinguals that one language is more dominant compared to the other (Genesee, 2009), although it is acknowledged that proficiency is often context- and content-specific (Kohnert, 2008).

**Native-like/monolingual-like.** These two terms are used interchangeably, whenever the acquisition process or outcome of bilingual children is compared to that of their monolingual peers. Note that this is not to suggest that monolingual-like performance is the ultimate target in bilingual language acquisition.

**Cross-linguistic influence.** It has largely been accepted that children who simultaneously acquire two languages from birth, or soon after, can acquire them as two independent but not necessarily autonomous systems that can interact and influence each other (Serratrice, 2013). Paradis and Genesee (1996) suggested three potential manifestations of cross-linguistic influence (i.e., transfer, acceleration, and delay) that may explain differences in the patterns and rate of language development between bilingual and monolingual children. In chapter 2 of this dissertation, aspects of cross-linguistic influence in combination with measures of language exposure on the phonological acquisition of Voice Onset Time in bilingual Italian-German children will be discussed.

When, by using this terminology, I report differences between monolingual and bilingual children in describing their current status of language development, in no way is this supposed to imply any form of incompleteness of acquisition, clinical delay in language development, or superiority of monolingual-like performance. The collection of studies presented in this thesis acknowledges that bilingual language performance exists on a continuum and is constantly in flux and thus does not have an ultimate endpoint as it is subjected to change throughout the entire course of a person's lifetime. However, especially in the clinical context, deviance from typical language acquisition that can be classified as language impairment needs to be detected as early as possible and approached accordingly. Considering that most language assessment tools available in clinical practice only exist in the societal language, the comparison between bilingual and monolingual children is necessary in order to argue for the necessity of the development of testing procedures specifically targeting the bilingual population and/or the relevance of providing bilingual norm data for monolingual assessment tools.

## 2. Bilingual Language Experience and Speech Perception and Production

Speech perception is a crucial prerequisite for language comprehension that entails (1) the detection of an acoustic signal as mechanical vibrations on the tympanic membrane in the ear and (2) the mapping this signal onto abstract representations in the brain through neurophysiological processes that underlie this perceptual analysis (Giraud & Poeppel, 2012). In order to be able to extract meaning from the acoustic signal that is speech, it is, thus, crucial to quickly and efficiently match highly variable acoustic cues with cortical representations of discrete phonemic categories. However, the acoustic signal, even for one and the same phonological unit, can vary greatly on many dimensions such as talker, speech rate, and phonetic context (Kuhl, 2004). Thus, it is difficult to rely on absolute acoustic values to determine the target of a particular speech sound.

Researchers agree that speech perception is not achieved by extracting invariant, static cues from the acoustic signal and seventy years of research have revealed that listeners perceive speech categorically (e.g., Pisoni & Lazarus, 1974). Categorical perception is the phenomenon whereby a listener groups variable acoustic patterns into one percept, also called a phoneme. A phoneme is defined as the smallest unit in speech that distinguishes one word from another. Phonemic category boundaries divide an acoustic continuum (e.g., from a voiced stop consonant to its voiceless counterpart) into qualitatively discrete regions. Adult listeners have been found to show good discrimination<sup>1</sup> across category boundaries specific to their native language, but not within categories of that language even when the acoustic difference suggests an equivalent degree of difference. However, there remains considerable disagreement on whether the underlying targets are motor, gestural, or auditory in nature (Diehl, Lotto, & Holt, 2004). Irrespective of this debate, acquiring a native language leads to the unique and language-specific configuration of a listener's phonological system.

The speech sounds of all the world's languages total about 600 consonants and 200 vowels (Ladefoged & Disner, 2012). However, individual languages employ a much smaller but unique sub-set of these phonetic elements and can, moreover, differ in the exact location of a phonemic category's boundary. Experience with the phonetic and phonological characteristics of a particular language has been found to exert a profound effect on an

---

<sup>1</sup> Note that the term "discrimination" can be at the perceptual level, but it can also be used to describe pre-conscious mechanisms. In order to specify which one I am referring to, the term "behavioral discrimination" will be used to indicate conscious discrimination. By contrast, the terms "neural discrimination" or "pre-attentive discrimination" will be used when I am referring to the underlying neural processes necessary to distinguish sensory input.

individual's perceptual abilities and can introduce constraints on the identification and behavioral discrimination of certain speech sounds (e.g., Jusczyk, 1997). Adults have been found to have greater difficulty perceiving foreign language contrasts as their perception relies on the category structure of their native language. For instance, while adult English listeners had no problem categorizing and behaviorally discriminating tokens from an /l/-/r/ continuum into the two respective distinct categories, Japanese-speaking adults demonstrated poor categorization and behavioral discrimination of the same sounds as the contrast is non-existent in their native language (Goto, 1971; MacKain, Best, & Strange, 1981). Furthermore, adult native speakers of American-English have shown difficulties in discriminating the cross-linguistically rare retroflex place of articulation of Hindi stop consonants that is not part of English phonology (Hayes-Harb & Barrios, 2022). Similarly, native speakers of Hindi have been found to show difficulties in accurately categorizing American-English /v/ versus /w/ even in Hindi L2 speakers of English who had reported high proficiency in English (Grover, Shafer, Campanelli, Whalen, & Levy, 2021).

Difficulties in distinguishing between the phonetic contrasts of foreign languages have been explained by various models. One model claims an effect of 'interference' by the native language (e.g., Kuhl, 2000). This model has been extended to suggest that extensive experience with a language causes neural commitment (i.e., dedicated neural networks that code the patterns of native-language speech) to the acoustic properties of that language, and that consequently, processing another language becomes more difficult if its acoustic properties and patterns do not conform to the initially learned one (Iverson et al., 2003). The concept of neural commitment is tightly linked to the notion of a "critical/sensitive period" for language acquisition (Lenneberg, 1967). The idea of the critical period model in L2 acquisition/learning is that there is a biologically based maturational cut-off that prevents older learners from achieving native-like performance in the L2. However, Knudsen (2004) pointed out that even after a sensitive period has ended a variety of mechanisms continue to operate and thereby support plasticity. Thus, even mature systems can modify their processing patterns. Similar to the interference model proposed by Kuhl and colleagues (Kuhl, 2000; Iverson et al., 2003), the Automatic Selective Perception (ASP) model suggests that interference from the L1 modulates L2 processing, but focuses more directly on the processes activated in speech perception (Strange, 2011; for more details see Section 2.2.2). According to the ASP model, listeners become highly automatic and skilled in quickly perceiving the speech sounds of the language they are exposed to from birth. Therefore, it becomes increasingly difficult to overcome these "overlearned" perception skills later on in life.

Considerable research shows that (monolingual) infants rely on (1) distributional patterns in language input to assist them with language learning at the phonetic level (Maye, Werker, & Gerken, 2002) and (2) tracking adjacent transitional probabilities in continuous speech to be able to parse speech (Saffran, 2002) to then acquire language-specific phonetic patterns. Language-specific phonetic perception is crucial for successful language acquisition because it promotes the detection of phonotactic patterns, which supports children in segmenting the continuous speech stream into words, that children will eventually need to associate with meaning (Tsao, Liu, & Kuhl, 2004).

Infants exposed to more than one language from birth need to learn and differentiate between two potentially opposing phonological systems in order to identify which phonological system is in effect and to further identify meaningful units in the continuous speech stream. It is likely that similar mechanisms are activated for both mono- and bilingual phonological learning. However, the possibility of conflicting phonological and phonetic patterns in the two languages indicates that some aspects of bilingual development must be different.

With regards to speech production, adults who learn an L2 after about 13 years of age have been found to usually speak with a detectable foreign accent (Flege, Munro, & MacKay, 1995) which, partially, stems from the non-native phonetic realization of consonant and vowels across different languages. Flege (1995) proposed the Speech Learning Model (SLM) to account for L2 speech production differences between native speakers and adults who are late learners of the same language. To date, no theoretical models have been proposed to specifically address the speech production of young simultaneous bilinguals. However, Flege's SLM may be extended even to early learners. The SLM proposes that it is a person's perceptual abilities that constrain their accurate production of foreign speech sounds; that is, without accurate perceptual targets to guide sensorimotor learning, the production of L2 sounds will align with the properties of the representation of the phonemic L1 categories, and thus result in inaccurate pronunciation. The model allows the possibility that perceptual constraints may also be governed by the degree of phonetic dissimilarity between an L2 sound, and the closest L1 sound and that sufficient naturalistic exposure will eventually lead to the formation of new distinct phonetic representations in the L2. Further, in addition to perceptual limitations, the SLM proposes that motoric output constraints may also play a role in L2 speech sound production. Evidence supporting the claim that perception drives production stems from perception trainings studies with adults (for an overview see Sakai & Moorman, 2018) that have shown that production of a novel sound improved after perceptual training (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Wang, Jongman, & Sereno, 2003).

In bilingual children who are acquiring two languages from birth, it is unclear to what extent experience-related constraints on speech sound production may come into effect as these children are in the course of setting up the phonemic categories for their two languages simultaneously. Further, it is unclear to what extent perception leads to improved production in the course of simultaneous bilingual development of young children. Flege (1987) postulated that the formation of a particular phonological category may be disrupted if there is a certain extent of overlap between an L1 and an L2 sound. With regards to simultaneous bilingual language development it is possible that a child acquires only one category for two sounds and that they perceive them to be alike. In recent years, several studies have compared the development of phonetic/phonological production skills of monolingual and bilingual children to examine how the two phonological systems of two languages that may differ to a certain extent develop and whether they influence one another (cf. crosslinguistic interaction; Paradis & Genesee, 1996). Fabiano-Smith and Barlow (2010), for instance, reported transfer (i.e., the incorporation of one phonetic unit from one language into the other) in the phonemic inventories of three-to-four-year-old bilingual Spanish-English children. Goldstein and Swasey Washington (2001) found evidence of delay in four-year-old Spanish-English bilinguals who were less accurate in producing spirants, flaps, and trills when compared to monolingual Spanish peers. To my knowledge, there is no evidence for acceleration in the acquisition of segmental phonemes. However, Tamburelli, Sanoudaki, Jones, and Sowinska (2015) reported that exposure to Polish (a language with complex word initial consonant clusters) accelerated the development of English consonant clusters in bilingual Polish-English seven- and eight-year-olds. The English clusters are less phonologically complex than those found in Polish.

The Articulatory-Map model (A-Map model; McAllister Byun, Inkelas, & Rose, 2016), which focuses on monolingual speech production development, may also be extendable to bilingual development. In addition to Flege's SLM (1995) it is important to consider this model as it emphasizes the role of children's language output, that is how often they have the opportunity to use and, thus, practice, their language(s). The A-Map accounts for differences between child productions and the respective adult-like target through anatomical and motor control differences between the two groups. The model suggests that children's speech productions are subject to two main constraints that interact with each other. The first constraint is *accuracy*, that is the pressure to match the adult-like target, while the second constraint is *precision*, that is the pressure to produce stable and well-practiced realizations even if they do not fully match the adult target. However, over time and through production experience (and a dynamic interplay between articulator movements and acoustic feedback) children's speech

productions become more precise and approach the adult target. In an extension of the A-Map model to the bilingual population, bilinguals may take longer to match the adult-like target for certain speech sounds if these sounds only occur in one of their languages. This claim is made because bilinguals have reduced opportunities to speak a given language, and therefore to practice speech sounds in that language. The A-Map model, thus, highlights the relevance of bilingual children's active language usage and the quantity of their language output for their phonological development.

It is an important question how bilingual children's language experience (i.e., language input and output), neural commitment (i.e., language attunement; see Section 2.1) and the developmental trajectory of speech sound production abilities interact. To approach this issue, I will first provide a detailed overview of the development of language attunement and its neurophysiological correlates before I proceed to the description of the phonemic feature of interest in this study, Voice Onset Time.

## 2.1 The Development of Language Attunement

Well before children are born, their linguistic environment shapes the perceptual foundation for language learning (e.g., Moon, Cooper, & Fifer, 1993). Studies on newborns who were exposed in-utero to two languages from different rhythmic classes have shown that infants language preferences are modulated by this early experience when tested soon after birth. For instance when comparing zero-to-five-day-old infants from monolingual English families and newborns of bilingual English-Tagalog-speaking mothers, using a high-amplitude preference-sucking technique, Byers-Heinlein, Burns, and Werker (2010) found that the monolingually-exposed infants showed a significant preference for English stimuli, whereas the bilingually-exposed newborns were equally interested in Tagalog and English stimuli. Furthermore, the authors showed that both monolingual and bilingual newborns could reliably exploit rhythmic and prosodic features to neurally discriminate between the two languages.

Less is known, however, about how bilingual experience shapes children's phonemic development very early on. Behavioral studies on monolingual infants have shown that children are born as "universal listeners" who have the ability to acoustically discriminate between the phonetic units of any given language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). However, this ability to perceive differences between speech sounds has been shown to soon be subjected to extensive change (Werker & Tees, 2002). At the time of birth, newborns are equipped with the perceptual prerequisites to acquire any given language but during the second half of their first year of life, through language experience, speech perception abilities are narrowed down

to the ambient language (Jusczyk, 1997). By that time, children's sensitivity to discriminate non-native speech sound contrasts (i.e., phonetic units that do not represent phonemic contrasts in a child's input language) diminishes. This process of perceptual reorganization, and with it the gradual modification of children's discriminatory abilities, has been described for both vowels and consonants. This transition from language-general to language-specific perception is observed between 9 and 12 months of age for consonants (Werker & Tees, 1983) with effects for vowels appearing earlier (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). However, Best, Goldstein, Nam, and Tyler (2016) pointed out that some speech contrasts were more difficult to perceive than others and argued that multimodal speech perception, involving the detection of articulatory information, may account for the maintenance of not only native but also some non-native speech contrasts.

Mixed findings have been reported for whether bilingual infants follow the same developmental trajectory as their monolingual peers with regard to their phonetic neural discrimination abilities. A study by Bosch and Sebastián-Gallés (2003) found that Spanish-Catalan bilingual children were going through a U-shaped learning curve. At four months of age all participating infants (monolingual Spanish, monolingual Catalan, and bilingual Catalan-Spanish) were sensitive to a Catalan vowel contrast (/e/ vs. /ɛ/). At eight months of age, Catalan-exposed children had maintained this sensitivity while Spanish-exposed children had not. Surprisingly, also in the bilingual Catalan-Spanish group, the sensitivity to the Catalan contrast temporarily declined at 8 months of age but recovered four months later. The authors attributed this developmental difference between monolingual and bilingual children to the higher processing demands that are linked with the complexity and variability in the input when dealing with two languages. Other studies that investigated bilingual children's phonetic development throughout the course of the first year of life, however, have found no such differences between monolingual and bilingual infants. Sundara, Polka, and Molnar (2008), for instance, compared infants' neural discrimination abilities of two place variants of /d/ (French dental /d/ vs. English alveolar /d/). At six to eight months all participants (monolingual French, monolingual English, bilingual French-English) succeeded in discriminating the consonant contrast, whereas at ten to twelve months monolingual English and bilingual French-English but not monolingual French infants were able to keep them apart. These discrepant results may partially be explained by methodological differences between the studies. Especially variation in the language combinations as well as variation across the phonemes under investigation need to be taken into account when interpreting these findings.



One explanation for why Bosch and Sebastián-Gallés (2003) found a delay in bilingual infants formation of language-specific, native vowel categories may be linked to the relative frequencies the phonemes occur in the languages. The Catalan /ɛ/, for instance, occurs a lot less frequently than the Spanish and Catalan /e/, whereas the consonant /d/ used in Sundara, Polka, and Molnar (2008) is one of the most frequent phonemes in both spoken English and Spanish. For monolinguals, the precise age of perceptual reorganization varies in relation to the frequency of occurrence of the target speech sounds in children's ambient language (Anderson, Morgan, & White, 2003). This proposal was tested in a study by Burns, Yoshida, Hill, and Werker (2007). They compared children from monolingual English environments and children from bilingual French-English environments on their neural discrimination abilities of the English- versus French-specific voicing contrast in bilabial stop consonants that occur with high frequency in both spoken English and French. At six to eight months both monolinguals and bilinguals responded alike; specifically, they showed a language-general pattern of neural discrimination that was driven by the saliency of the acoustic properties of the stimuli, rather than being influenced by language experience. While bilinguals were able to neurally discriminate the phonemic contrasts consistent with both French and English both at ten to twelve months of age and fourteen to twenty months of age, monolinguals were unable to neurally discriminate the French-specific phonemic contrast past the age of six to eight months. Their findings suggest that in monolingual and bilingual infants, language-specific neural discrimination that is in accordance with their respective language environment has emerged by ten to twelve months of age and that it matches adult phonetic categories.

The maturation of speech perception, however, is not complete at one year of age and extends well beyond early childhood. This has been found for both monolingual (Polka, Colantonio, & Sundara, 2001) and bilingual language development (Sundara, Polka, & Genesee, 2006), at least for some phonemes. Sundara and colleagues (2006) compared simultaneous bilingual French-English children with age-matched monolingual English and monolingual French controls and adults with respect to their ability to neurally discriminate between the English /d/ versus /ð/ contrast. The authors found that monolingual English, but not bilingual four-year-olds, were significantly better than the monolingual French four-year-olds at discriminating the English /d/ versus /ð/ contrast, whereas by adulthood this difference between the monolingual English speakers and the bilingual French-English speakers had disappeared.

In general, our understanding of age effects on the development of phonetic categories is limited by the differences in the application of age-appropriate, behavioral tasks that are

suitable to test infants, young children, and adults alike. As opposed to the high-amplitude sucking technique (e.g., Byers-Heinlein et al., 2010), the conditioned head-turn paradigm (e.g., Polka, Colantonio, & Sundara, 2001), or the visual habituation procedure (e.g., Sundara et al., 2008) that can be used successfully with infants, research with older children often relies on tasks that require differential association of a spoken word and a certain referent object and for adults even more complex and cognitively demanding auditory behavioral discrimination tasks, where participants have to make a conscious decision in the categorization of a stimulus.

## 2.2 Neurophysiological Correlates of Language Attunement

Studies are needed in which similar/identical tasks are used to test infants, children, and adults to allow a better understanding of how language experience shapes the developmental trajectory of humans' phonetic perception abilities across the lifespan. The use of techniques that index brain activity allow the investigation of the neural mechanisms and processes that underlie speech perception (including neural discrimination and categorization). Electroencephalography (EEG) is a non-invasive neurophysiological method that records electrical activity via electrodes placed on a participant's scalp. This method can be used to investigate developmental changes in the neural correlates of speech processing without the need of a consciously controlled behavioral task. EEG methods allow measurement of a nearly instantaneous neural activity to a stimulus or event (i.e., event-related potentials; ERPs). ERPs are time-locked to the EEG to a certain stimulus or event (Vaughan & Kurtzberg, 1992). The ability to record ERPs without a behavioral task allows the same method to be used independent of a participant's age or even conscious state. ERPs provide an opportunity to study the time course of neural processes with great temporal resolution. Their components are classified according to the time in milliseconds of the occurrence of peaks (latency), their polarity (positive or negative), and their topographic distribution over the scalp (e.g., frontal vs. occipital sites). The spatial resolution of the EEG is fairly low, especially when compared to other neurophysiological methods, such as functional magnetic resonance imaging, but ERPs still provide some information about underlying brain sources, particularly when interpreted in conjunction with evidence from research on other species and other imaging methods.

### 2.2.1 The Mismatch Negativity

Neurophysiological auditory neural discrimination studies often use the Mismatch Negativity (MMN) method (Näätänen & Alho, 1995; Näätänen, Paavilainen, Rinne, & Alho, 2007). The MMN is an ERP that indexes the brain's pre-attentive detection of an infrequent change in an

auditory stimulus after being presented with a series of identical stimuli. In adults, the MMN is observed as a frontally-distributed negativity peaking between 100 and 250 ms after stimulus onset (Näätänen, 2001). It is computed by subtracting the brain's responses to the repeated stimulus (i.e., standard) from those of the infrequent stimulus (i.e., deviant). As neural discrimination between a standard and a deviant becomes more difficult, the MMN shifts later in time and becomes smaller in amplitude (Näätänen et al., 2007).

Several studies suggest that the MMN can be elicited in infants. Alho and colleagues (1990) first reported an MMN among sleeping newborns after a pure tone change from 1000 to 1200 Hz that peaked between 160 and 400 ms after stimulus onset. Other studies have reported an MMN in infants to changes related to pitch (Cheour, Kushnerenko, Čeponienė, Fellman, & Näätänen, 2002), duration (Brannon, Wolfe Roussel, Meck, & Woldorff, 2004), and phonetic properties (Cheour et al., 1998). Nevertheless, despite the well-established properties of the MMN in adults, the polarity and latency of this ERP reported for infants and young children are highly inconsistent across studies (Cheng et al., 2015). For instance, when compared to the adult MMN, the MMN found in infants has usually been described to occur in a relatively late time window and to persist for a longer period of time.

Other studies have reported a positive rather than a negative deflection related to a change in an auditory stimulus (e.g., Friederici, Friedrich, & Weber, 2002). Due to these inconsistencies in the literature, when referring to infants and young children I will henceforth be using the term Mismatch Response (MMR; pMMR refers to a positive response, nMMR refers to a negative response). One factor that is hypothesized to impact the polarity of the MMR is neural maturation (Cheng et al., 2013). Additionally, stimulus-related factors such as the degree of contrast between the standard and the deviant stimulus have been found to influence the polarity of the MMR. For instance, Morr and colleagues (2002) found that the majority of their preschool participants failed to show adult-like MMNs to a change in stimulus frequency by the age of four years but that the amplitude of the pMMR declined with age while more adult-like negative characteristics emerged. Note that this was despite this change being behaviorally detectable by this age group. Furthermore, Ahmmed, Clarke, and Adams (2008) found that children's individual language abilities also influenced the presence of a pMMR or an adult-like MMN. Seven-to-eleven-year-old language impaired children were more likely to exhibit a pMMR whereas age matched controls showed an MMN to the same stimulus contrast. However, since these findings all rely on studies with non-speech stimuli, it remains unclear to what extent changes in perception are exclusively related to general developmental changes and

maturation processes and to what extent experience with the specific phonology of a particular language (or languages) may influence this developmental trajectory for speech processing.

### 2.2.2 Automaticity in Speech Sound Perception

In his model of speech perception, Jusczyk (1997) claimed that speech perception becomes increasingly automatic in the course of language development. As pointed out previously, automatic speech perception has been addressed in the context of differences in processing/discriminating native versus non-native speech sounds and has been elaborated further by Strange (2011) in her Automatic Selective Perception (ASP) model. In more detail, the model explains differences in the processing a first (L1) and second (L2) language. Speech perception is defined as a “purposeful, information-seeking activity whereby adult listeners detect the most reliable acoustic parameters that specify phonetic segments and sequences” (p. 456). In order to efficiently process their native language (L1), adults make use of “highly over-learned selective perception routines” (SPRs). In contrast, the extraction of sufficient information necessary to differentiate phonetic contrasts that do not occur in the listener’s native language requires greater attentional resources. SPRs become highly over-learned, and thus automatized through years of experience with a language. Once established, SPRs permit robust and rapid (i.e., automatic) perception that does not require listeners’ focal attention.

In adults, this automatic perception is indexed by a robust negative discriminative brain response (i.e., the MMN) in the EEG. Evidence supporting the ASP model has been found for late learners of an L2 (e.g., Shafer et al., 2021) but only a few studies have examined the development of SPRs and developing automaticity in neural speech sound discrimination in children who grow up as dual language learners of two languages (e.g., Yu et al., 2019). The presence of a pMMR in infants and young children may indicate that perceptual routines have not fully been established, and therefore speech perception requires greater attentional demands (Shafer, Yu, & Garrido-Nag, 2012), whereas the transition to a robust adult-like MMN reflects a more automatic stage of processing that is in accordance with the accumulated linguistic experience with the specific phonetic properties of one or several languages.

### 2.2.3 Electrophysiology of Neural Speech Sound Discrimination in Infancy and Early Childhood

A study by Cheour and colleagues (1998) was among the first to provide neurophysiological evidence for the development of language-specific memory traces for speech sounds in the brains of infants. When comparing monolingual Finnish infants’ brain responses to a native

Finnish vowel contrast (/e/ vs. /ö/) to their responses to a non-native, Estonian vowel contrast (/e/ vs. /õ/) that compared to the Finnish contrast is acoustically more salient, they found that at six months of age, Finnish infants showed nMMRs to both, the native and the non-native vowel contrast. At 11 months of age, however, the same Finnish babies displayed an increased nMMR for the native contrast but a decrease for the non-native contrast, despite its acoustic salience. Similarly, Rivera-Gaxiola, Silva-Pereyra, and Kuhl (2005) found that at seven months of age, their monolingual American-English participants showed neural indices of pre-attentive discrimination for both a native (English) and a non-native (Spanish) consonant contrast. However, by eleven months of age, when analyzed at the group-level, infants' ERPs showed increased receptiveness only to the phonetic contrast that conveys meaning in English (i.e., the native contrast) and failed to neurally discriminate the non-native contrast. Taken together these studies confirmed the behavioral findings concerning the transition from language-general to language-specific processing in monolingual infants during their first year of life. This suggests that language experience results in neural commitment to the specific sounds of the native language early on during development (Kuhl, 2004).

Applying the same methods and stimuli as Rivera-Gaxiola et al. (2005) but extending them to two groups of simultaneous Spanish-English bilingual infants (one aged six to nine months and another aged ten to twelve months), Garcia-Sierra and colleagues (2011) found that bilinguals did not show the same developmental trajectory observed for English monolinguals in their ability to neurally discriminate the sounds of English and Spanish. The authors found that bilinguals developed phonetic representations at a slower pace than monolingual infants. Their bilingual participants did not show any neural indices of pre-attentive discrimination before the age of ten to twelve months. Furthermore, their study was the first to examine the relationship between infants' relative amount of language exposure and their sensitivity to phonetic contrasts. Language experience in both languages was assessed through a parental language background questionnaire and was found to correlate positively with the strength of the neural commitment (i.e., the size of the negative amplitude of the ERP) at ten to twelve months.

This relationship between language input and infants' brain responses was further substantiated by findings reported by Garcia-Sierra, Ramírez-Esparza, and Kuhl (2016) who collected more direct measures of language input using the Language Environment Analysis System (LENA foundation, Boulder Colorado; see [www.lena.org](http://www.lena.org)). Based on the counts of words parents addressed to their children (in each respective language), participants were divided into a high-input versus low-input group. The authors found that monolingual infants

in the low language input group and bilingual infants in the high English language input group showed similar word counts for English. In the monolingual group, only high language input was associated with a more robust nMMR. Similarly, in bilinguals, language input quantity was associated with advances in the process of neural commitment (i.e., a greater nMMR). However, significant nMMRs were observed in neither the high English nor high Spanish input group, suggesting that bilinguals were not fully committed to either of their two languages. These findings suggest that the differences previously found in the timeline for neural commitment between monolingual and bilingual infants are the result of differences in language input quantity. Bilingual and monolingual infants with similar amounts of English input (i.e., high English input bilinguals and low-input monolinguals) showed similar neural trajectories. The authors suggested that these children would transition to full neural commitment later in life once they had gathered sufficient experience with their language(s).

Only a few studies have examined this development beyond infancy (i.e., in toddlerhood). Shafer, Yu, and Datta (2010) found a robust presence of a discriminatory brain response to the English vowel contrast /i/ vs. /ɛ/ in four-to-seven-year-old monolingual English children that was consistent with the adult pattern, suggesting maturation of speech perception and automaticity of speech processing. However, there is evidence that for very complex phonetic properties, the shift towards the adult-like MMN occurs even later in life (Liu, Chen, & Tsao, 2014). Furthermore, the inconsistent presence of an nMMR under the age of four years suggests that representations of many contrasts are not established well enough to allow for pre-attentive, automatic discrimination in infants and toddlers. Examining the neural discrimination abilities of a German vowel contrast (/e/ vs. /ɛ/) in five-year-old bilingual Turkish-German children, Rinker, Alku, Brosch, and Kiefer (2010) found that, compared to their German monolingual peers, the MMN amplitude in bilinguals was significantly reduced despite their immersion in a German environment for at least two years. However, their brain response to a vowel contrast that exists both in German and in Turkish (/i/ vs. /y/) did not yield significant group differences. Conversely, a group of Finnish children was found to show native-like neural discrimination of a foreign vowel contrast after only a few months of being immersed in this foreign language context (Cheour, Shestakova, Alku, Ceponiene, & Näätänen, 2002). These contradictory findings with (early) sequential bilingual children further highlight the influence of child-external factors (i.e., input quantity and quality) on children's perceptual abilities. Datta and colleagues (2019) used the English vowel-contrast /i/ versus /ɛ/ (that is not present in Spanish) to compare neural measures of automatic change detection between early Spanish-English bilingual adults and children. They found that their bilingual child participants

were indistinguishable from their monolingual peers. However, these bilingual children were eight to eleven years of age, and the study did not assess the bilingual participants' neural discrimination abilities in Spanish, in addition to English. The questions of whether and how automaticity in speech sound processing evolves for two languages when acquiring them both simultaneously in the first five years of life therefore remained unanswered.

#### 2.2.4 Voice Onset Time

As I have established in Section 2, languages differ in both the inventory of speech sounds (phonemes) and in how these phonemes are produced (i.e., their phonetic realization). Voice Onset Time (VOT), the onset and timing of voicing (i.e., vibration of the vocal folds), is an interesting example for both phonemic and phonetic differences across languages. VOT is a laryngeal feature which means that its acoustic and perceptual properties are characterized by the state of the larynx or the vocal folds (Blevins, 2004). VOT refers to the length of time that passes between the release of a stop consonant and the onset of voicing (usually of a following vowel but also sonorant consonant). VOT realizations for a phoneme contrast may differ across languages, and thus, it is both interesting and useful in the study of bilingualism (e.g., Kupisch & Lleó, 2017). For instance, both Germanic (e.g., German) and Romance (e.g., Italian) languages have a phonemic two-way voicing contrast for stop consonants. Specifically, they show a phonological contrast between voiced and voiceless stops. The two language groups differ, however, in the way in which this contrast is phonetically realized. Germanic languages contrast a long lag VOT with short lag VOT, whereas Romance languages contrast short lag and voicing lead consonants. The Germanic long lag stops are aspirated (i.e., accompanied by a turbulent burst of air) (Cho, Whalen, & Docherty, 2019). By contrast, voicing lead is characterized by phonation during closure before the stop is released. Short lag stops have a VOT at or near zero (typically up to 20 ms); thus, the phonation of a following speech sound (e.g., a vowel) begins at or near to the release of the stop (i.e., opening of the lips). Long lag stops are produced with VOT, that are typically longer than 40 ms (often as great as 80 to 100 ms), and in Germanic languages, reveal a considerable period of aspiration (Kleber, 2018). The presence of a two-way phonemic contrast in Germanic and Romance languages (e.g., /p/ vs. /b/), that at the same time is substantially different in its language-specific phonetic realization makes VOT an interesting and important focus of study in the phonological development of bilinguals who acquire a Germanic and a Romance language.

In monolingual German language development, the short lag bilabial plosive is considered fully acquired (i.e., matching the adult-like target in 90% of children across all

permissible positions) by the age of 18 to 23 months, and the long lag bilabial plosive at the age of 24 to 30 months (Fox, 2000). In English (which, like German, is a Germanic language), three stages in the acquisition of VOT have been identified (Macken & Barton, 1979). In stage 1, children produce English target voiced (i.e., short lag) and (i.e., long lag) voiceless stops in the short lag region without any statistically significant difference between the VOT values for the two target sounds. In stage 2, they still produce both stops in the short lag region, but they produce target voiceless stops with significantly longer VOTs than target voiced stops. However, perceptually this distinction is not yet noticeable. Finally, they reach stage 3 and produce (perceivable) distinct VOT values for voiced versus voiceless targets. However, often with a tendency, in the case of the latter, to overshoot and produce values more extreme than adult-like measures. In contrast, in Italian monolingual language development, the short lag bilabial plosive is considered fully acquired by the age of 18 months and the voicing lead bilabial stop by 24 months; however in word initial and intervocalic position, the child productions did not match the adult target, even at 27 months of age (Zmarich & Bonifacio, 2005). Other studies that look at languages with a short lag-voicing lead distinction report that voicing lead is acquired late and may not be fully acquired with adult-like values even by the age of five years (e.g., Macken & Barton, 1980 for Spanish, which like Italian is a Romance language). Macken and Barton (1980) pointed out that children may make use of other cues to realize the voiced versus voiceless contrast. These cues include a significant VOT distinction in the short lag region. Cross-language differences in acquisition are often explained by the relative difficulty of the laryngeal gesture necessary to realize a certain VOT feature (Lisker & Abramson, 1964) and correspondingly, across languages, short lag is commonly acquired first, followed by long lag and finally voicing lead (Jong Kong, Beckman, & Edwards, 2012).

Bilingual children who acquire a language like Italian (that contrasts short lag and voicing lead) and a language like German (that contrasts short lag and long lag) have to acquire the language-specific VOT distinctions for both languages. In addition, they need to resolve the phonological ambiguity within the short lag VOT category that corresponds to ‘voiceless’ plosives in Italian, and to ‘voiced’ plosives in German. There are no clear-cut developmental norms for the acquisition of VOT categories in bilingual children. Bilingual children who acquire a Germanic and a Romance language simultaneously might not have difficulties with short lag stops but, rather, may have difficulty with voicing lead and/or long lag respectively, because they are only present in one of their two languages (Kupisch & Lleó, 2017).

Only a few studies have examined VOT acquisition in bilingual development. Deuchar and Clark traced the phonetic realization of word initial VOT between the ages of 1;7 and 2;3



years in a Spanish-English speaking child living in England. They found that at 2;3 years of age the child did realize the English long lag-short lag distinction and showed signs of starting to distinguish between the Spanish short lag and voicing lead. The authors suggested that the imbalance between English and Spanish development may stem from the relatively greater acoustic saliency of the English VOT contrast compared to the Spanish distinction. Kehoe, Lleó, and Rakow (2004) examined the VOT values of four simultaneous German-Spanish two-to-three-year-old bilinguals living in Germany and found that the phonetic/phonological systems of bilingual children interacted throughout the course of development. Kehoe and colleagues reported evidence for (bidirectional) transfer and delay but also one case of no cross-linguistic influence in the formation of VOT categories in their bilingual German-Spanish participants. Thus, the effects of bilingualism were not uniform across children. The authors pointed out that dominance factors to a certain extent may account for this finding. Fabiano-Smith and Bunta (2012) compared the formation of VOT categories in eight three-to-four-year-old bilingual Spanish-English speaking children living in the US to that of their monolingual Spanish and English peers. They found that bilinguals did not differ from their monolingual peers in the realization of Spanish short lag but did differ in the realization of English long lag. This finding suggests influence from Spanish to English. In comparison, Watson (1990) found that VOT production of simultaneous English-French bilingual children of six years and older in their two languages was similar to that of monolinguals in both languages, suggesting that in between four and six years of age bilingual children successfully form distinct VOT categories for their two languages.

### 2.3 The Present Study

This study addresses the topic of developing automaticity in processing VOT in simultaneous Italian-German bilingual children and how this relates to children's production thereof and the influence of linguistic experience (i.e., language exposure and dominance). I consider the questions of whether and how it is possible for young children to become automatic in processing two phonetic systems, if the cues for a category between the two languages are in conflict.

### 2.4 Research Questions/Hypotheses

This study aims to examine the relationship between children's discriminatory brain responses, vocal productions, and measures of language experience and language use in early dual language development. Specifically, I address the following research questions:

**RQ1:** What impact does language experience have on bilingual children's automatic speech sound processing in their two languages?

Hypothesis 1: I predict a robust presence of a negative component of the MMR to German-like VOT deviants, indicating automaticity in speech processing, in four-to-five-year-old monolingual German children (Shafer, Yu, & Datta, 2010). Considering that bilingual children's language experience with either one of their two languages is inevitably reduced compared to their monolingual peers, I further predict less automatic neural speech sound discrimination of German-like VOT in the bilingual Italian-German group. This pattern will be indexed by a reduced and/or delayed (n)MMR or the presence of a pMMR (Morr et al., 2002). Conversely, I hypothesize that bilingual Italian-German children will show a discriminatory (n)MMR to Italian-like VOT deviants. In contrast, monolingual German children are predicted to show no evidence of neural discrimination for this foreign contrast, at least in the difficult condition, as predicted by the SLM (Flege, 1995). I predict monolingual German children to assimilate the difficult voicing lead speech sound into their short lag VOT category, and thus, to show poor or absent neural discrimination of the subtle, phonetic difference between these two speech sounds.

Hypothesis 2: Generally, in the group of bilingual children, I predict signs of increased (in)voluntary attention to speech sounds, as indexed by overall more negative brain responses when compared to the group of monolingual children (Cheour et al., 2002). This would be in line with previous research suggesting that bilingual speakers more commonly need to rely on details of the surrounding speech stream in order to identify the target language (Ortiz-Mantilla et al., 2010).

**RQ2:** What is the minimum input threshold that allows for monolingual-like automatic speech processing in both languages?

Hypothesis 3: I predict the polarity (negativity vs. positivity) and amplitude of the MMR to be modulated by children's Italian versus German experience; specifically, bilinguals who fall below a minimum input threshold will exhibit more immature signs of processing (i.e., reduced nMMR amplitude/presence of a pMMR and/or later onset; Shafer, Yu, & Datta, 2010).

**RQ3:** In the production of VOT, are children's VOT values target-like (i.e., compared to their monolingual peers) and do they realize voicing opposition in both languages?

Hypothesis 4: Considering the difficulty in the voicing lead-short lag distinction even in monolingual children (Macken & Barton, 1980) and the greater acoustic saliency of long lag compared to voicing lead (Deuchar & Clark, 1996), I predict bilingual Italian-German children

to show a monolingual-German-like voicing contrast in German but greater deviance from the monolingual Italian target pattern for the Italian voiced versus voiceless distinction.

**RQ4:** How does language experience (relative amount of input and output, and measures of input quality) relate to bilingual children's realizations of VOT categories in Italian and German?

Hypothesis 5: I predict that children's realization of language-specific VOT is dependent on both the quality and amount of language input in the respective language (Kehoe, Lleó, & Rakow, 2004) and their use of each language (i.e., language output).

## 2.5 Methodology

### 2.4.1 Participants

A total of 40 TD children between the ages of 3;11 and 6;3 participated in this study. Twenty-four of the children were simultaneous or early-sequential bilingual Italian-German speaking children (18 females) with a mean age of 59.4 months ( $SD = 8.5$  months) and 16 were monolingual German speaking children (6 females) with a mean age of 61.1 months ( $SD = 6.4$  months). At the time of their participation in this study, all children were living and attending a kindergarten in Germany. All bilingual participants had at least one native Italian-speaking parent and were exposed to Italian on a daily basis, although to varying degrees. Twenty-two of the bilingual children were born and raised in Germany and two in Italy; these two had moved to Germany before the age of 3. Participants with two Italian-speaking parents had been exposed to German for a minimum of two years (Table 1 provides an overview of parents' language background; for bilingual children's individual measures of language input and output see Table A1 in Appendix A). While one of the bilingual Italian-German children was attending a monolingual German kindergarten<sup>2</sup>, 23 were enrolled in a bilingual Italian-German kindergarten program. Their dual language environment thus provided them with frequent language input from multiple speakers in both Italian and German.

---

<sup>2</sup> I acknowledge that because of this difference, this child's everyday language experience may be different from the rest of the sample. However, with respect to their quantitative measurements of current language exposure (see section 2.4.3), this participant (participant ID 306612; see Table A1 in Appendix A) did not deviate substantially from the other children in the group of bilingual Italian-German children and was thus included in the analysis.

**Table 1.** Parents’ language background as assessed by a Language Background Questionnaire (LBQ; see Section 2.4.3 for more details). In the Can-Do-Questionnaire (Section 2.4.3), parents additionally had to rate their language proficiency in German and Italian for a variety of different oral and written competences on a scale from 1 to 5 (5 = very good, native-like language skills). Note that a great proportion of parents are considered heritage speakers of Italian themselves (58.9% of the mothers and 41.67% of the fathers); meaning that they had been born in Germany to Italian speaking parents.

		Mothers ( <i>n</i> =23)	Fathers ( <i>n</i> = 22)
Language(s) spoken	German only	17.39%	31.82%
	Italian only	8.70%	13.64%
	German and Italian*	73.91%	54.55%
	*including:		
	Second language learners of German	41.18%	33.33%
	Heritage speakers of Italian	58.82%	41.67%
	Heritage speakers of German		8.33%
	Early sequential bilinguals (L1 Italian)		16.67%
Self-rated language skills	German	<i>M</i> = 4.57, <i>SD</i> = 0.87	<i>M</i> = 4.12, <i>SD</i> = 1.32
	Italian	<i>M</i> = 3.71, <i>SD</i> = 1.63	<i>M</i> = 3.57, <i>SD</i> = 1.86

All monolingual German children had two monolingual German-speaking parents, had been born and raised in Germany and were attending a monolingual German kindergarten program. The monolingual German children thus had no prior experience with the Italian language.

Additionally, for the VOT production task only, a TD monolingual Italian control group (*n* = 5; 2 females; age *M* = 59.4 months, *SD* = 6.1 months) was included in the study. All children in this group had two monolingual Italian-speaking parents, were born and raised in Italy and were attending a monolingual Italian kindergarten in Italy. The monolingual Italian children thus had no prior experience with the German language.

The study was approved by Ethics Committee of the Catholic University Eichstätt-Ingolstadt and all parents signed informed consent according to the Declaration of Helsinki.

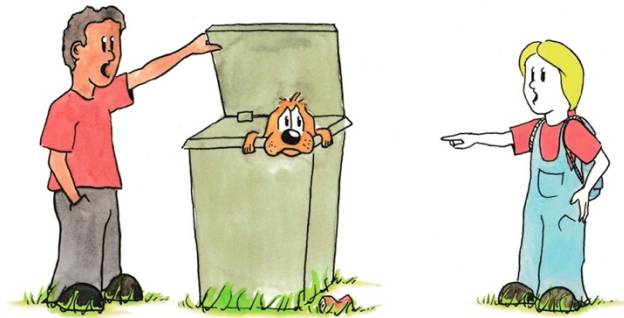
#### 2.4.2 Behavioral Measures

**Colored Progressive Matrices (CPM).** The German adaptation of Raven’s CPM (Bulheller & Häcker, 2001) was used for the assessment of children’s nonverbal intelligence. The test consists of 3 blocks each including 12 tasks and hence a total of 36 items. For each task, children are presented with a visual geometric pattern with a missing piece. Out of six options, children are supposed to choose the piece that needs to be added in order to complete the pattern. CPM scores are divided into five categories: from level I (“outstanding intellectual capacity”; score falls above the 95<sup>th</sup> percentile rank) to level V (“mentally disabled”; score falls below the 5<sup>th</sup>

percentile rank). Participants whose nonverbal intelligence scores fell below the 25<sup>th</sup> percentile (level IV or V) were excluded from the experiment.

**Linguistische Sprachstandserhebung Deutsch als Zweitsprache** (LiSe-DaZ; Schulz & Tracy, 2011). Children's morpho-syntactic abilities in German were assessed using the elicited production task of the LiSe-DaZ that provides norms for monolingual German children as well as for L2 learners of German in accordance with their length of exposure to the L2. The elicited production task of the LiSe-DaZ addresses several areas of German morpho-syntax and was thus used to verify children's linguistic status as TD. The task is administered as a semi-structured storytelling interview and is presented in form of a picture book. To ensure that the child is familiar with the labels for the most important objects and animals that appear in the story, the test is preceded by a brief vocabulary task that is not included in the final evaluation. For this purpose, the child is presented with a total of 16 picture cards one after the other and asked to name the pictures that are depicted. If the child is not familiar with the one of these objects or animals, the examiner provides the corresponding German label. Similarly, throughout the entire assessment, the examiner can provide labels for objects that appear in the story if the child seems to not know them. After two practice sentences that are designed to introduce the child to the story and the type of questioning, the test begins. Test items are either questions that elicit certain morpho-syntactic structures (e.g., "Was fragt Lise?" [What is Lise asking (them)] or beginnings of sentences that the child has to complete (e.g., "Du darfst nur mitkommen, (wenn)..." [You are only allowed to come, (if)...]). Children's utterances are evaluated according to four categories (1) production of complex sentences and verb placement, (2) subject-verb-agreement, (3) word classes, and (4) case markings.

*Production of complex sentences and verb placement* is judged on a scale ranging from 1, indicating a single-word-utterance, to 4, indicating the use of embedded sentences with the verb in sentence-final position in a subordinate clause. For instance, the question "Warum macht der Hund so ein trauriges Gesicht?" [Why is the dog making such a sad face?] is used to elicit children's production of a subordinate clause (see Figure 1 for the corresponding picture in the LiSe-DaZ). A common, correct level 4 response is "Weil er in der Mülltonne ist" [Because he is (caught) in the garbage can]. By contrast, an utterance such as "Weil ist traurig" with the verb in second position is assigned level 3. A level is considered mastered if the child produces at least three utterances corresponding to that level throughout the entire test.



**Figure 1.** Picture 3 in the LiSe-DaZ used to elicit a subordinate clause. Elicitation question: “Warum macht der Hund so ein trauriges Gesicht?”

*Subject-verb-agreement* scores are calculated by first identifying all utterances that contain a subject and a verb (i.e., obligatory contexts for the correct realization of agreement) and in a next step counting all the utterances with correct subject-verb agreement. Finally, the ratio between the number of obligatory contexts for subject-verb-agreement and the correctly realized instances is calculated. For instance, a common, correct response to the question elicitation “Was passiert hier?” [What is happening (in this picture)?] is “Die (Kinder) spielen<sub>[play 3rd person plural]</sub> Fußball” [They are playing football], whereas “Die (Kinder) spielt<sub>[play 2nd person singular]</sub> Fußball” is counted as incorrect (see Figure 2 for the corresponding picture in the LiSe-DaZ). By comparison an utterance such as “Spielen Fußball“ is not included in the analysis as it is missing the subject and thus, agreement marking cannot be assessed. The maximum score is 1.0; children’s performance again is represented on a four-point scale.



**Figure 2.** Picture 6 in the LiSe-DaZ used to elicit a main clause including a subject and a verb. Elicitation question: “Was passiert hier?”

*Word classes* scores reflect the number of elements corresponding to one of the target word classes (prepositions, focus particles, conjunctions, full verbs, and modal verbs). For each of these five word classes, a separate raw score is obtained.

*Case marking* evaluates children's realizations of the accusative and the dative. In eight sentences, the accusative is elicited four times and the dative is elicited five times. In order to allow unambiguous case markings, nouns were chosen which have visible case features in German at least for the definite article (e.g., Ball [ball], Hase [rabbit], See [lake]). Raw scores reflect the sum of the correctly realized accusative and dative markings.

Children for which the LiSe-DaZ indicated language learning needs in two or more subtests, were excluded from the experiment.

**Target Sound Elicitation Task.** Children's vocal productions of word initial VOT were elicited in both Italian and German through a brief picture naming task. For each language, the task comprised ten pictures (to elicit five tokens in the voiced and five tokens in the voiceless category; for German: Baby [baby], Bär [bear], Ball [ball], Baum [tree], Bett [bed], Pullover [pullover], Puppe [doll], Polizist [policeman], Pinguin [penguin], Pommes [fries]; for Italian: bicicletta [bicycle], bambola [doll], biscotti [cookie], bocca [mouth], bottone [button], palla [ball], pinguino [penguin], porta [door], pane [bread], pesce [fish]). The target items were selected based on the German and Italian adaptation of the MacArthur Bates Communicative Development Inventories (for the German adaptation see Szagun, Stumper, & Schramm, 2014; for the Italian adaptation see Caselli, Bello, Rinaldi, Stefanini, & Pasqualetti, 2015) that reflect lexical knowledge of monolingual children under the age of three. In order to elicit the target words in isolation, the respective determiner was provided in the elicitation question (for German: "Das ist ein/e...?"; for Italian: "Questo/a è un/a...?" [This is a/n...?]). German and Italian word initial VOT was elicited separately. Children's vocal productions were recorded using a Philips VoiceTracer DVT6010 at a sampling rate of (44000 Hz); VOT was measured in Praat (Boersma & Weenink, 2018).

#### 2.4.3 Measures of quantity and quality of children's language input

Children's current relative language exposure was measured by means of a parental questionnaire to gain an objective estimate of the average proportion of the time participants heard and spoke their heritage (Italian) compared to the societal language (German) during a

typical week of their lives. The language background questionnaire (LBQ) was adapted from the two questionnaires used in Rinker, Shafer, Kiefer, Vidal, and Yu (2017) and, according to the caregivers' preference, was provided either in Italian or in German.

The main information that was gathered can globally be classified into two categories: (1) language use in the home; that is how much Italian and/or German each family member (e.g., each parent/caregiver, sibling, or any other adult living in the home) spoke to the child (input) and how much Italian and/or German the child spoke to each family member (output); and (2) language use outside the home; that is how many hours per week the child spent outside of the home (e.g., in kindergarten, with another caretaker outside of the core-family context, on leisure activities, or with friends) and how much Italian and/or German the child heard (input) and spoke (output) during these times. Parents were asked to use a seven-point scale defined by the frequency adverbs (below) and a percentage scale to estimate the proportion of their children's Italian compared to German exposure in different contexts. In order to preclude the sum of percentages adding up to more than 100% of total language exposure, the two languages were combined in the same scale (see below).

1. Only German (100% German, 0% Italian)
2. Predominantly German, hardly any Italian (90% German, 10% Italian)
3. Mostly German, sometimes Italian (75% German, 25% Italian)
4. The same amount of German and Italian (50% German, 50% Italian)
5. Sometimes German, mostly Italian (25% German, 75% Italian)
6. Hardly any German, predominantly Italian (10% German, 90% Italian)
7. Only Italian (0% German, 100% Italian)

Based on parents' responses a compound score representing children's current language exposure (one for their language input and output respectively) was calculated similar to the procedure described by Cattani and colleagues (2014). First, the number waking hours per week ( $h_{awake}$ ) was calculated based on an average estimate of a mean sleep duration per night 10.47 hours in preschool children (Scharf, Demmer, Silver, & Stein, 2013).

$$h_{awake} = \frac{\text{total number of hours}}{\text{week}} - \frac{\text{mean sleep duration}}{\text{night}} * 7$$

Second, for every activity outside of the family home, the average time spent on this activity per week was calculated: (1) hours per week spent in kindergarten ( $h_{kindergarten}$ ), (2) hours per



week spent with another caretaker outside of the core-family context ( $h_{\text{caretaker}}$ ), (3) hours per week spent with leisure activities ( $h_{\text{leisureact}}$ ), and (4) hours per week spent with friends ( $h_{\text{friends}}$ ).

$$h_{\text{kindergarten}} = \frac{\text{hours in kindergarten}}{\text{day}} * \frac{\text{days in kindergarten}}{\text{week}}$$

$$h_{\text{caretaker}} = \frac{\text{hours with caretaker}}{\text{day}} * \frac{\text{days with caretaker}}{\text{week}}$$

$$h_{\text{leisureAct}} = \frac{\text{hours of leiusure activities}}{\text{day}} * \frac{\text{days with leisure activitiey}}{\text{week}}$$

$$h_{\text{friends}} = \frac{\text{hours with friends}}{\text{day}} * \frac{\text{days with friends}}{\text{week}}$$

Third, the time children spent awake with their families ( $h_{\text{family}}$ ) was calculated by subtracting the time they spent outside of the family home from the number of waking hours.

$$h_{\text{family}} = h_{\text{awake}} - (h_{\text{kindergarten}} + h_{\text{caretaker}} + h_{\text{leisureAct}} + h_{\text{friends}})$$

Next, I computed an estimate of the number of children's Italian-hearing hours per week (i.e., their Italian input<sup>3</sup>) by multiplying the weekly hours spent in the different contexts with percentage scale information provided by the parents for each respective activity (i.e., 100%, 90%, 50%, 25%, 10%, or 0%). Below, I will demonstrate this step for the calculation of the number of hours per week that children hear Italian within the kindergarten context ( $[\text{h}]Input_{\text{ItalianKindergarten}}$ ). For example, if a child spends 10 hours per week in the kindergarten where the language, they hear is 75% Italian, these 10 hours will be converted into 7.5 hours of Italian (and 2.5 hours of German respectively).

$$[\text{h}]Input_{\text{ItalianKindergarten}} = h_{\text{kindergarten}} * [\%] \text{ Italian child hears in kindergarten}$$

The same procedure was applied for Italian input from other caretakers ( $[\text{h}]Input_{\text{ItalianCaretaker}}$ ), during leisure activities ( $[\text{h}]Input_{\text{ItalianLeisureAct}}$ ) and during time spent with friends ( $[\text{h}]Input_{\text{ItalianFriends}}$ ). Further, when calculating the number of hours children hear Italian within

---

<sup>3</sup> Note that the exact same approach described next was followed in order to obtain the number of hours children spoke Italian during a typical week of their everyday lives (i.e., their relative Italian output)

the family context  $[h]Input_{ItalianFamily}$ ), unlike Cattani and colleagues (2014) I did not assign weight to the input provided by different family members. Rather, I calculated a mean family Italian input score by averaging across the individual data provided for each family member living in the same household as the child.

$$[h]Input_{ItalianFamily} = h_{family} * \bar{X}(\%Italian_{mother} + \%Italian_{father} + \%Italian_{siblings} + \%Italian_{otheradults})$$

Finally, the total number of hours of Italian input per week was calculated by adding up all of the Italian input measures; that sum was then divided by children's average waking hours ( $h_{awake}$ ) in order to obtain the relative amount of Italian input.

$$[h]Italian_{input} = [h]Input_{ItalianFamily} + [h]Input_{ItalianKindergarten} + [h]Input_{ItalianCaretaker} + [h]Input_{ItalianLeisureAct} + [h]Input_{ItalianFriends}$$

$$[\%]Italian_{input} = \frac{[h]Italian_{input}}{h_{awake}}$$

The relative amount of German input could then easily be derived on the basis of the relative Italian input by subtracting it from 100%.

Additionally, in order to obtain a rough estimate of the quality of children's language input, caregivers were asked to fill out a "Can-Do Checklist" and evaluate their own oral and written language skills in both languages (Italian and German) on a scale from 1-5 (5 = very good, native-like skills) in both the productive and receptive domain.

#### 2.4.4 Electrophysiological Data

Natural speech stimuli were recorded by a native speaker of Bengali, a language that includes all three speech sounds used in this experiment to ensure that the stimuli were not biased towards German or Italian<sup>4</sup>. Recordings were obtained in a sound-shielded booth. Stimuli were comprised of stop consonant-vowel syllables: voicing lead [ba], short lag [pa], and long lag [p<sup>h</sup>a]. The short lag stop ([pa] with 0 ms VOT), common to both German and Italian was used as the frequent stimulus (standard). A behavioral ABX perception task with monolingual Italian ( $n = 11$ ) and monolingual German ( $n = 8$ ) adults was used to select the VOT values to serve as

---

<sup>4</sup> Note that despite the stimuli having been recorded by a speaker of Bengali, the stimuli will be referred to as either Italian-like versus German-like because they share the same VOT features applied in the Italian and/or German language.

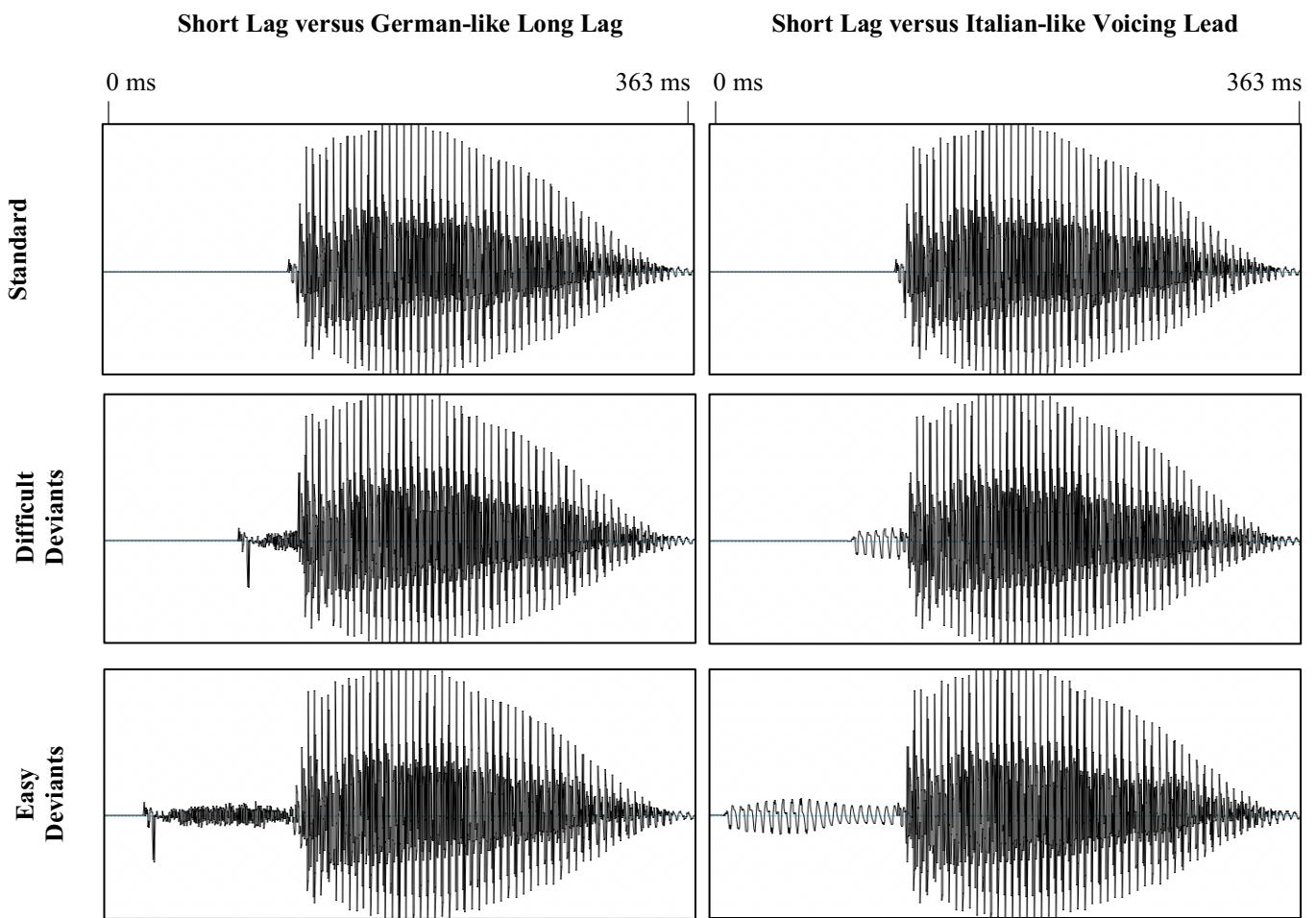
the infrequent (deviant) stimuli in the oddball paradigm. For each language-specific VOT contrast (i.e., voicing lead-short lag for Italian; long lag-short lag for German), adult participants were presented with 5 trials of three consecutive stimuli (target stimulus A, target stimulus B, unknown stimulus X) and had to decide whether the third and last stimulus resembled the first (A) or the second (B) target stimulus. Based on adult monolingual speakers' responses four deviant stimuli were selected. Two EEG oddball paradigms were created, Easy and Difficult for each language (see Table 2 for an overview of the stimuli used in the EEG experiment). The stimuli used in the Easy EEG paradigm were successfully identified as the target by almost all adult speakers independently of their language background (German-like Long Lag of 92 ms identified as long lag on 96.9% of the trials by the monolingual German adults and on 90.9% of the trials by the monolingual Italian adults; Italian-like Voicing Lead of -112 ms identified as voicing lead on 81.2% of the trials by the monolingual German adults and on 93.2% of the trials by the monolingual Italian adults. The deviants for the difficult EEG paradigm were selected such that monolingual adult speakers were able to successfully perceive the stimulus as the target (> 70% correct trials) if it was their native contrast, whereas success rate in allocating the stimulus as the target was at chance level for the foreign distinction (German-like Long Lag of 36 ms identified as long lag on 71.9% of trials by the monolingual German adults and on 47.7% of the trials by monolingual Italian adults; Italian-like voicing lead -36 ms identified as voicing lead on 50.0% of the trials by the monolingual German adults and 77.3% of the trials by the monolingual Italian adults).

**Table 2.** Overview of the Stimuli used in the EEG experiment. In the Easy paradigm the acoustic difference between the standard and the deviants was more pronounced, whereas in the Difficult paradigm the standard and deviants were acoustically more similar and thus exhibited a smaller degree of contrast.

Stimulus type	Language	Properties	VOT Easy paradigm	VOT Difficult paradigm
[ba]	Italian-like /b/	Voicing lead	-112 ms VOT	-36 ms VOT
[pa]	Italian-like /p/, German-like /b/	Short lag	0 ms VOT	0 ms VOT
[p <sup>h</sup> a]	German-like /p/	Long lag	92 ms VOT	36 ms VOT

A double-deviant oddball design was used to examine children's early brain responses to the contrast between the three different realizations of the bilabial stop consonant (cf. Shafer et al., 2021). By applying a double-oddball paradigm, I was able to examine the two deviant stimuli under the exact same conditions so that fatigue or other external conditions would not account

for differences in children’s brain responses to the two deviants (Rivera-Gaxiola et al., 2005). During a passive listening task, children were presented with a train of the standard stimulus [pa], occasionally interrupted by either one of the two deviants [ba] and [p<sup>h</sup>a]. Eighty percent of all stimuli were the repeated [pa] standards. The deviants [p<sup>h</sup>a] and [ba] were equally distributed (10% each) among the remaining 20% of tokens. Stimuli were presented in a pseudo-randomized order to allow for at least three consecutive standard stimuli between the presentation of a deviant. The stimuli were aligned according to the vowel onset rather than the onset of acoustic information (i.e., prevoicing or aspiration) with the goal to present them with a sense of regular rhythm (see Figure 3). The inter-stimulus interval was 600 ms from the offset of the vowel to 122 ms before the onset of the next vowel. At the end of the oddball paradigm, each deviant sound was repeated 100 times for use as a control-deviant (also referred to as the deviant’s identity) to which the deviant response was compared for the subsequent analyses (cf. Garcia-Sierra et al., 2016).



**Figure 3.** The waveform of the standard (Short Lag) and deviant stimuli (German-like Long Lag vs. Italian-like Voicing Lead; Easy vs. Difficult) used in the EEG experiment. The same standard was used in the two paradigms.

**EEG Acquisition.** The EEG signal was recorded at a 500 Hz sampling rate using a BrainProducts Inc. EEG system via a PC laptop running BrainVision Recorder software. This system includes the LiveAmp 32 amplifier to record the continuous EEG from the scalp using 32 actiCAP slim electrodes mounted in the actiCAP snap electrode cap. Electrode placement included standard placements in the 10/10 montage. Electrodes were filled with SuperVisc electrolyte gel to reduce impedances below 50 k $\Omega$ . Active circuits for impedance conversion are directly integrated in the actiCAP slim electrodes. Impedance conversion at the electrode level makes it possible to achieve high signal quality with higher impedances. FCz served as the reference during data collection.

During the EEG recording, no active task was required from the children; they were allowed to watch a muted cartoon on an Ipad screen, while the auditory stimuli were presented binaurally through headphones at 60 dB SPL, delivered via Eprime software (Psychology Software Tools, Pittsburgh, PA, United States).

Participants were randomly presented with either the Easy paradigm or the Difficult paradigm first. Due to fatigue and non-compliance, not all children managed to complete both EEG paradigms. Out of the 24 bilingual Italian-German participants, two did not complete the Easy paradigm and another two did not complete the Difficult paradigm. Note that these are not the same children. Similarly, one monolingual German child did not complete the Difficult paradigm.

**EEG Preprocessing.** The continuous EEG data were processed offline using BrainVision Analyzer software v2.1 (BrainProducts Inc.). After visual inspection of the raw data for each participant, channels contaminated by noise were reconstructed using triangulation and linear interpolation. The signal was re-referenced to the mean-mastoid reference. An IIR filter (low cut-off: 0.10; high cut-off: 30 Hz) was applied to the signal, followed by a 50 Hz notch filter. Independent Component Analysis (ICA) was used to perform ocular correction. The frontal electrodes (FP1 and FP2) served as a blink marker channels for vertical activity. The difference between FT9 and FT10 electrodes served as a marker for horizontal activity. The procedure was conducted in semi-automatic mode. For each participant, ICA components were inspected visually with respect to their topographic location and relative impact on the data. The components that were contributing to blinks were set to zero. Next, the data was segmented into epochs with interval durations of 200 ms pre- and 900 ms post-stimulus. Then artifact rejection was carried out with the criterion of no voltage step of more than 100  $\mu$ V in the segment. Baseline correction was performed using the 200 ms pre-stimulus amplitude. For each

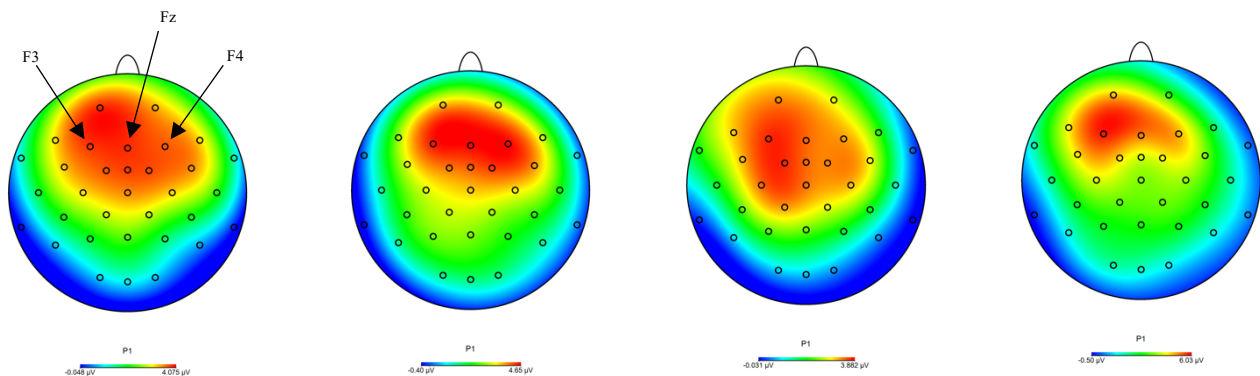
stimulus type (standard, voicing lead deviant, and long lag deviant), segments (-200 to 900 ms) were averaged separately. To ensure that the ERP to the standard did not include any change-related aspects, post-deviant standards were not included in the standard averages. For each deviant, all children had more than 80 trials; 90% had more than 90 trials that were included in the averages. The mean number of trials for the standard and deviant ERPs did not differ significantly between the two groups (all  $p_s > .05$ ; see Table 3).

**Table 3.** Overview of the number of trials for each stimulus according to group (monolinguals vs. bilinguals) and paradigm (Easy vs. Difficult).

		Monolinguals	Bilinguals
Short Lag Standard	Difficult	$M = 572.55, SD = 24.52$	$M = 572.98, SD = 20.24$
	Easy	$M = 574.11, SD = 18.84$	$M = 575.81, SD = 13.57$
Voicing Lead Deviant	Difficult	$M = 95.33, SD = 4.27$	$M = 95.58, SD = 3.34$
	Easy	$M = 94.11, SD = 2.65$	$M = 94.02, SD = 2.54$
Long Lag Deviant	Difficult	$M = 95.16, SD = 4.33$	$M = 95.34, SD = 3.34$
	Easy	$M = 97.56, SD = 3.87$	$M = 97.82, SD = 2.54$
Voicing Lead Identity	Difficult	$M = 94.19, SD = 5.38$	$M = 94.56, SD = 4.92$
	Easy	$M = 95.07, SD = 3.99$	$M = 95.56, SD = 3.63$
Long Lag Identity	Difficult	$M = 95.38, SD = 4.55$	$M = 94.80, SD = 4.67$
	Easy	$M = 95.36, SD = 3.25$	$M = 96.29, SD = 3.14$

The identity MMR (iMMR) was generated by subtracting the control stimulus from the deviant of the same stimulus.

**EEG Analysis.** The first step in the analysis was to reduce the ERP data from 32 sites to a model representing the MMR. I utilized a principal component analysis (PCA) to examine the topography of children’s iMMRs and to determine the optimal electrodes to include in the subsequent analysis. The PCA identified similar topography for component 1 (that accounted for the most variance), which included sites F3, Fz, and F4 (see Figure 4). These sites have also



**Figure 4.** Results from the PCA for the iMMR from left to right for Voicing Lead -36 ms VOT, Voicing Lead -112 ms VOT, Long Lag 36 ms VOT, and Long Lag 92 ms VOT for all children ( $n = 40$ )

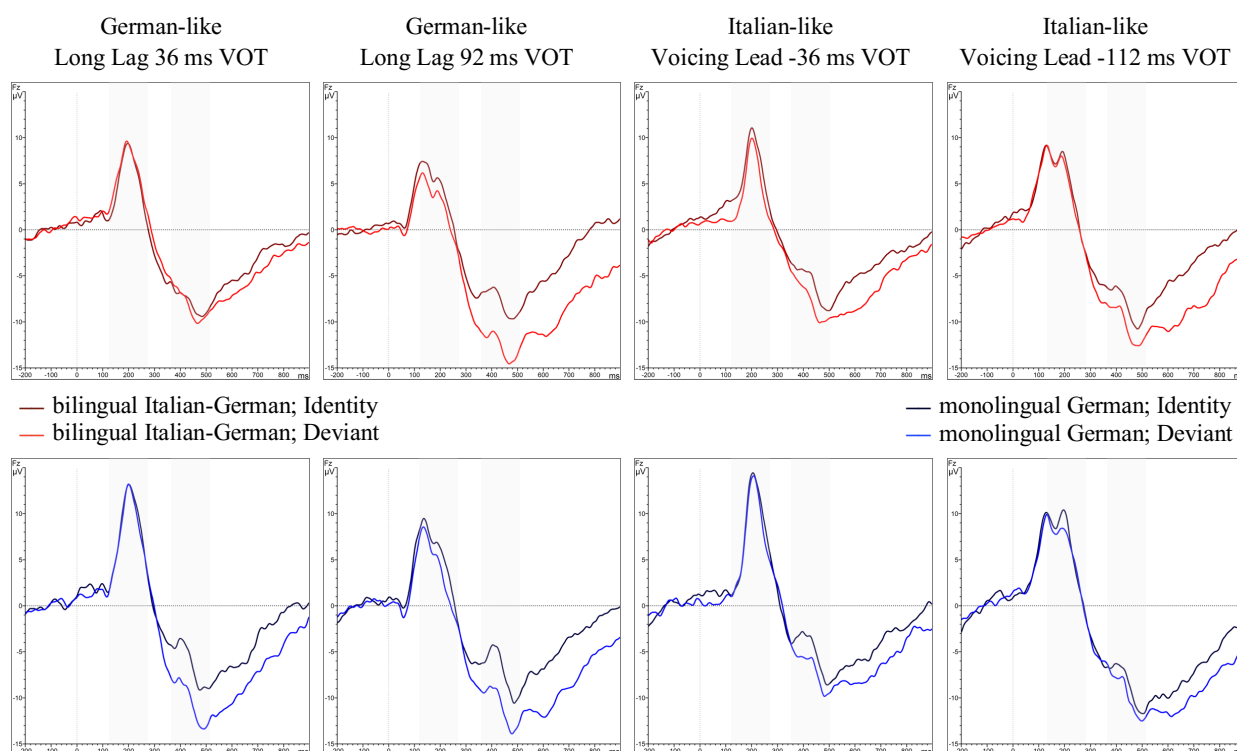
been reported to show the largest amplitude MMRs in previous studies (e.g., Yu et al., 2019). In order to reduce noise unrelated to the stimulus I averaged across the ERPs recorded at F3, Fz, and F4 for each participant to create the MMR measure used in statistical analyses.

#### 2.4.5 Procedure

Each child was tested individually in a quiet room in one of the kindergartens where they were recruited or at the university. Three monolingual German participants' behavioral language skills were assessed in their home and in the case of one bilingual child, the entire testing protocol (behavioral tests and EEG) was carried out in a quiet room in their home due to Covid-19 contact restrictions. Data collection extended across 2 to 3 testing sessions (45 to 60 minutes each) and took place on 2 to 3 consecutive days. Due to Covid-19, data collection had to be interrupted abruptly for several months during the spring of 2020, and thus for 6 children there is a wider gap (of about 3-4 months) between the collection of their EEG data and their behavioral measures. Bilingual children's parents completed the LBQ and Can-Do-Questionnaire prior to their participation. All children passed a hearing screening at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz (pure tone threshold, 25 dB HL) immediately preceding the EEG recording.

#### 2.4.6 Data Analysis

Based on visual inspection of children's iMMRs, two time windows of interest were selected (see grey boxes in Figure 5). These roughly matched what had previously been reported in the literature. Time window 1 (henceforth the early iMMR) between 120-280 ms (cf. 140-260 ms in Shafer et al., 2010 for monolingual English-speaking four-to-five-year-olds) and time window 2 (henceforth the late iMMR) between 360-520 ms (cf. 401-490 ms in Datta, Shafer, Morr, Kurtzberg, & Schwartz, 2010, although for a group of older children).



**Figure 5.** Bilingual and monolingual children’s brain responses to the different Deviant Stimuli compared to their respective Identities. Bilingual’s brain responses are plotted in the upper row; monolingual’s brain responses are plotted in the lower row. From left to right for German-like Long Lag 36 ms VOT, German-like Long Lag 92 ms VOT, Italian-like Voicing Lead -36 ms VOT, and Italian-like Voicing Lead -112 ms VOT. The grey boxes indicate the time windows of analysis. Early iMMR 120-280 ms and Late iMMR 360-520 ms.

To determine whether the ERPs to the deviants and their respective identity-control stimulus differed (see Figure 5), for each deviant, for both the early iMMR and the late iMMR time window respectively, a three-way ANOVA with stimulus (identity vs. deviant) and time as the within-subject measures and group (bilingual vs. monolingual) as the between-subject variable was performed. I tested four successive 40-ms time intervals between 120 and 280 ms to identify when the early iMMR began and another four 40 ms time intervals between 360 and 520 ms to determine the onset of the late iMMR. Significant effects of Stimulus (i.e., the difference between children’s ERP amplitude to a deviant compared to its identity) at time window 2 were followed up with another three-way ANOVA with difficulty level (easy vs. difficult) and target language (Italian-like Voicing Lead vs. German-like Long Lag) as the within-subject measures and group (bilingual vs. monolingual) as the between-subject variable.

Pearson’s  $r$  correlations (for children’s relative amount of language input and output in %) and nonparametric Spearman  $\rho$  correlations (for parents’ self-reported language skills) were run to determine whether there was a relationship between children’s language experience and the mean amplitude of their iMMR.



Considering that not all production data was normally distributed and the small sample size of the control group of monolingual Italian children (only 5 participants), nonparametric Mann-Whitney U tests were run in order to determine whether bilingual children differed from their monolingual peers in terms of the production of language-specific VOT distinctions. Further, nonparametric Friedman and post-hoc Wilcoxon Signed Rank Tests were used to examine differences in VOT production across the different categories within the group of Italian-German bilinguals.

In the group of bilingual children, Spearman *rho* correlations were run to determine the relationship between children’s language experience (German/Italian input and output, and input quality) and their mastering of monolingual-like target VOT productions.

Finally, to examine the relationship between the perception and production of language-specific VOT, Pearson’s correlations between the amplitude and latency of the iMMR and children’s word initial realization of the respective target VOT were run.

No Bonferroni correction was applied when the analysis was based on a priori hypotheses, nor when it involved a set of mutually-correlated variables. In all other cases, Bonferroni correction and the value of alpha are specified in the results section.

## 2.6 Results

### 2.5.1 Description of the Sample

Children’s age did not differ across the three groups ( $p = .696$ ). The group of bilingual Italian-German children differed significantly from their monolingual peers in their relative amount of German and Italian input and output (German input:  $t(22) = 11.435, p < .001$ , German output:  $t(22) = 7.099, p < .001$ , Italian input:  $t(22) = 14.724, p < .001$ , Italian output:  $t(22) = 12.072, p < .001$ ); see Table 4 for an overview of bilingual children’s relative amount of language input and output (for monolingual German children both German input and output  $M = 100.00, SD = .00$ ; for monolingual Italian children Italian input and output  $M = 100.00, SD = .00$ ).

**Table 4.** Overview of bilingual Italian-German speaking children’s current language experience as assessed with the LBQ. Measures of relative amount of language input and output are displayed in percent (%). Due one family failing to return the completed questionnaire  $n = 23$ .

Relative amount of current language input	Italian	$M = 43.71, SD = 18.33$
	German	$M = 56.29, SD = 18.33$
Relative amount of current language output	Italian	$M = 37.03, SD = 25.02$
	German	$M = 62.97, SD = 25.02$

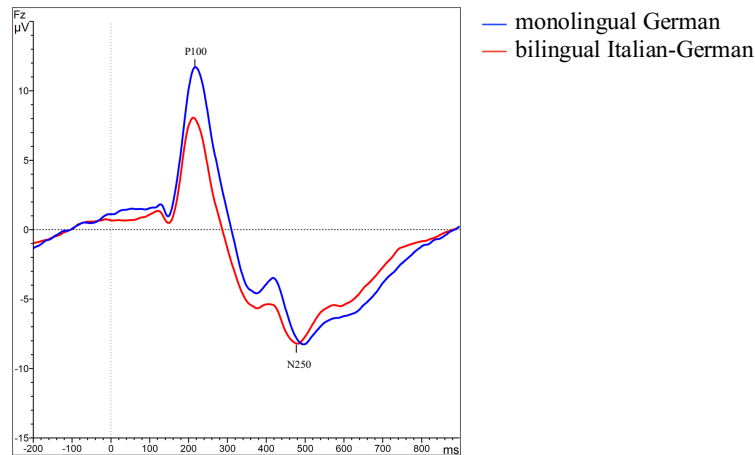
None of the measures of German language performance in the LiSe-DaZ differed significantly between monolingual German children and bilingual Italian-German children and neither did their CPM scores ( $p_s \geq .126$ ; see Table 5).

**Table 5.** Overview of children’s German language performance assessed with the LiSe-DaZ and nonverbal intelligence assessed with the CPM according to group (monolingual German vs. bilingual Italian-German). For the LiSe-DaZ subtests verb placement and subject-verb-agreement the maximum score is 4; for the subtests word classes and case markings T-scores are displayed.

		Monolingual German ( $n = 16$ )	Bilingual Italian-German ( $n = 24$ )
LiSe-DaZ	verb placement	$M = 4.00, SD = .00$	$M = 3.71, SD = .46$
	subject-verb-agreement	$M = 3.56, SD = .96$	$M = 3.54, SD = .98$
	word classes	$M = 51.76, SD = 3.71$	$M = 51.84, SD = 8.77$
	case markings	$M = 56.81, SD = 10.12$	$M = 55.29, SD = 14.12$
CPM	raw scores	$M = 16.56, SD = 3.43$	$M = 16.17, SD = 5.33$

### 2.5.2 Brain Responses to VOT Differences

All children showed typical ERPs to all VOT stimuli, consisting of a large initial positivity (P100) followed by a negativity (N250) (cf. Shafer et al., 2010); for instance, see Figure 6 for monolinguals’ versus bilinguals’ ERP to the Short Lag (0 ms VOT) standard stimulus. An independent samples t-test revealed a significant difference in children’s mean P100 amplitude ( $t(38) = 4.456, p < .001$ ). On average bilingual participants ( $n = 24$ ) showed a more negative P100 response ( $M = 3.38, SD = 1.90$ ) than their monolingual peers ( $n = 16, M = 6.16, SD = 1.98$ ). Examination of monolinguals’ versus bilinguals’ mean ERP amplitudes to the different stimuli in MMR relevant time windows (see Table A2 and A3 in Appendix A) further shows that in general bilingual children’s brain responses were more negative than those of their monolingual peers.



**Figure 6.** Children's ERP at Fz to the standard (0 ms VOT) stimulus used in both the easy and the difficult EEG paradigm.

For the mixed three-way ANOVAs comparing the ERPs to the deviants to their identity stimuli, outliers of 3 *SDs* above/below the mean were excluded from the analyses. The assumption of normality was assessed using Shapiro-Wilks test. As  $p_s > .05$  the null hypothesis that the data were normally distributed was accepted. Levene's test confirmed homogeneity of variances ( $p_s > .05$ ). The assumption of covariance of matrices was also met. Only the assumption of sphericity for variables with more than two levels (i.e., time) was not met. Thus, Greenhouse-Geisser correction was used where applicable.

**Early iMMR.** An overview of children's mean ERP amplitudes to the various stimuli can be found in Table A2 in Appendix A. For the Difficult VOT Long Lag deviant, there was no significant effect of stimulus (deviant vs. identity;  $p = .731$ ) but a significant effect for time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.65, 57.85) = 82.152, p < .001, \eta_p^2 = .701$ ). Additionally, there was a significant effect of group (monolinguals vs. bilinguals;  $F(1, 35) = 5.184, p = .029, \eta_p^2 = .129$ ). Generally, independent of stimulus, bilinguals showed more negative ERPs than their monolingual peers across all time points (bilinguals  $M = 5.48, SE = .64$ ; monolinguals  $M = 7.78, SE = .78$ ). Furthermore, there was a significant interaction effect of time x group,  $F(1.65, 57.85) = 4.921, p = .015, \eta_p^2 = .123$ . The results of a post-hoc independent samples *t*-test showed that the difference between monolinguals and bilinguals was non-significant at time point 1 (120-160 ms;  $p = .859$ ) and time point 2 (160-200 ms;  $p = .073$ ) but significant at time point 3 (200-240 ms;  $t(35) = 2.877, p = .007$ ) as well as at time point 4 (240-280 ms;  $t(35) = 2.156, p = .038$ ). The monolinguals showed a more positive response than

the bilinguals (time point 3: monolinguals  $M = 11.68$ ,  $SE = 1.03$ ; bilinguals  $M = 7.83$ ,  $SE = .85$ ; time point 4 monolinguals  $M = 6.03$ ,  $SE = 1.15$ ; bilinguals  $M = 2.82$ ,  $SE = .95$ ).

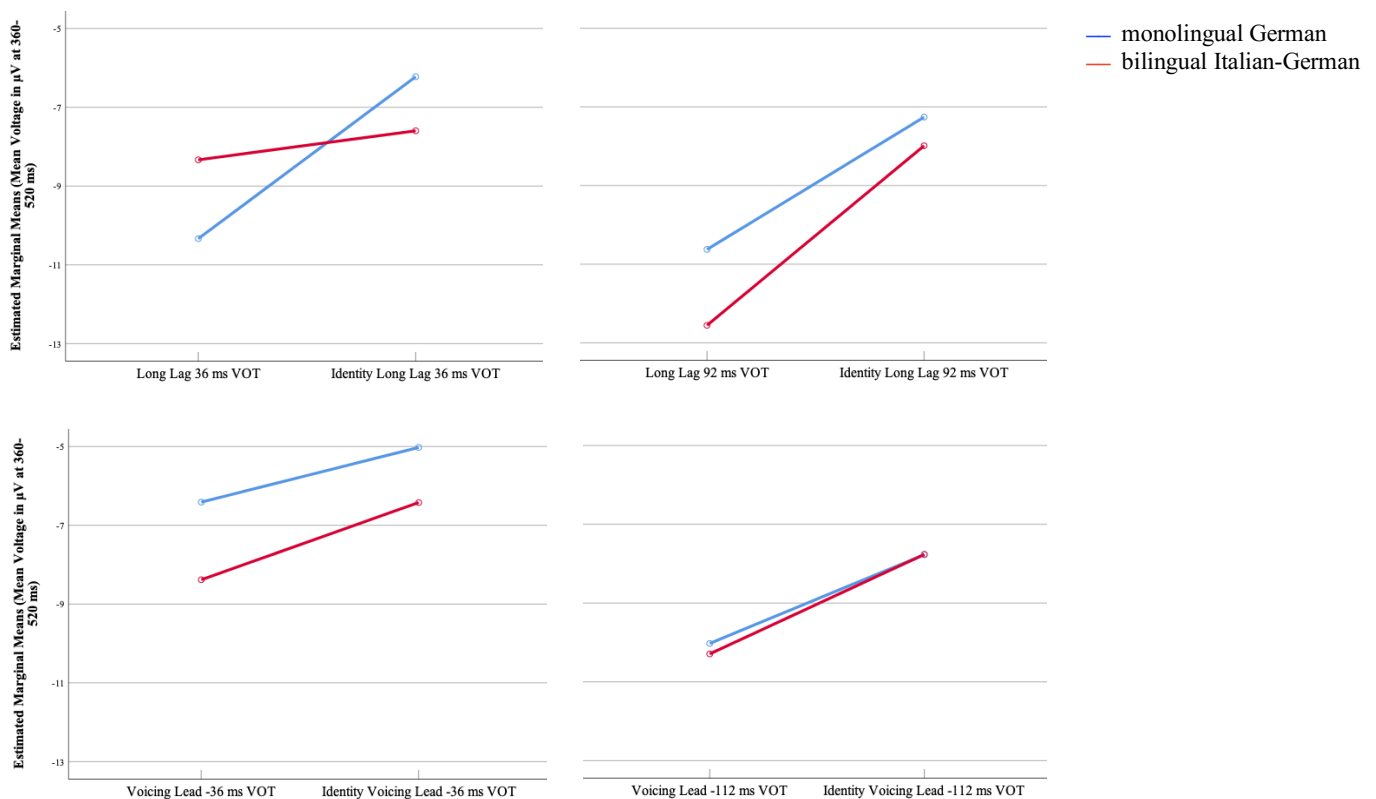
For the Easy VOT Long Lag deviant, there was no significant effect of stimulus (deviant vs. identity;  $p = .083$ ) but a significant effect for time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.78,62.58) = 109.127$ ,  $p < .001$ ,  $\eta_p^2 = .757$ ). There was no significant effect of group (monolinguals vs. bilinguals;  $p = .293$ ). No significant interactions between the three factors were found ( $p_s > .156$ ).

For the Difficult VOT Voicing Lead deviant, there was a marginally significant effect of stimulus (deviant vs. identity;  $F(1,34) = 4.033$ ,  $p = .053$ ,  $\eta_p^2 = .106$ ). All children showed more negative ERPs to the deviant ( $M = 6.06$ ,  $SE = .49$ ) than to its identity ( $M = 7.05$ ,  $SE = .49$ ). Further, there was a significant effect for time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.75,59.32) = 120.333$ ,  $p < .001$ ,  $\eta_p^2 = .780$ ). Additionally, there was a significant effect of group (monolinguals vs. bilinguals;  $F(1,34) = 11.274$ ,  $p = .002$ ,  $\eta_p^2 = .249$ ). Generally, independent of stimulus, bilinguals showed more negative ERPs than their monolingual peers across all time points (bilinguals  $M = 5.12$ ,  $SE = .54$ ; monolinguals  $M = 8.01$ ,  $SE = .67$ ). Finally, there was a significant interaction effect for time x group ( $F(1.75,59.32) = 11.521$ ,  $p < .001$ ,  $\eta_p^2 = .253$ ). The results of a post-hoc independent samples t-test showed that the difference between monolinguals and bilinguals was non-significant at time point 1 (120-160 ms;  $p = .823$ ) but significant at time point 2 (160-200 ms;  $t(34) = 2.044$ ,  $p = .049$ ; monolinguals  $M = 9.37$ ,  $SE = .84$ , bilinguals  $M = 7.20$ ,  $SE = .65$ ), time point 3 (200-240 ms;  $t(34) = 4.192$ ,  $p < .001$ ; monolinguals  $M = 13.36$ ,  $SE = .65$ ; bilinguals  $M = 8.53$ ,  $SE = .82$ ) as well as at time point 4 (240-280 ms;  $t(34) = 4.098$ ,  $p < .001$ ; monolinguals  $M = 7.28$ ,  $SE = .66$ ; bilinguals  $M = 2.55$ ,  $SE = .82$ ); bilinguals showed significantly more negative ERPs than their monolingual peers independent of stimulus (deviant vs. identity).

For the Easy VOT Voicing Lead deviant, there was no significant effect of stimulus (deviant vs. identity;  $p = .179$ ) but a significant effect for time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.53,54.90) = 82.152$ ,  $p < .001$ ,  $\eta_p^2 = .692$ ). There was no significant effect of group (monolinguals vs. bilinguals;  $p = .189$ ). Neither were there any significant interactions between the three factors ( $p_s > .301$ ).

**Late iMMR.** An overview of children's mean ERP amplitudes to the various stimuli can be found in Table A3 in Appendix A For the Difficult VOT Long Lag deviant, there was a significant effect of stimulus (deviant vs. identity;  $F(1,35) = 14.491$ ,  $p = .001$ ,  $\eta_p^2 = .293$ ), and a significant effect for time (360-400 ms vs. 400-440 ms vs. 440-480 ms vs. 480-520 ms;

$F(1.59,55.47) = 43.958, p < .001, \eta_p^2 = .557$ ) but no significant effect of group (monolinguals vs. bilinguals;  $p = .858$ ). Finally, there was a significant interaction effect of stimulus x group,  $F(1,35) = 7.021, p = .012, \eta_p^2 = .167$  (monolinguals: deviant  $M = -10.34, SE = 1.49$ , identity  $M = -6.23, SE = 1.37$ ; bilinguals: deviant  $M = -8.334, SE = 1.23$ , identity  $M = -7.59, SE = 1.13$ ; see Figure 7). The results of a post-hoc paired samples  $t$ -test showed that the difference between children's ERP amplitude to the deviant and its identity was significant for monolinguals but not for the group of bilinguals (monolinguals:  $t(14) = -4.242, p = .001$ ; bilinguals:  $t(21) = -.901, p = .378$ ).



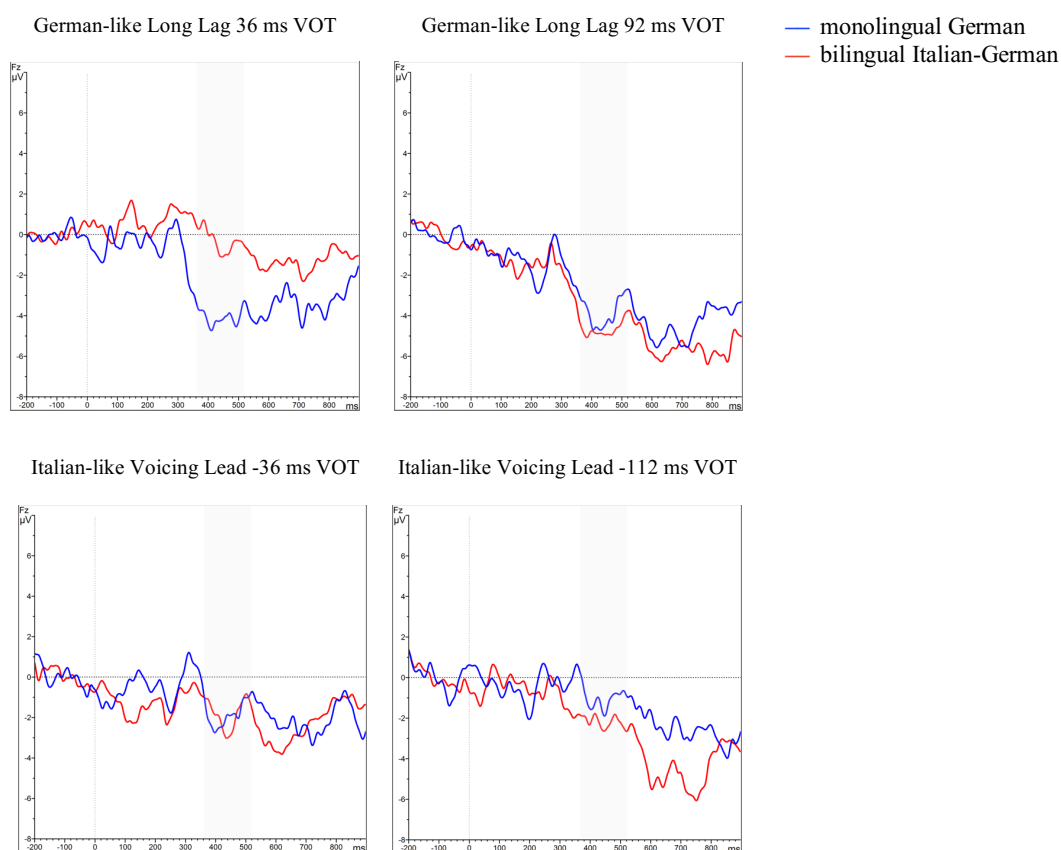
**Figure 7.** Children's mean ERP voltage (averaged across electrode sites F3, Fz, and F4) to each deviant vs. its respective identity across the time window of 360-520 ms after stimulus onset according to group (monolinguals vs. bilinguals). German-like Long Lag plotted in the upper row; from left to right for Long Lag 36 ms VOT and Long Lag 92 ms VOT; Italian-like Voicing Lead plotted in the lower row; from left to right for Voicing Lead -36 ms VOT, and Voicing Lead -112 ms VOT.

For the Easy VOT Long Lag deviant, there was a significant effect of stimulus (deviant vs. identity;  $F(1,35) = 25.026, p < .001, \eta_p^2 = .417$ ) and a significant effect of time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.58,55.39) = 48.770, p < .001, \eta_p^2 = .582$ ) but no significant effect of group (monolinguals vs. bilinguals;  $p = .492$ ). Furthermore, there was a significant interaction effect of time x group,  $F(1.58,55.39) = 3.547, p = .046, \eta_p^2 = .092$ . The results of a post-hoc independent samples  $t$ -test showed that the difference between monolinguals and bilinguals was non-significant at any of the four time windows ( $p_s > .262$ ).

For the Difficult VOT Voicing Lead deviant, there were significant effects of stimulus (deviant vs. identity;  $F(1,34) = 4.872, p = .034, \eta_p^2 = .125$ ) and time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(1.53, 52.14) = 49.245, p < .001, \eta_p^2 = .601$ ), but not for group (monolinguals vs. bilinguals;  $p = .592$ ). Furthermore, there were no significant interaction effects for any of the factors ( $p_s > .225$ ). The interaction between stimulus x time approached significance ( $p = .074$ ). A paired samples *t*-test with Bonferroni correction applied (new alpha-level set at .0125) revealed that the difference between the Difficult Voicing Lead deviant and its identity was significant at time point 2 (400-440 ms;  $t(35) = -2.041, p = .006$ ; deviant  $M = -6.38, SE = .92$ ; identity  $M = -4.10, SE = .85$ ), and time point 3 (440-480 ms;  $t(35) = -2.769, p = .009$ ; deviant  $M = -9.09, SE = .92$ ; identity  $M = -6.86, SE = .87$ ), but not at time point 1 (360-400 ms;  $p = .090$ ) and time point 4 (480-520 ms;  $p = .174$ ).

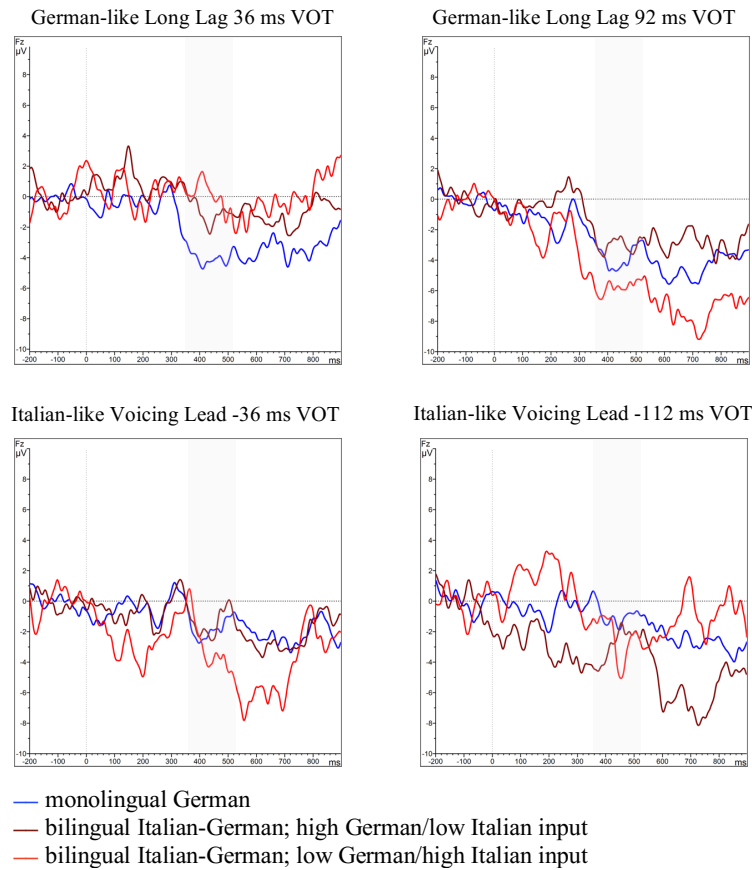
For the Easy VOT Voicing Lead deviant, there was a significant effect of stimulus (deviant vs. identity;  $F(1,33) = 15.957, p < .001, \eta_p^2 = .326$ ) and a significant effect of time (120-160 ms vs. 160-200 ms vs. 200-240 ms vs. 240-280 ms;  $F(2.25, 74.39) = 58.499, p < .001, \eta_p^2 = .63.9$ ). There was no significant effect of group (monolinguals vs. bilinguals;  $p = .940$ ). Furthermore, there was a significant interaction effect of stimulus x time,  $F(1.69, 55.62) = 4.751, p = .017, \eta_p^2 = .126$ . A paired samples *t*-test with Bonferroni correction applied (new alpha-level set at .0125) revealed that the difference between the deviant and its identity was significant at time point 1 (360-400 ms;  $t(37) = -2.626, p = .0125$ ; deviant  $M = -8.23, SE = 1.03$ ; identity  $M = -6.32, SE = .99$ ), time point 2 (400-440 ms;  $t(37) = -3.127, p = .003$ ; deviant  $M = -8.58, SE = 1.02$ ; identity  $M = -4.19, SE = .85$ ), and time point 3 (440-480 ms;  $t(37) = -3.405, p = .002$ ; deviant  $M = -11.48, SE = 1.05$ ; identity  $M = -9.67, SE = .98$ ), but not at time point 4 (480-520 ms;  $p = .026$ ).

As the difference between the ERPs to the deviants and their respective identities was only significant in the later time window (360-520 ms after stimulus onset), the iMMR within that time range was explored further and compared between the two groups (see Figure 8 for monolinguals' vs. bilinguals' iMMRs). A three-way mixed ANOVA with language (German-like VOT vs. Italian-like VOT) and difficulty level (easy vs. difficult contrast) as the within-subject variables and group (monolinguals vs. bilinguals) as the between-subject variable revealed a significant interaction for language x difficulty level x group  $F(1,30) = 4.637, p = .039, \eta_p^2 = .134$  (see Table A2 in Appendix A for children's mean amplitudes and *SE* values for the late iMMR).



**Figure 8.** Monolinguals' vs. bilinguals' iMMRs for each deviant at Fz. German-like Long Lag plotted in the upper row; from left to right for Long Lag 36 ms VOT and Long Lag 92 ms VOT; Italian-like Voicing Lead plotted in the lower row; from left to right for Voicing Lead -36 ms VOT, and Voicing Lead -112 ms VOT. The grey box indicates the time window of interest (late iMMR at 360-520 ms).

To follow up on the question concerning how much language exposure is necessary for bilingual children to show processing that is similar to their monolingual peers, a sub-selection of children in the group of bilinguals was made. According to Cattani and colleagues (2014) bilingual Spanish-English toddlers who received 60% or more of their language input in English matched their monolingual English peers with regards to their English language skills. Thus, in the current study, children were grouped into a high German (i.e., low Italian) versus low German (i.e., high Italian) experience cohort, as quantified with the LBQ (high German = more than 60% German input,  $n = 10$  in the difficult condition,  $n = 8$  in the easy condition; low German = less than 40% German input,  $n = 3$  in the difficult condition,  $n = 4$  in the easy condition; bilinguals with more balanced language input were not included in this selection) to compare their iMMRs to the different VOT deviants to those of their monolingual German peers ( $n = 15$  in the difficult condition,  $n = 16$  in the easy condition) (see Figure 9).



**Figure 9.** Monolinguals' vs. high German/low Italian vs. low German/high Italian input bilinguals' iMMRs for each deviant at Fz. German-like Long Lag plotted in the upper row; from left to right for Long Lag 36 ms VOT and Long Lag 92 ms VOT; Italian-like Voicing Lead plotted in the lower row; from left to right for Voicing Lead -36 ms VOT, and Voicing Lead -112 ms VOT. The grey box indicates the time window of interest (late iMMR at 360-520 ms).

Examination of children's mean iMMR amplitudes averaged across Fz, F3 and F4 within the time window of 360-520 ms (see Table 6 for an overview) further shows that depending on the target language and the magnitude of the acoustic difference between the standard and the deviant, monolingual German versus bilingual high German/low Italian input versus bilingual low German/high Italian input children processed the different VOT stimuli differently. Of the three groups, monolingual German children showed the most negative MMR to the Difficult Long Lag 36 ms deviant. The high German/low Italian input showed a negative MMR, but less so than the German Monolingual group. The low German/ high Italian input children showed the least negativity and even showed a positive peak within the target time window. The monolingual German and high German/low Italian input children showed a similar amplitude MMR to the Difficult Voicing Lead -36 ms deviant, whereas the low German/high Italian input children showed the most negativity MMR. Note that due to the small group sizes, no group statistical analyses were conducted. However, further examination of children's individual



mean amplitude brain responses in the time range of interest to the different deviants (i.e., late iMMR 360-520 ms post stimulus onset) showed that in the monolingual subgroup, 32.25% ( $n = 5$ ) of the participants showed a positive iMMR to the Easy Voicing Lead Deviant (mean amplitude ranging from  $-9.85 \mu\text{V}$  to  $6.99 \mu\text{V}$ ), 25.0% ( $n = 4$ ) showed a positive iMMR to the Easy Long Lag Deviant (mean amplitude ranging from  $-12.36 \mu\text{V}$  to  $2.64 \mu\text{V}$ ), 42.86% ( $n = 6$ ) showed a positive iMMR to the Difficult Voicing Lead Deviant (mean amplitude ranging from  $-12.17 \mu\text{V}$  to  $6.31 \mu\text{V}$ ), and 12.0% ( $n = 3$ ) showed a positive iMMR to the Difficult Long Lag Deviant (mean amplitude ranging from  $-9.72 \mu\text{V}$  to  $4.36 \mu\text{V}$ ). In the subgroup of high German/low Italian input bilinguals, 14.29% ( $n = 1$ ) of the participants showed a positive iMMR to the Easy Voicing Lead Deviant (mean amplitude ranging from  $-8.41 \mu\text{V}$  to  $.28 \mu\text{V}$ ), 14.29% ( $n = 1$ ) showed a positive iMMR to the Easy Long Lag Deviant (mean amplitude ranging from  $-13.48 \mu\text{V}$  to  $1.81 \mu\text{V}$ ), 33.33% ( $n = 3$ ) showed a positive iMMR to the Difficult Voicing Lead Deviant (mean amplitude ranging from  $-3.73 \mu\text{V}$  to  $5.59 \mu\text{V}$ ), and 33.3% ( $n = 3$ ) showed a positive iMMR to the Difficult Long Lag Deviant (mean amplitude ranging from  $-6.39 \mu\text{V}$  to  $2.15 \mu\text{V}$ ). Finally, in the subgroup of low German/high Italian input bilinguals, 50.0% ( $n = 2$ ) of the participants showed a positive iMMR to the Easy Voicing Lead Deviant (mean amplitude ranging from  $-7.92 \mu\text{V}$  to  $4.98 \mu\text{V}$ ), 25.0% ( $n = 1$ ) showed a positive iMMR to the Easy Long Lag Deviant (mean amplitude ranging from  $12.56 \mu\text{V}$  to  $3.93 \mu\text{V}$ ), none of the children showed a positive iMMR to the Difficult Voicing Lead Deviant (mean amplitude ranging from  $-5.62 \mu\text{V}$  to  $-2.06 \mu\text{V}$ ), and 33.33% ( $n = 1$ ) showed a positive iMMR to the Difficult Long Lag Deviant (mean amplitude ranging from  $-4.53 \mu\text{V}$  to  $7.38 \mu\text{V}$ ).

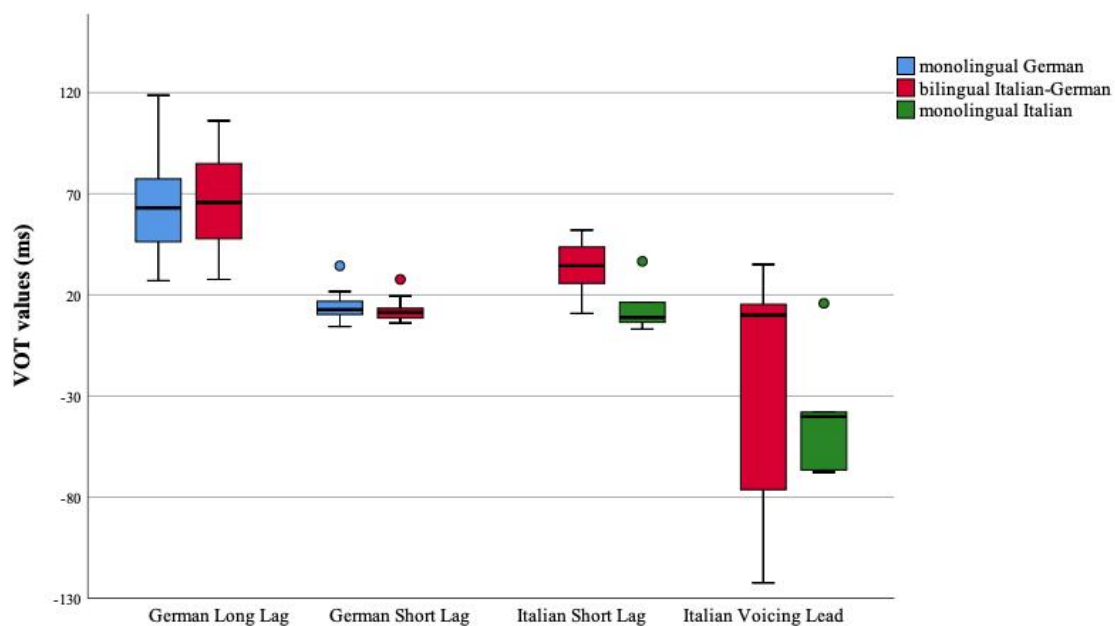
**Table 6.** Overview of children’s mean late iMMR amplitudes averaged across Fz, F3, and F4 according to their language input situation.

		Bilinguals low German/ high Italian input	Bilinguals high German/ low Italian input	Monolinguals
German-like	iMMR “easy” (92 ms VOT)	$M = -5.19,$ $SD = 6.82$	$M = -3.56,$ $SD = 4.46$	$M = -3.37,$ $SD = 1.71$
	Long Lag iMMR “difficult” (36 ms VOT)	$M = -.54,$ $SD = 9.86$	$M = -1.73,$ $SD = 2.96$	$M = -4.11,$ $SD = 3.75$
Italian-like	iMMR “easy” (-112 ms VOT)	$M = -.26,$ $SD = 5.53$	$M = -3.12,$ $SD = 3.08$	$M = -1.72,$ $SD = 3.96$
	Voicing Lead iMMR “difficult” (-36 ms VOT)	$M = -3.86,$ $SD = 1.78$	$M = -1.25,$ $SD = 3.99$	$M = -1.39,$ $SD = 5.14$

Pearson  $r$  correlation analyses yielded no significant relationships between bilingual children's language experience (i.e., their relative amount of Italian vs. German language input and output) and any of their mean late iMMR amplitudes ( $r_s$  ranging from  $-.257$  to  $.128$ ,  $p_s > .261$ ).

### 2.5.3 VOT Production

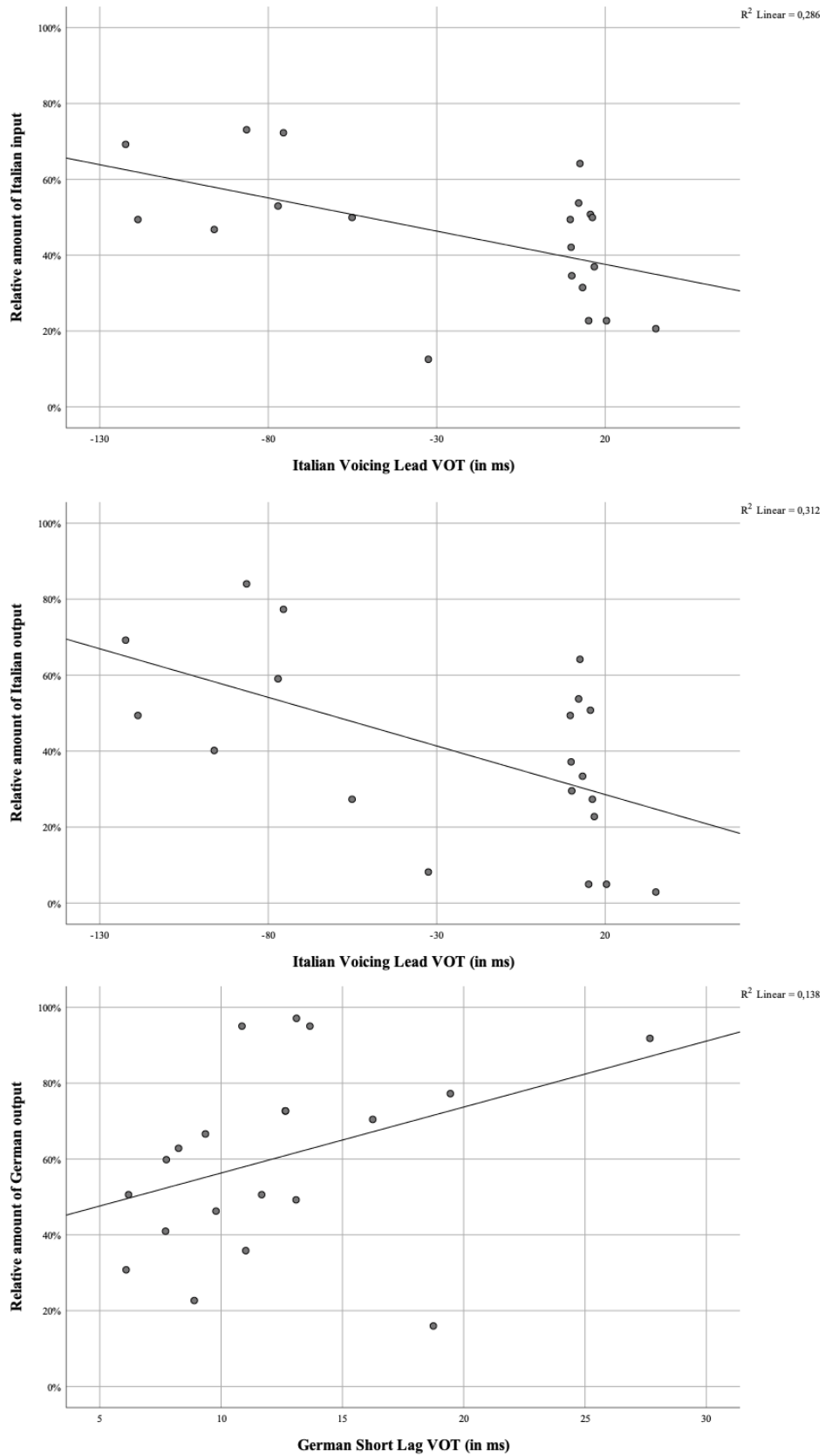
Figure 10 provides an overview of children's realizations of the different VOT categories according to group. Mann Whitney-U tests revealed that bilingual Italian-German children did not differ significantly from monolingual German children neither with respect to their productions of German Long Lag (monolingual German  $M = 63.95$ ,  $SD = 24.92$ ; bilingual Italian-German  $M = 66.45$ ,  $SD = 21.57$ ;  $p = .582$ ) nor with respect to their realizations of German Short Lag (monolingual German  $M = 14.23$ ,  $SD = 6.92$ , bilingual Italian-German  $M = 12.24$ ,  $SD = 5.19$ ;  $p = .305$ ). Comparing bilinguals' realizations of Italian Short Lag and Voicing Lead in Italian words with those of their monolingual Italian peers, another Mann Whitney-U test yielded a significant difference for the Italian Short Lag category (monolingual Italian  $M = 14.27$ ,  $SD = 13.36$ , bilingual Italian-German  $M = 33.93$ ,  $SD = 13.18$ ;  $U = 15.00$ ,  $Z = -2.38$ ,  $p = .017$ ) with a moderate effect size,  $r = -.48$ . By contrast, the two groups did not differ significantly in their realization of Italian Voicing Lead (monolingual Italian  $M = -39.29$ ,  $SD = 33.86$ , bilingual Italian-German  $M = -23.87$ ,  $SD = 53.08$ ;  $p = .59$ ).



**Figure 10.** Children's mean VOT values for the different VOT categories according to group (monolingual German vs. bilingual Italian-German vs. monolingual Italian).

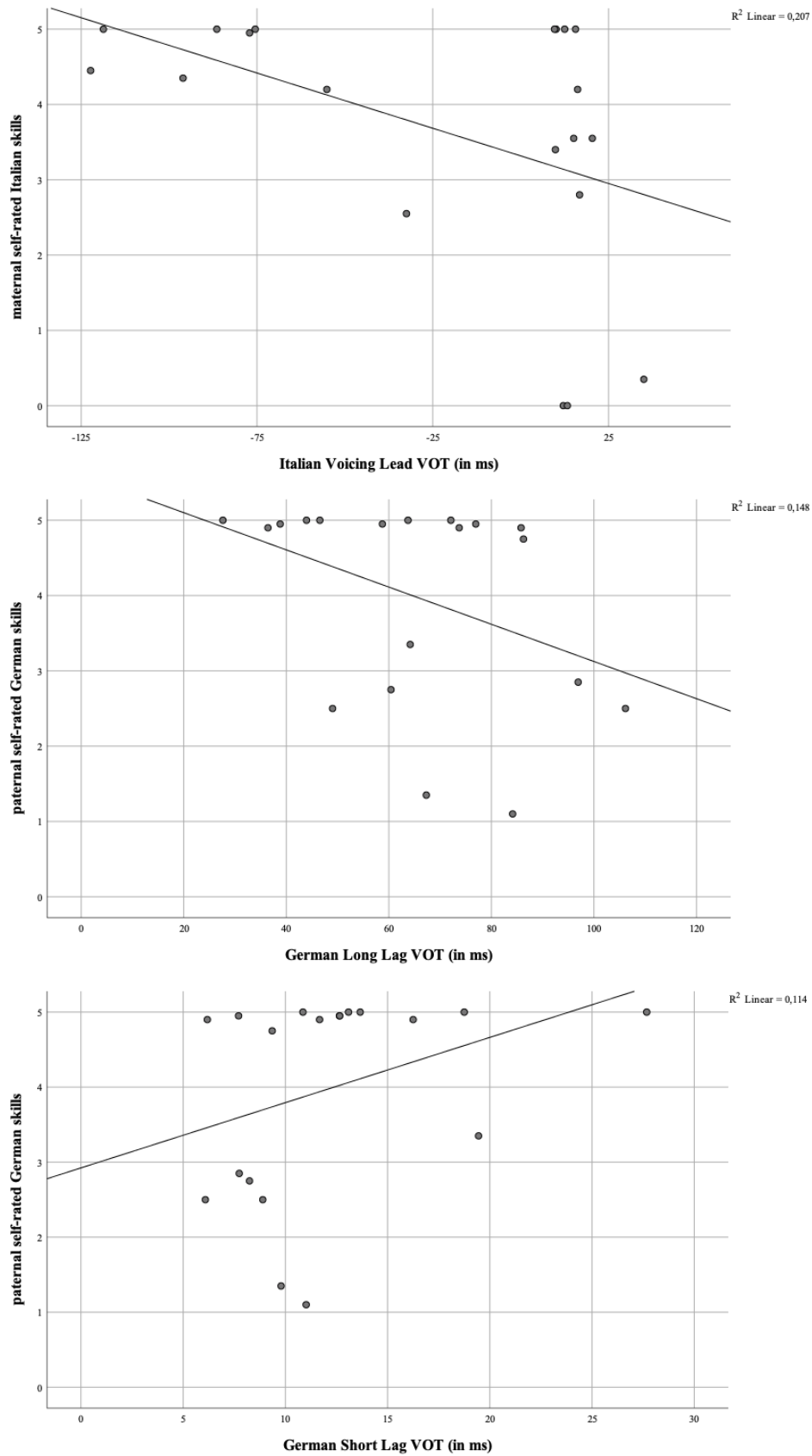
Within the group of bilinguals, a nonparametric Friedman's test revealed that VOT values differed significantly according to the target category,  $\chi^2(3) = 47.40, p < .001$ ). Due to multiple comparisons, post hoc Wilcoxon signed-rank tests analysis was conducted applying the Bonferroni correction, resulting in the significance level set at  $p < .01$ . There was a significant difference between bilingual children's realizations of German Long Lag versus German Short Lag ( $Z = -3.92, p < .001$ ), Italian Voicing Lead versus Italian Short Lag ( $Z = -3.92, p < .001$ ), German Short Lag versus Italian Short Lag ( $Z = -3.77, p < .001$ ), and German Long Lag versus Italian Short Lag (i.e., the two voiceless categories;  $Z = -3.62, p < .001$ ), but not for German Short Lag versus Italian Voicing Lead (i.e., the two voiced categories;  $p = .073$ ). Moreover, it was noted that within the bilingual Italian-German-speaking group, there was considerable variance with respect to the realization of Voicing Lead. When comparing variability in the Voicing Lead category with variability in the Long Lag category (where there is also a possibility for variability), a Levene F-test was significant ( $F(1,38) = 29.24, p < .001$ ), indicating unequal variance between these two VOT categories.

Spearman *rho* correlation analyses revealed a significant relationship between the amount of children's Italian input and their realization of Italian Voicing Lead ( $rho = -.496, p = .026$ ). Furthermore, significant correlations were found for children's active use of Italian (i.e., the amount of Italian output with their productions of Italian Voicing Lead ( $rho = -.633, p = .003$ ). Children's active use of German (i.e., the amount of German output) was significantly correlated with their realization of German Short Lag ( $rho = .464, p = .039$ ); see Figure 11.



**Figure 11.** Scatter plots visualizing the relationship between bilingual Italian-German children's language experience and their realization of VOT ( $n = 20$ ). From top to bottom for Italian input and Italian Voicing Lead, Italian output and Italian Voicing Lead, and German output and German Short Lag.

Moreover, global measures of parental Italian versus German language proficiency (as assessed through the Can-Do-Questionnaire) were correlated with children's VOT realizations. Specifically, maternal self-rated Italian skills correlated with children's productions of Italian Voicing Lead ( $\rho = -.452, p = .045$ ), while paternal self-rated German skills correlated with children's productions of German Long Lag ( $\rho = -.484, p = .036$ ), and German Short Lag ( $\rho = -.522, p = .022$ ); see Figure 12.



**Figure 12.** Scatter plots visualizing the relationship between bilingual Italian-German children’s parents’ language proficiency ( $n = 20$ ). From top to bottom for maternal Italian skills and Voicing Lead in Italian word onsets, paternal German skills and Long Lag in German word onsets, and paternal German skills and Short Lag in German word onsets.

## 2.7 Discussion

This study explored the relationship between bilingual children's language experience and the development of automaticity in neural speech sound discrimination (as indexed by the iMMR) and the influence of the degree of stimulus difference. It has been previously suggested that automaticity of speech perception in (monolingual) children is not robustly established for fine-grained phonetic contrasts until four years of age (Shafer et al., 2010). More specifically, it has been proposed that neural speech sound discrimination is not initially automatic because it takes time and experience to establish robust, selective perceptual routines (SPRs) even for native-language phonological categories (Strange, 2011). During a passive listening task, this study looked at four-to-five-year-old monolingual German and bilingual Italian-German children's brain responses when processing natural German- versus Italian-like VOT stimuli that differed in their magnitude of acoustic difference from the standard. The results confirmed my hypotheses. Specifically, bilingual children differed from their monolingual peers with respect to the iMMR amplitudes, supporting language-experience dependent effects in (native) speech sound processing (RQ1, Hypothesis 1). Generally, as a group, bilinguals showed a tendency towards more negative ERPs compared to their monolingual peers. This finding may indicate increased (in)voluntary attention to the acoustic signal (RQ1, Hypothesis 2; cf. Yu et al., 2019). Furthermore, high-Italian-input bilingual children who fell below a threshold of 40% of current German input did show immature signs (more positive MMR) of processing the difficult German VOT contrast (RQ2, Hypothesis 3). Additionally, this study looked at children's VOT productions and found that Italian-German bilinguals matched their monolingual German peers, whereas when compared to age-matched Italian controls their Italian-specific VOT realizations were a lot more variable (RQ3, Hypothesis 4). Finally, measures of children's relative amount of Italian versus German input and output, as well as their parents' language proficiency were associated with their language-specific realizations of VOT (RQ4, Hypothesis 5). In the following, these results will be discussed in more detail. Additionally, the interplay between children's productions of VOT and their individual patterns of language-specific VOT processing will be considered.

### 2.7.1 VOT Perception

**Early versus Late iMMR.** Children's ERPs were examined in an early time window (120-280 ms; early MMR) and a late time window (360-520 ms; late iMMR) (cf. Yu et al., 2019). While there were no significant differences in amplitude between children's brain responses to the

deviants and their respective identities in the early time window, there was robust evidence of neural speech sound discrimination for all deviant stimuli in the late time window. Although the nature and functional mechanisms indexed by late MMR have not yet been fully understood, Yu and colleagues (2019) suggested that this late response may have the potential to evaluate the development of speech processing in toddlers, especially since “the MMR in the earlier time frame (150–400 ms) is often not significant to the subtle speech contrasts that are of particular interest in studies of language development” (Yu et al., 2019, p. 17). The results of this current study clearly support this claim. Alternatively, as suggested by Morr, Shafer, Kreuzer, and Kurtzberg (2002), the early negativity may have been overlapped, and thus masked, by a larger magnitude positive response, suggesting that children’s processing was still immature.

**Development of Automaticity as indexed by the MMR.** Results indicate that Italian-German bilingual four- and five-year-old children and age-matched German monolingual controls behave similarly when processing Long Lag and Voicing Lead VOT, with the exception of the Difficult Long Lag 36 ms VOT stimulus. This is in line with the proposal of Crick and Koch (1990). They suggested that highly salient differences can be processed (i.e., neurally discriminated) with fewer attentional resources than less salient distinctions. The authors further pointed out that less salient information can be made more apparent through the process of over-learning. The lack of a robust negative MMR to the Difficult long lag stimulus in bilinguals may suggest that over-learning of this subtle difference has not yet been achieved by bilinguals at four years of age. That is, bilinguals may not have yet accumulated sufficient experience with the German language to become fully automatic in processing German-like VOT (García-Sierra et al., 2016).

The lack of a bilingual advantage in processing Italian Voicing Lead VOT stimuli was unexpected. One interpretation may be that the bilingual participants were not yet automatic in discriminating the Italian VOT contrasts, considering that sufficient experience (in terms of time and amount of input) is necessary to establish automaticity of SPRs for phonological categories (Strange, 2011). This account leads to the hypothesis that, when processing speech sounds at an attention-independent level, child bilinguals show automaticity earlier in their dominant language (i.e., German in the current study) compared to their non-dominant language (Cutler, Norris, & Williams, 1987; Snijders, Kooijman, Cutler, & Hagoort, 2007). However, an alternative interpretation may be that the Italian-German children were, in fact, automatic at processing, but that the German linguistic context during the experiment primed them to process according to the German categories (Elman, Diehl, & Buchwald, 1977; Gonzales &



Lotto, 2013; Wig & García-Sierra, 2021), although Winkler, Kujala, Alku, and Näätänen (2003) found no effect of a Finnish versus Hungarian linguistic context on their adult Finnish and Hungarian participants. Nevertheless, in this relatively young age group linguistic context might be more important as children are still in the process of acquiring their two languages. Finally, one possible explanation for why monolingual German children showed neural signs of pre-attentive speech sound discrimination to the Italian Voicing Lead contrasts is that they may in fact have come across instances of voicing lead in their everyday German language input. Although German voiced plosives are usually produced in the short lag region, voicing lead has been found to be possible in the speech of German adults (Jessen, 1998).

A pMMR (cf. Friederici, Friedrich, & Weber, 2002) was only apparent in the subgroup data. One explanation for the absence of a pMMR could be that the distinction was easy enough for the MMR to dominate for all but the bilingual low-German-input children in the Long Lag difficult condition. Alternatively, the pMMR may only be present in younger children (Morr et al., 2002). Finally, it is possible that the pMMR is more prominent at F3 than at Fz and F4 (Shafer et al., 2010). Further examinations should thus focus on the brain response's topography.

**Bilingual experience and the late iMMR.** It was predicted that bilingual children's language experience (i.e., their relative amount of language input and output respectively) would modulate their indices of automatic neural discrimination. None of the measures of language experience were significantly correlated with children's iMMR, which suggests that other factors may play a role in the development of automaticity for speech sound processing (such as e.g., maturation effects, Morr et al., 2002; or attention, Datta et al., 2010). Nevertheless, when dividing the bilingual children into two subgroups of high German/low Italian input versus low German/high Italian input it was found that the group of bilinguals with relatively high rates of German input (> 60%) and the group of monolingual German children had a similar iMMR amplitude to the Voicing Lead -36 ms VOT deviant, whereas children with higher rates of Italian input (> 60%) showed a more negative response. Conversely, the same group of high Italian input (i.e., low German input) children responded with a positive peak to the German-like Long Lag 36 ms VOT stimulus within the late iMMR time window and even high German input children did not show a response as negative as their monolingual German peers.

Taken together, these results suggest that more subtle VOT differences are more susceptible to the influence of language experience. Shafer and colleagues (2010) suggested

that with increasing age, the adult-like MMN moves earlier in latency, increases in amplitude and overlaps with, and thus reduces the amplitude of the positive MMR. This explanation indicates that four-to-five-year-old bilingual Italian-German children were not as automatic in processing the “difficult” German-like VOT contrast as their age-matched monolingual German peers. The difference between monolinguals and bilinguals for the Italian-like voicing lead was less, although the group of high Italian (i.e., low German) input children showed a more negative response to the Voicing Lead -36 ms VOT deviant, than the low input Italian (i.e., high German) and monolingual German children. This finding may indicate that the Italian-German bilinguals are not yet automatic in processing the Italian stimuli, and thus, do not yet show a clear advantage over the monolingual German children.

### 2.7.2 VOT Production

VOT realizations of the short lag-voicing lead distinction in Italian words differed from monolingual Italian control children. In contrast, VOT productions for the short and long lag German words did not differ from German monolingual children. Differences between monolingual Italian and bilingual Italian-German children in the Italian short lag and voicing lead category (especially the increased VOT for Italian Short Lag stops) are interpreted as signs of crosslinguistic influence (i.e., transfer; Paradis & Genesee, 1996) from the majority (i.e., German) to the minority (i.e., Italian) language. Specifically, the short lag-long lag distinction is in German, and not in Italian. Alternatively, according to Gawlitzek-Maiwald and Tracy (1996) bilingual Italian-German children have the advantage to make use of a contrast that is not available in the monolingual Italian context when the Italian target is more difficult to produce (Lisker & Abramson, 1964). Bilingual Italian-German children can thus increase their VOT for the Italian short lag /p/ in order to make it more distinct from the voiced category if they struggle with producing voicing lead. This is also in line with McAllister Byun and colleagues' (2016) A-Map model. This model claims that due to reduced opportunities for bilingual children to actively use their heritage language (in this case, Italian) and thus motorically practice the Italian voicing lead feature, their productions do not (yet) match the monolingual target. In contrast, the motorically easier German long lag feature (cf. Lisker & Abramson, 1964) may need less practice in order to achieve adult-like precision. Nevertheless, it is noteworthy that the group difference may also be explained by VOT characteristics in bilingual children's language input, considering that it has been found that children's VOT patterns correlate with those of their parents (Stoehr, 2018). Similar to Stoehr (2018), the bilingual children in the current study as well, may have been exposed to Italian input in which

Voicing Lead was not always employed. That is, many of the Italian-speaking parents were born in Germany and grew up fully immersed in the majority language German, and thus are considered heritage language speakers of Italian themselves. Heritage speakers may show differences from those who acquire the language in a largely monolingual environment where their native language is dominant in terms of having an accent in the heritage language (Lloyd-Smith, Einfeldt, & Kupisch, 2020). This is especially relevant, considering the significant correlation found in the current study between children's quantity and quality of Italian input and their production of voicing lead. However, note that in this study, input quality was only measured indirectly via parents' self-reports. One caveat in the current study therefore was that the quality of children's input VOT (i.e., whether or not the input matched the Italian adult monolingual-like target) was not measured directly and rather was inferred from their self-ratings in the Can-Do-Questionnaire. Therefore, no definite conclusions can be drawn about the nature of the differences between monolingual Italian and bilingual Italian-German children's realization of the Italian VOT contrast. Nevertheless, lower self-ratings may indicate that parents' heritage language phonetic realizations, in fact, differed from monolingual adult norms.

### 2.7.3 VOT Perception x Production

Measures of VOT production for Italian and German words in bilingual children support the interpretation that as a group the bilinguals were not yet automatic in processing the Italian-like VOT contrasts. Specifically, in production, the children showed mastery of German VOT but great variability in Italian voicing lead VOT (both across each child's individual productions and across the group) in addition to an increase of VOT in the Italian Short Lag category. Although bilinguals' productions of Italian versus German short lag were significantly different, it may be that despite being able to mark the distinction between their two languages productively, the perceptual boundary between long lag and short lag in bilingual children may have shifted, suggesting that they treat German short lag and Italian short lag as one perceptual category and that they assimilated the short lag and the German-like long lag "difficult" 36 ms VOT stimulus into one and the same category (cf. SLM; Flege, 1995). This interpretation of the results is in line with previous research applying a behavioral XAB perception task with bilingual German-Dutch children (Stoehr, 2018) that reported a shift bilinguals' perceptual boundaries towards higher VOT values. Unfortunately, the current study could not fully address the complex interplay between speech sound perception (including processing) and production because it did not include a behavioral perception task to establish and subsequently take into

account each child's individual perceptual boundaries between the voiced and the voiceless category in German and Italian respectively.

#### 2.7.4 Limitations and Future Directions

One of the main limitations of this study was that due to Covid-19 contact restrictions and thus a repeatedly interrupted data collection process, sample sizes especially for sub-groups (i.e., high vs. low input distinction) in the bilingual sample were relatively small which prevented the application of several statistical analyses. Thus, some of the observations made have to be interpreted with caution as they rely solely on visual inspection of the data.

Further, estimates of children's language input quality were based on a global measure obtained through parental self-reports (i.e., the Can-Do-Checklist). A direct measure of VOT input quality in German and Italian obtained through recordings of parental speech samples would have been more informative concerning the model parents provide for their children in both languages. Similarly, an indicator of children's individual perceptual boundaries in each language would have been beneficial to interpreting and explaining group differences.

Finally, including a control group of four-to-five-year-old monolingual Italian children not only in the production but also in the EEG study would have enabled me to make more explicit inferences about the development of bilingual Italian-German children's automaticity in speech sound perception in their heritage language.

Taken together, in addition to speech sound production and measures of attention-independent speech sound perception (e.g., using EEG), future studies should collect data on the quality of the target speech sounds in children's language input as well as a behavioral equivalent to measuring speech sound perception in order to estimate their individual perceptual boundaries in order to obtain a complete picture of the complex interplay between language experience and speech sound processing.

## 2.8 Conclusion

This study replicated previous findings of automaticity in native speech sound perception in monolingual children by the age of four (Shafer et al., 2010). However, for the group of bilingual children, a different developmental trajectory was observed. Even the group of high German input bilinguals differed from age-matched monolingual German children when processing the subtle German-like VOT contrast, although it might be expected that with increasing age and greater cumulative German experience bilingual children's brain responses and thus indices of automaticity would match that of their monolingual peers (cf. Datta et al.,

2019 for 8-to-10-year-old bilingual children), suggesting that language input quantity is a driving factor in the formation of SPRs (Strange, 2011). Alternatively, a shift in bilingual children's perceptual boundary between the short lag and the long lag VOT category may have also accounted for the differences between the two groups.

### 3. Bilingual Language Experience in the Context of Speech and Language Pathology<sup>5</sup>

Children's academic and future professional achievements depend to a great extent on successful language acquisition and language performance (Johnson, Beitchman, & Brownlie, 2010; Snowling, Bishop, Stothard, Chipchase, & Kaplan, 2006). Therefore, being able to detect and address atypical language development as early as possible is crucial. Furthermore, in both family and academic contexts, appropriate language competence is relevant in order for children to develop a cultural and social identity. The United Nations addressed this topic, among others, when proposing their 17 Sustainable Development Goals (SDGs; United Nations, 2015) with the ambition "to end poverty, protect the planet and ensure prosperity for everyone (...)". In this chapter, special emphasis is placed on four of these 17 goals, first and foremost "3: Good health and well-being" and "4: Quality education" which in the wider sense are further related to "8: Decent work and economic growth" and "10: Reduce inequality".

In particular, this chapter addresses the sustainability of SLPs' approaches but also attitudes when assessing and treating DLD in children who grow up in Germany acquiring more than one language.

Data provided by the statistical office of the European Union show that in Germany about 18.1% of the country's population are considered foreign-born<sup>6</sup> (Eurostat, 2021). Over the past decades, Germany has been subjected to several distinct waves of immigration. These include, among others, the resettlement of ethnic Germans from Eastern Europe after World War II and Germany's guest worker program of the 1950s to 1970s. Germany also took in a large number of refugees from the Yugoslav wars in the 1990s and most recently experienced the country's sharpest spike of immigration by asylum seekers from Syria, Afghanistan, and Iraq fleeing from war scenarios in their home countries. Based on data provided by Eurostat (2021), the ethnical groups currently most prevalent among German immigrants are from Turkey (12.7%) followed by Poland (7.4%), Syria (7.3%), Romania (6.8%), and Italy (5.7%).

---

<sup>5</sup> This chapter is based on Bloder, T., Eikerling, M., Rinker, T., & Lorusso, M. L. (2021). Speech and Language Therapy Service for Multilingual Children: Attitudes and Approaches across Four European Countries. *Sustainability*, 13(21), 12143. <https://doi.org/10.3390/su132112143>

Note that in this chapter, I only review the situation in Germany and present data based on a more extensive version of the questionnaire. The data collected in Austria, Italy, and Switzerland, reported in Bloder et al. (2021) are not included in this chapter.

<sup>6</sup> Note that the numbers reported do not contain all people of foreign descent but with citizenship of one of the respective countries. Actual numbers of people with migration background are thus expected to be higher.

### 3.1. Speech and Language Pathology in Bilingual Contexts

Almost half of the children who receive SLP services in Germany grow up in bilingual settings (with a tendency for a steady increase in this number; cf. 42.7% in Lücke & Ritterfeldt, 2011; compared to 49% in Scharff Rethfeldt, 2017). Even though bilingual language acquisition differs from monolingual language acquisition (Zmarich, Lena, & Pinton, 2014), bilingualism per se does not put successful language development at risk (Armon-Lotem, de Jong, & Meir, 2015; Paradis, Crago, Genesee, & Rice, 2003). In fact, the same prevalence of DLD (approximately 7%; Tomblin et al., 1997) across all children, be they monolingual or bilingual, can be expected.

Nevertheless, the data of Lehti, Gyllenberg, Suominen, and Sourander (2018) show that poor language development and as a result poor academic performance are more frequently found among children of foreign-born parents. Similar observations were made by De Lamo White and Jin (2011) as well as Saenz and Huer (2003) but the authors argued that the differences found between children born in monolingual families and children born in families with a migration background should be traced back to inadequacies and biases in the assessment of linguistic abilities of bilingual children rather than being representative of their actual language abilities. Another possible explanation for the higher proportion of bilingual children showing SLP needs could be linked to socio-economic-status (SES), presuming that lower SES is more prevalent among families with migration and, thus, more likely a bilingual background. Indeed, SES exerts an influence on (bilingual) language development through the degree of richness and diversity in the speech caregivers address to their children (Hart & Risley, 1995; Hart & Risley, 2003; Hurtado, Marchman, & Fernald, 2008; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010), and the quality and amount of complex language input in the form of shared book-reading, story-telling, or varied social discourse relevant for the development of narrative competences (Grosjean, 1982). The effects of such input-related factors are not to be confused with the effects of bilingualism (the two factors act upon language development independently from each other, as shown by Calvo and Bialystok, 2014). However, a combination of the two can make it even more difficult to adequately identify language impairment and increase the risk of misdiagnoses, which may at least partially account for the perceived increased prevalence of DLD in bilingual children.

Despite the high proportion of bilingual children receiving SLP services, common SLP practices predominantly involve monocultural and monolingual assessment and intervention routines (Scharff Rethfeldt, 2016), which can lead to either over- or underdiagnosis of DLD in bilingual children, by either overlooking clinically relevant deviance from typical language

development or (which appears to be more common) providing SLP services to children with as yet insufficient skills in the societal language due to insufficient language experience, which may lead to unnecessary stigmatization (von Suchodoletz & Macharey, 2006) as well as inappropriate allocation of health system resources.

Circling back to the topic of sustainability, these shortcomings in the provision of SLP services for bilingual children that in particular are in conflict with SDG “10: Reduced inequalities” (United Nations, 2015) have already been approached in various attempts through so-called policy papers, in which best-practice recommendations and guidelines for the assessment and treatment of DLD in bilingual children are summarized. For instance, the MULTI-SLI position statement (Blumenthal, Scharff Rethfeldt, Salameh, Muller, Vandewalle, & Grech, 2015) demands equal access to SLP services for monolingual and bilingual children. It is also widely known that DLD always affects all languages spoken by a child (Bishop et al., 2017; Garraffa, Vender, Sorace, & Guasti, 2019; Blumenthal et al., 2015; Scharff Rethfeldt et al., 2020). Therefore, children’s oral skills need to be assessed in all of their languages in order to reliably distinguish DLD from typical variations in bilingual language acquisition (De Lamo White & Jin, 2011; Garraffa et al., 2019). However, there is a disconnect between this theoretical knowledge and applied clinical practice in SLP services when it comes to working with bilingual children. This is evidenced by the fact that there are only very few SLPs who provide bilingual assessment and intervention (Jordaan, 2008) and that children who live in homes where a heritage language is spoken are most commonly assessed and treated in the societal language only (Friedrich & Knebel, 2017; Williams & McLeod, 2012), even though the majority of SLPs declares not to agree with this practice (Marinova-Todd et al., 2016). In general, many SLPs and parents feel especially under pressure to assess and treat children’s competences in the societal language as they are considered most important in fostering children’s school-readiness (Jordaan, 2008). However, research has shown that children’s improvements in the language targeted by the intervention (i.e., the societal language) often do not transfer to their heritage language (Restrepo, Morgan, & Thompson, 2013; Thordardottir, Cloutier, Ménard, Pelland-Blais, & Rvachew, 2015). Furthermore, bilingual children with DLD have been found to be at greater risk of heritage language attrition, stagnation, and incomplete acquisition than their bilingual TD peers (Restrepo, 2003). Taken together these findings suggest the great importance of specifically targeting bilingual children’s heritage language in Speech and Language Pathology intervention in order for children to be able to successfully communicate both within and outside the family context. Especially, since previous research has shown that bilingual language intervention does not reduce the rate of children’s acquisition



of the societal language (Crowe, Cuervo, Guiberson, & Washington, 2021; Harvey, Allaway, & Jones, 2018).

Several studies have focused on identifying the factors that hinder SLPs from applying what may be labeled the “best practice protocol” when working with bilingual children. Based on these studies I identified three main types of barriers to effective assessment and treatment of bilingual children with DLD.

### 3.1.1. SLPs’ Lack of Language Proficiency beyond the Societal Language

Several efforts have been made to address the issue of providing assessment and intervention in the societal language while at the same time accounting for children’s L2-learner status. The BiSLI COST Action IS0804, for example, has developed so-called “language-universal” assessment tasks that are supposed to be independent of the language(s) a child speaks and Boerma and Blom (2017) showed that performance on so-called language-universal tasks was indeed correlated with children’s language proficiency. Another diagnostic approach is to rely exclusively on the societal language, but to refer to norms that were standardized specifically on the bilingual population (e.g., the LiSeDaZ; Schulz & Tracy, 2011). Nevertheless, Scharff Rethfeldt and colleagues (2020) as well as Garraffa and colleagues (2019) still highlighted the need for better knowledge about the language-specific sequences in the acquisition of linguistic elements (e.g., phonemes) and structures (e.g., morphosyntax) and language-specific clinical markers for DLD in order to provide adequate assessment and intervention. For this reason, Chilla (2014) recommended that the assessment of children’s heritage language skills should be conducted by trained, native-language speaking professionals. However, this requirement poses challenges for clinical and educational staff due to a lack of linguistic knowledge of the various heritage languages potentially represented in all of their bilingual patients (Armon-Lotem & de Jong, 2015; Garraffa et al., 2019; Grimm & Schulz, 2014; Scharff Rethfeldt et al., 2020). In fact, Roseberry-McKibbin, Brice, and O’Hanlon (2005) found that the problem that occurred most frequently in service delivery to bilinguals was therapists’ lack of knowledge of the children’s heritage languages. Only a small percentage of SLPs provide therapy in more than one language and even fewer are L1 speakers of two or more languages (7% in Friedrich & Knebel, 2017 or 8.3% in Lüke & Ritterfeld, 2011 respectively). Additionally, the overlap between the languages spoken by SLPs and the children on their caseloads is limited. Although one-third of the respondents in Wintruff, Orlando, and Gumpert (2011) indicated that they had native-like language skills in another language in addition to German that could potentially be used diagnostically, in the majority of cases these were English (25%) and French (5%),

languages that are represented in only 5% to 7% of their bilingual patients. Beyond the restrictions of lack of language proficiency in a language other than the societal language affecting assessment and treatment of DLD in bilinguals, many clinicians also feel limited in their abilities to provide information and instructions to children's caretakers (Grandpierre et al., 2018) and many believe that, even though it should not be compulsory, it would be beneficial to SLPs to be fluent in more than one language (Williams & McLeod, 2012). Clinicians who speak several languages indeed (and rather obviously) reported assessing and treating children in all their languages more often than their monolingual colleagues (D'Souza, Kay-Raining Bird, & Deacon, 2012).

### 3.1.2. Lack of Satisfactory Diagnostic Tools Available for the Bilingual Population

Many SLPs in Germany feel rather inadequately equipped with the diagnostic materials that are currently available, particularly for the bilingual population (Lüke & Ritterfeld, 2011), and many try to cope by instead using informal procedures when assessing the speech and language development of children from bilingual backgrounds (Williams & McLeod, 2012). Some, for instance, use self-constructed or translated versions of instruments in the societal language to elicit utterances in the children's heritage languages, and only in isolated cases do SLPs use assessment tools that have been explicitly normed for monolingual children in the respective heritage language (Wintruff et al., 2011). For the purpose of language assessment of their patients' heritage languages, many SLPs alternatively focus mainly on anamnestic data (Friedrich & Knebel, 2017) or questionnaires, specifically designed for bilingual children's parents (Bonifacci et al., 2020; Paradis, Schneider, & Sorenson Duncan, 2013). However, collecting this information is made difficult when the therapist and the caregiver(s) do not speak the same language. Difficulties in collecting a patient's case history were reported to especially arise when a child's primary caregiver (typically the person who is most knowledgeable of the child's (language) development) was unable to speak the therapist's language, leaving the other parent to act as the family spokesperson and interpreter (Grandpierre et al., 2018). This communicative barrier is often maintained because the majority of SLPs do not have access to professional interpreters at all (D'Souza et al., 2012). This is why position statements and policy papers recommend the use of indirect measures to gain information on the linguistic background of the child (Blumenthal et al., 2016; Garraffa et al., 2019; Scharf Rethfeldt et al., 2020). Many SLPs, in fact, rely on parental questionnaires in the family's native language (if available) (Wintruff et al., 2011). Another clear recommendation is that language performance of bilingual children should not be compared to monolingual language learners (Blumenthal et

al., 2016). At the same time, the number of standardized tests providing norms for bilinguals is insufficient and – even more serious – due to heterogeneity of the bilingual population (Scharff Rethfeldt et al., 2020), it is difficult to create these norm data (Garraffa et al., 2019). When assessing children’s language performance in the societal language, SLPs thus usually rely on diagnostic materials designed and normed for monolingual children (Friedrich & Knebel, 2017). Unfortunately, SLPs normally (have to) rely on the assessment of only the societal language using tests that were standardized exclusively for monolingual children (Lüke & Ritterfeld, 2011). Still, since monolingual norms generally are unsuitable for bilingual children (Armon-Lotem & de Jong, 2015), these test results have to be interpreted with caution and should be limited to the assessment of the educational issues such as the child’s school-readiness or their need for additional support in school, while they are of little use to address clinical, diagnostic issues. Lüke and Ritterfeld (2011) highlight that test scores obtained from materials designed for monolinguals should only be considered as a reference point in the course of the diagnostic process and have to be related to different factors of bilingualism. Otherwise, SLPs risk misdiagnoses and erroneous decisions concerning a child’s need for therapy (Grimm & Schulz, 2014). On average, bilingual children receive speech therapy over a longer period of time than their monolingual peers and their rate of successfully completing SLP services is lower (Triarchi-Herrmann, 2009), which, in part, may be explained by SLPs’ difficulties to quantifiably track their bilingual patients’ intervention progress and outcome due to the paucity of available standardized and normed assessment materials (Jordaan, 2008).

As already pointed out earlier, SLPs are not only involved in the diagnosis and treatment of DLD, but also in advising caregivers (also referred to as caregiver counseling or consultation) and are thus responsible for providing recommendations about how parents can best support their children’s language development. Garraffa and colleagues (2019) highlight that language use within the family might change when being immersed into a different linguistic context but emphasize that heritage languages should not be abandoned. They also clearly state that bilingualism does not negatively impact language acquisition, even if a child has language impairment (as mentioned earlier in section 1.1). The majority of SLPs agree with this practice, as shown by Williams and McLeod (2012) in their investigation of parent counseling. An encouraging finding was that the majority of their respondents indicated that they advised parents of bilingual children to speak the languages they know best to their children (61.6%). About a quarter said they usually give the advice to use the family’s L1 only (25.6%), which, according to Jordaan (2008) could be interpreted as a strategy to ensure the development of both languages (as exposure to the societal language is ensured in any context outside of a

family's home). Beyond that, only very few SLPs (7.0%) recommended the use of the societal language only or gave other advice (5.8%) (Williams & McLeod, 2012).

### 3.1.3. SLPs' Preparedness and Confidence when Assessing and Treating Bilingual Children

The gap between clinical practice and theoretical knowledge is also evident in SLPs' perception of their own preparedness and, thus, confidence to diagnose and treat bilingual children with DLD. Unfortunately, many (monolingual) SLPs believe that they do not have the necessary expertise to work with culturally and linguistically diverse families (Stankova et al., 2020). In a German study by Wintruff and colleagues (2011), only 3% of SLPs felt that their training had adequately prepared them to work with bilingual children. Of the respondents, 80% said that the topic of bilingualism was covered only "very little" or "not at all" throughout the course of their training or studies (Wintruff et al., 2011). In their 2005 study, Roseberry-McKibbin and colleagues investigated how frequently different problems in SLP service delivery were encountered by clinicians in the USA. In terms of SLPs' knowledge about the phenomenon of bilingualism, they found that the lack of knowledge of developmental norms in children's primary language(s) and difficulties in distinguishing a language difference from an actual disorder were encountered most frequently. These findings suggest that SLPs have a need for additional support and information about language-specific linguistic phenomena and their developmental trajectory in different languages as well as specific procedures for bilingual children in order to feel more confident in working with them. In a more recent study in Germany, 43.9% of the participating SLPs also indicated to feel insecure when assessing bilingual children (Friedrich & Knebel, 2017). Half of the participants in this study stated that they had attended workshops on bilingualism, a finding that substantiates the connection that has been made between SLPs' engagement in the topic of bilingualism and their perception of their own competence when working with bilinguals (Lüke & Ritterfeld, 2011). Comparing the results from a 1990 survey with their more recent data that were collected 11 years later, Roseberry-McKibbin and colleagues (2005) found that in 1990, 23.6% of respondents had engaged in university courses addressing bilingualism, whereas in 2001, 73% of respondents indicated that they had been taking similar classes. This may indicate an increase in the attention the topic of bilingualism has received in recent years but also the challenges that SLPs face when trying to adequately assess and treat bilingual children. Another factor that has been found to impact on SLPs' confidence in their cultural and language competence is a bilingual socialization (influenced by their country of origin's history of immigration, internal minority groups, and patterns of language use). Taken together, these findings suggest that SLPs are very

much aware of the barriers and try to overcome them by acquiring more knowledge and expertise.

### 3.2. The Present Study

This study aimed to explore German SLPs' beliefs and approaches towards treating bilingual children and to examine which factors are most influential to adopting bilingually-oriented and -sensitive assessment and treatment approaches for this population by means of an online questionnaire.

### 3.3 Research Questions/Hypotheses

More specifically but very globally, I asked the question of how SLPs' attitudes towards bilingualism and DLD and their approaches in assessing the language performance of bilingual children were affected by:

- their overall professional experience (in years),
- their specific experience in working with bilingual children (assessed by how frequently they worked with bilingual children and the proportion of bilingual children on their caseloads),
- their personal language background (being monolingual vs. bilingual).

Based on the sources reviewed above, or - in the absence of previous literature - on what were logical consequences or the optimal way to face the challenges described, I expected that:

1. There is an association between SLPs' experience in working with bilingual children and their understanding of the relevance of bilingual approaches and the application of assessment tools specifically designed for the bilingual population.
2. Longer overall work experience (in years) is linked with more bilingually-oriented approaches in SLP service provision.
3. SLPs' personal language background (being monolingual vs. bilingual) is linked with the value assigned to bilingual approaches in SLP service for bilingual children.
4. The knowledge and the frequency of using assessment tools specifically designed for bilingual children is associated with the number of different diagnostic materials SLPs use in their everyday practice with bilingual children.
5. SLPs who advocate the notion that DLD is not exacerbated by bilingual language acquisition are more open to promoting and encouraging a linguistically diverse language environment.

6. SLPs who are supportive of the use of the heritage language at home are more likely to adopt bilingual approaches and are more open to the use bilingual materials.

### 3.4. Methodology

#### 3.4.1. Participants

The target group of this study was defined as certified Speech and Language Pathologists and other occupational groups working in the field of Speech and Language Pathology (e.g., speech therapists, clinical linguists, or state-certified breath, speech, and voice teachers), who diagnose and conduct therapy with children with developmental language disorder in Germany.

#### 3.4.2. Questionnaire Design

The data were collected by means of an online questionnaire, that was originally created and piloted in Italy and consisted of 24 questions. This Italian questionnaire was translated and adapted to German. Furthermore, a set of questions was added to address participants diagnostic approaches in more detail with an in-depth focus on the relevance of assessing children's language exposure. This led to a total number of 37 questions in the German version. Overall, the questions addressed participants' general beliefs towards bilingualism in the context of SLP, their professional and personal experience with bilingualism, and their common clinical practice when assessing and treating bilingual children.

The questionnaire comprised 21 multiple-choice questions, eight open-ended questions, and eight closed (yes–no) questions. In terms of its content, the questionnaire more specifically covered four thematic areas: (1) participants' experience with childhood bilingualism, (2) participants' attitudes and beliefs about DLD in bilingual children, (3) participants' personal clinical practice and their individual practical approaches to procedures for bilingual children with DLD, (4) barriers to effective assessment or treatment of bilingual children with DLD and the consequences of misdiagnoses. A full list of the questions (Q1–Q37) and response options in their English translation are reported in Appendix B (Table B1).

#### 3.4.3. Procedure

Between October 2020 and January 2021, the questionnaire was advertised online through (1) the two German national professional associations (the Bundesverband für Logopädie e.V., dbf; and the Deutscher Bundesverband für akademische Sprachtherapie und Logopädie, dbs), (2) other mailing lists (e.g., SES interdisziplinär, alumni list Klinische Linguistik Bielefeld University), (3) various social media channels, (4) and word of mouth. In sum, 99 German SLPs

took part in the study and a total number of 65 participants were included in the analysis after the exclusion of (1) incomplete responses,  $n = 22$  and (2) responses from participants with no prior professional experience (i.e., students),  $n = 12$ .

The data were collected in a completely anonymous form. Before initiating the actual survey, all respondents confirmed that they were participating in the study voluntarily and that they had been informed that they could abort their participation at any time. Finally, they agreed to the storage and processing of their anonymous responses as well as to the use of their data for scientific purposes and potential publications.

#### 3.4.4. Data Analysis

Data were analyzed by means of IBM SPSS Statistics v.26. The prevalence of certain responses was assessed through chi-square goodness-of-fit analyses, whereas associations between the variables were assessed by chi square tests of independence. Since some of the variables provided answers that could be led back to an ordinal scale (i.e., all the questions for which possible responses were “never, rarely, sometimes, yes” or similarly, other response options with three or four different ordinal levels), the reciprocal effects of such variables were further characterized through ordinal-x-ordinal logistic regression statistics such as Somer’s D. First, the results were analyzed in the whole sample in order to describe the general attitudes and practices of SLPs in the group. Then, specific analyses were conducted to check the effects of the variables that were hypothesized to act as independent variables on other responses, following the hypotheses postulated above.

Participants’ comments and remarks to open-ended questions were evaluated qualitatively. In a first step, all responses were screened and, based on participants’ individual responses, general response-types were derived based on which thematic categories were created. Subsequently, all responses were assigned to at least one of these categories, whereby multiple classifications were permissible. For the final quantitative presentation of the data, a simple frequency count was performed. Furthermore, some of the participants’ most common responses are presented as examples to illustrate, justify and support the quantitative results.

### 3.5 Results

#### 3.5.1 SLPs’ Characteristics

In general, different levels of overall professional experience in the field of SLP (assessed in years) were equally represented among the study participants ( $Q1; \chi^2(3, N = 65) = 3.99, p > .05$ ); 18.5% had less than 5 years of overall professional experience in providing SLP services,

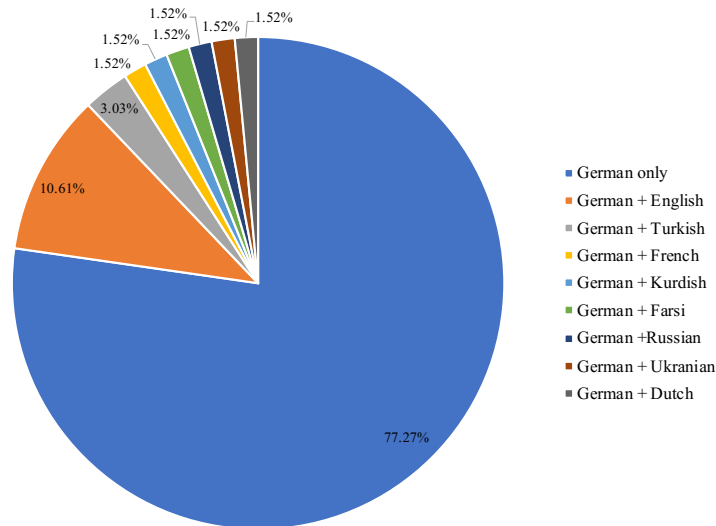
27.7% had between 5 and 10 years, 20.0% between 10 and 20 years, 33.8% had more than 20 years of overall professional experience in providing SLP services to children. More specifically, concerning respondents' particular experience in working with bilingual children (Q2), only 3.1% declared that they had no experience at all in working with bilingual children ( $X^2(2, N = 65) = 112.65, p < .001$ ), while 1.5% declared that they had at least some experience, and 95.4% regularly provided SLP services for bilingual children with DLD. A more in-depth measure of SLPs' experience particularly in working with bilingual children (assessed through the proportion of bilingual children with DLD represented on their total caseloads; Q3) is displayed in Table 7.

**Table 7.** Proportion of bilingual children with DLD on participants caseloads.

	Percentage of bilingual children assessed/treated for DLD					
	0–5%	6–25%	26–50%	51–75%	76–95%	96–100%
proportion of respondents	1.5%	27.1%	24.6%	23.1%	18.5%	4.6%

Furthermore, of all respondents, only the minority of 21.5% indicated that they could speak a language other than German at a native-like level (Q4;  $X^2(1, N = 65) = 21.06, p < .001$ ). Of these, in addition to German, the majority spoke languages that are typically taught in German education contexts (i.e., English or French) or languages of Germany's neighbor countries and, thus, border regions (e.g., Dutch), while overall less than 10% spoke at least one language that could be considered more prevalent among the German migrant community (e.g., Turkish, Farsi, Polish) (Q5). In accordance with their personal language knowledge and abilities beyond German, 18.5% offer SLP services in multiple languages (Q6;  $X^2(1, N = 65) = 25.86, p < .001$ ) among which are English, French, Russian, Turkish, Croatian, Czech (Q7; see Figure 13). While the majority of SLPs (72.3%) declared to feel competent when working with bilingual children, almost a third of the respondents (27.7%) did not (Q28;  $X^2(1, N = 65) = 12.94, p < .001$ ) and almost all of the respondents (93.8%) are seeking more information and knowledge concerning the assessment and treatment of bilingual children (Q29;  $X^2(1, N = 65) = 49.99, p < .001$ ).





**Figure 13.** Overview of the languages used by the participants for SLP service purposes (Q7).

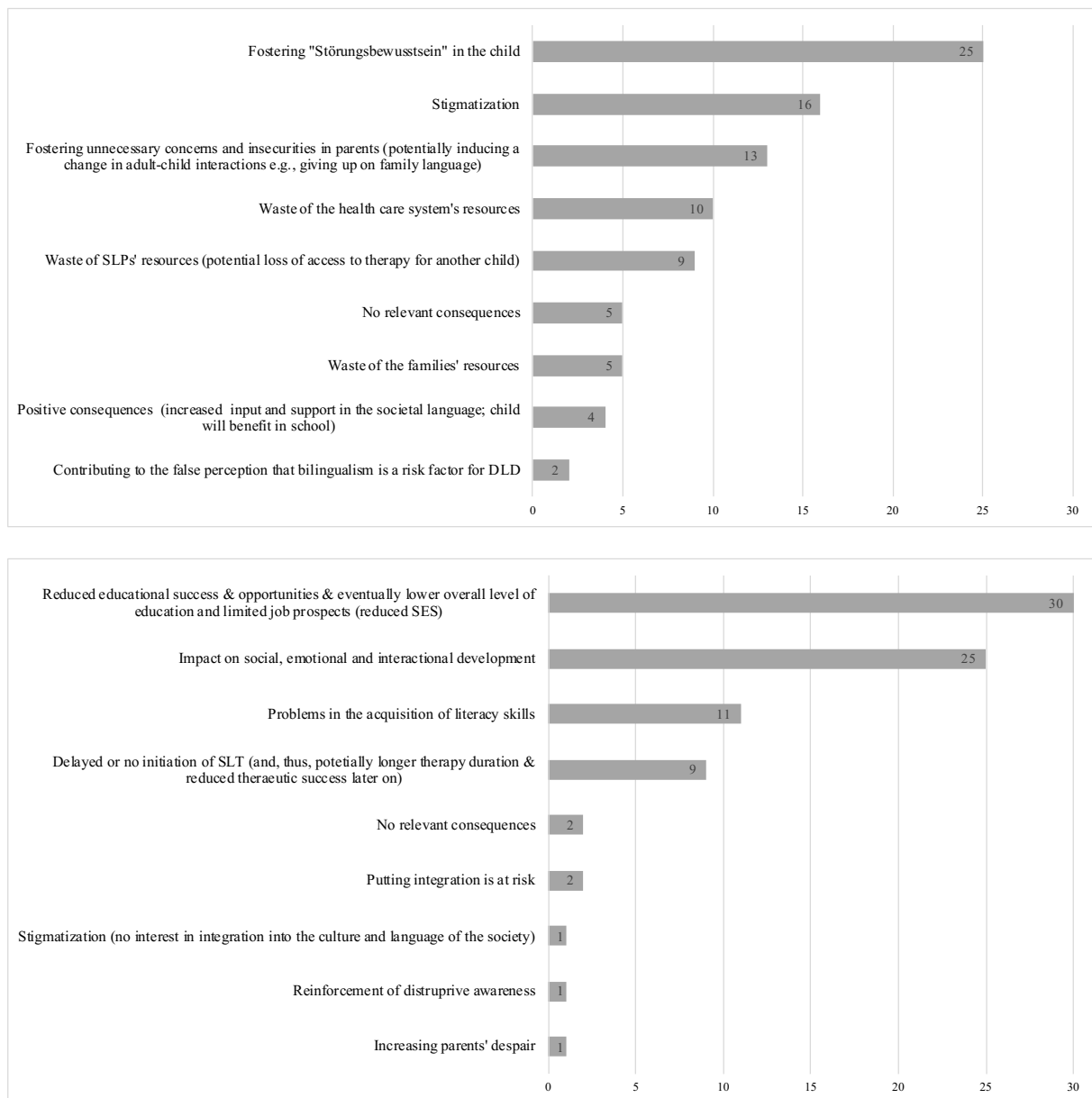
### 3.5.2. SLPs' Attitudes and Beliefs About the Assessment of Bilingual Children

Almost all participants (95.4%) believed that it is (highly) relevant to distinguish between DLD and language difficulties due to limited exposure, whereas the minority of respondents (4.6%) considered this distinction to be (completely) irrelevant (Q34;  $X^2(3, N = 65) = 58.02, p < .001$ ). In their comments (Q35), participants who were supportive of this distinction, reasoned with the adequate allocation of resources and the professional expertise different occupational groups have in order to either best support children with limited exposure or treat children with language impairment. They further mentioned that depending on the reason behind children's language difficulties, different approaches were necessary; while therapy should also take into account children's heritage language, language support should exclusively be concerned with the societal language. In contrast, in their comments, SLPs who did not consider a distinction between DLD and expressions of limited exposure to be of relevance, pointed out that (1) sometimes "pathologizing" children's language performance was necessary in order for them to be recognized by the (health care) system and that (2) the first priority for bilingual children should be to acquire adequate societal language skills in order to perform well in school and, thus, should be pursued by all occupational groups alike.

Concerning potential consequences of an overdiagnosis of DLD (i.e., identifying a bilingual child who is actually linguistically typically developing as being language impaired; Q36), among participants' most common responses were "Pathologisierung"<sup>7</sup>, generating

<sup>7</sup> The German term "Pathologisierung" means that a typical variation of language development is being classified as a pathological condition.

unjustified “Störungsbewusstsein”<sup>8</sup> in healthy, typically developing children, , and a waste of resources. Conversely, reduced educational achievements, eventually a lower overall level of education and limited job prospects, as well as a negative impact on children’s social, emotional and interactional development were most commonly mentioned as consequences of an underdiagnosis of DLD (i.e., not recognizing DLD in a bilingual child who actually needs therapy; Q37). See Figure 14 for an overview of participant’s responses to Q36 and Q37.



**Figure 14.** Systematic overview of participants’ responses to the questions concerning the consequences of an overdiagnosis (Q36; top Figure) and underdiagnosis (Q37; bottom Figure) of DLD in bilingual children. Responses were assigned to overarching thematic categories, X-axis displays the absolute count of responses allocated to each category.

<sup>8</sup> The German term “Störungsbewusstsein” indicates that a child becomes aware of their disorder and is often associated with negative consequences such as fear of speaking the (second) language, withdrawal from social and communicative interactions etc.

In general, the significant majority of SLPs was supportive of heritage language use in the context of DLD (Q11;  $X^2(1, N = 65) = 46.54, p < .001$ ). While 92.3% believed that children’s communication partners should always speak in the language, they know best, 7.7% thought that for bilingual children with DLD language input should be reduced to a single language (both inside and outside the family environment). Furthermore, respondents thought that compared to monolingual children, different approaches were always (69.2%) or at least sometimes (30.8%) needed in the assessment and treatment of bilinguals with DLD (Q8;  $X^2(1, N = 65) = 9.62, p = .002$ ). Furthermore, the majority of SLPs thought (52.3%) and tended to think (20.0%) that DLD is independent of bilingual language acquisition (Q10;  $X^2(3, N = 65) = 29.59, p < .001$ ). Additionally, the majority of all SLP respondents also indicated that intervention should take both, children’s heritage as well as the societal language, into account (Q9;  $X^2(3, N = 65) = 21.95, p < .001$ ) (see Table 8).

**Table 8.** Distribution of SLPs’ responses to Q9 and Q10, concerning general attitudes towards bilingualism.

	A	A more than B	B more than A	B
Q9 Do you think that (A) the therapy of DLD in bilingual children should be limited exclusively to the societal language, or that (B) the child’s heritage language should also be taken into account?	3.1%	21.5%	33.8%	41.5%
Q10 Do you believe that (A) DLD is independent of speaking a second language, or that (B) a disorder can be exacerbated by second language acquisition?	52.3%	20.0%	21.5%	6.2%

As shown in Table 9, the significant majority (67.7%;  $X^2(3, N = 65) = 74.39, p < .001$ ) of all respondents thought that it was always useful to compare bilingual children’s language performance between the societal and the heritage language (Q12). More specifically, most participants declared that it is useful to compare children’s performance in the various linguistic domains (Q21- phonology; Q22 – morphosyntax; Q23 and Q24 - vocabulary: see Table 5 for complete results) in their heritage language with their performance in the societal language (Q21-Q23;  $X^2(3, N = 65) = 87.80-117.95, all p_s < .001$ ).

**Table 9.** Distribution of SLPs’ responses concerning the usefulness of comparison of different aspects of children’s heritage and societal language performance.

Do you think that it would be useful/helpful to...	Never	Rarely	Sometimes	Yes
--	-------	--------	-----------	-----

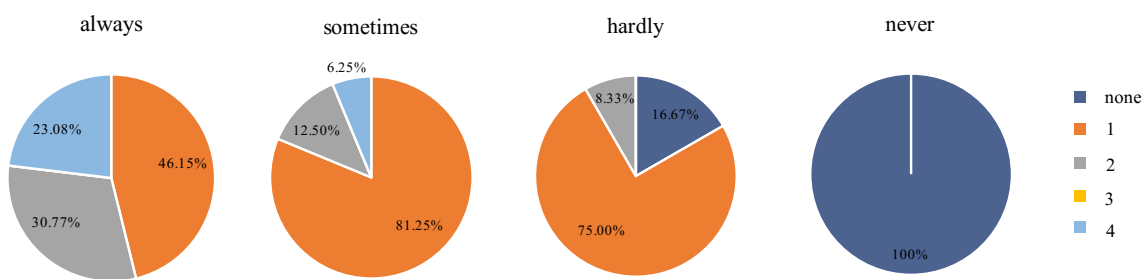
Q12 ...compare a child's language performance in their heritage and societal language?	1.5%	3.5%	27.7%	67.7%
Q21 ...check whether the phonemes that cause difficulties to the child in the societal language are present/are also affected in the heritage language?	3.1%	3.1%	13.8%	80.0%
Q25 ...have a summary chart of the phoneme inventory of the child's heritage language?	0.0%	4.6%	20.0%	75.4%
Q22 ...check whether the syntactic and morphological structures that cause difficulties to the child in the societal language are present/are also affected in the heritage language?	4.6%	1.5%	20.0%	73.8%
Q26 ...have a summary table of the main syntactic structures and constructions in the child's heritage language?	0.0%	1.5%	13.8%	84.6%
Q23 ...check whether words that the child uses semantically/lexically incorrectly in the societal language are similar or very different from the heritage language?	4.6%	4.6%	7.7%	83.1%
Q24 ... check whether words that the child uses semantically/lexically incorrectly in the societal language are also used incorrectly in the heritage language?	3.1%	3.1%	10.8%	83.1
Q27 ...have an overview table with a list of the most important prepositions (translations & usage) in the child's heritage language?	3.1%	10.8%	15.4%	70.8%

### 3.5.3. SLPs' Clinical Practices in the Language Assessment of Bilingual Children

Most respondents indicated to draw (26.2%) or a tendency to draw (41.5%) their comparisons of children's language performance between their two languages based on information provided by the parents, whereas only a small proportion of clinicians were usually (16.9%) or at least sometimes (13.8%) able to directly observe and compare children's language behavior in both languages (Q13;  $X^2(3, N = 64) = 12.25, p = .007$ ).

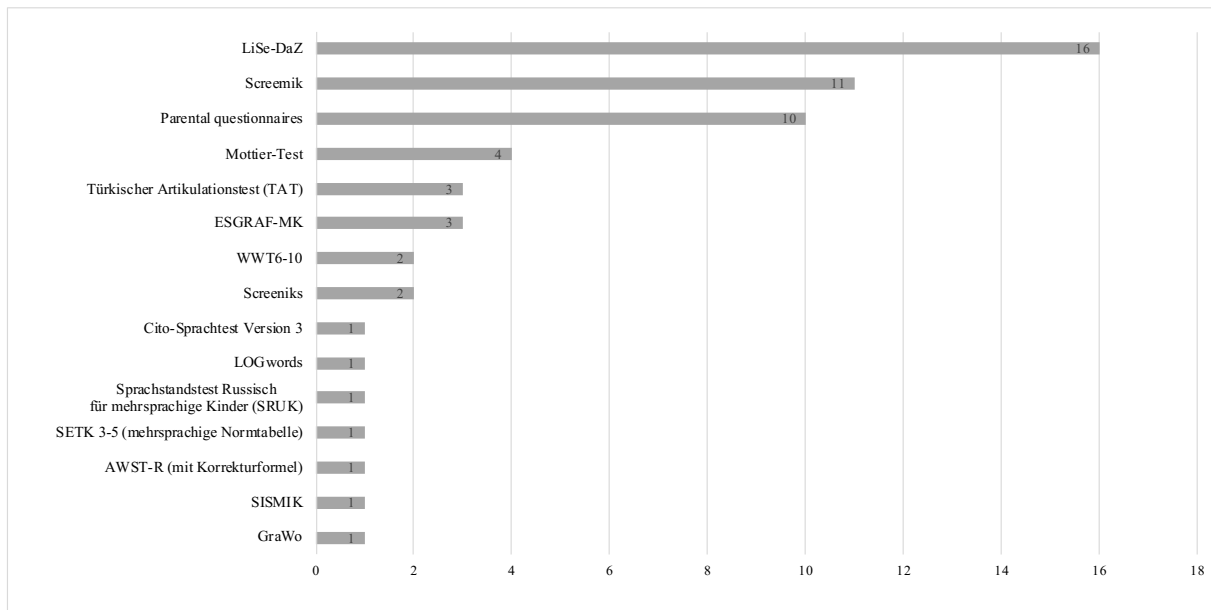
When assessing bilingual children's language performance in the societal language German, only 9.2% considered it sufficient to refer to monolingual norm data (Q16) and the vast majority (90.8%) of all respondents considered monolingual norm data to be (completely) insufficient for the bilingual population ( $X^2(2, N = 65) = 18.86, p < .001$ ). However, in their comments, some participants pointed out that monolingual German norm data may provide a very general reference point for typical (German) language development that if interpreted with caution could be useful for the qualitative evaluation of bilingual children's language skills. Along the same lines, 73.8% considered the use of test or screening procedures in children's heritage languages to be (very) informative, while 26.2% thought that such materials had little to no informative value (Q14;  $X^2(3, N = 65) = 34.02, p < .001$ ). However, only 4.6% of the SLPs judged the use of tests or screening procedures in the heritage language to be very feasible/applicable (Q15;  $X^2(3, N = 65) = 62.94, p < .001$ ) but almost the entire sample (93.3%) assigned (high) relevance to the provision of bilingual norm data for existing test and screening materials (Q17;  $X^2(2, N = 65) = 19.23, p < .001$ ). While the significant majority (84.6%) was

aware of diagnostic material specifically designed for bilingual children, 15.4% of the respondents were not (Q18;  $\chi^2(1, N = 65) = 31.15, p < .001$ ). However, only a few of the respondents (21.5%) indicated to regularly use such materials when assessing bilingual children's language performance (Q19;  $\chi^2(3, N = 65) = 2.75, p > .05$ ). While 30.8% declared to use such materials sometimes and 18.5% rarely, about a third of all respondents (29.2%) declared not to use them at all. Further, those SLPs who indicated to always use special diagnostic tools with bilingual children applied a greater variety of different materials than their colleagues who indicated to only sometimes or rarely use such materials ( $\chi^2(9, N = 65) = 51.269, p < .001, \text{Somers's } D = .599$ ; see Figure 15).



**Figure 15.** Overview of the variety of different diagnostic materials/tools specifically designed for bilingual children applied by the participants according to their response to Q19 (Do you use special diagnostic materials/tools for bilingual children with DLD?)

Figure 16 provides an overview of the frequency counts for each bilingual test or screening tool that was mentioned by the participants (Q20). While participants' responses indicate that there is a considerable number of tools available to SLPs in their everyday clinical practice, the majority of them is actually used by only a small proportion of clinicians.



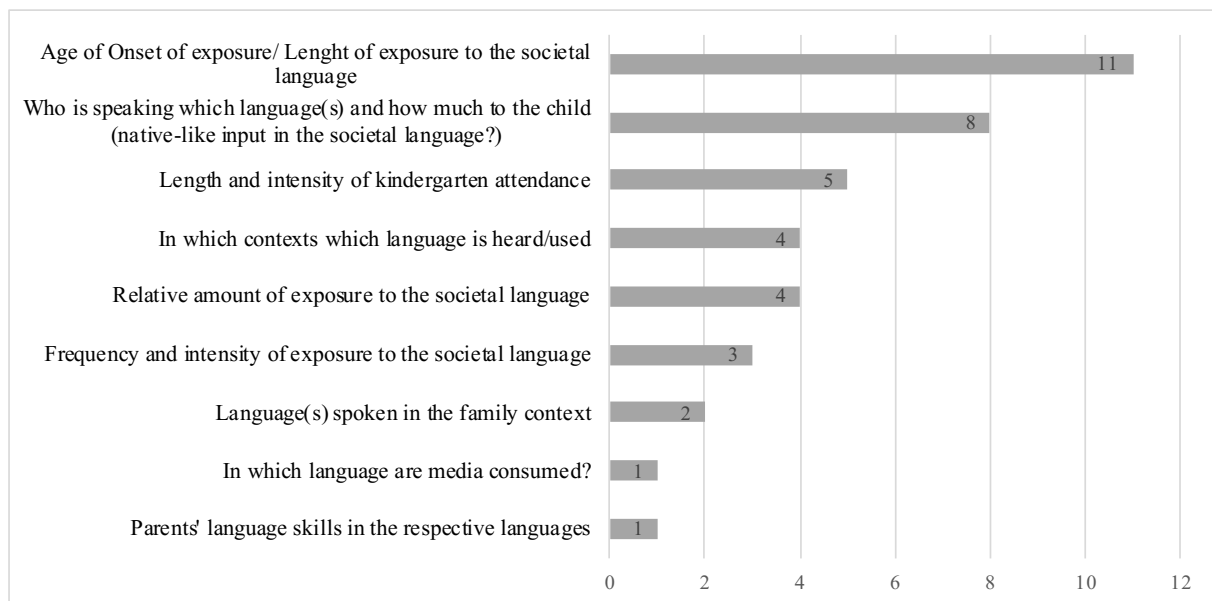
**Figure 16.** Frequency count of screening tools/test materials mentioned by the study participants in response to the open-ended question Q20 (What diagnostic material/tools do you use for bilingual children with DLD?)

Despite being aware of diagnostic tools specifically designed for bilingual children, 32.7% indicated to not apply any of them when assessing bilingual children’s language abilities ( $\chi^2(3, N = 65) = 8.188, p < .05, Cramer’s V = .355$ ). Only the minority of SLPs (7.7%) indicated to be familiar with the LITMUS Cost Action IS0804 test-battery (Q30;  $\chi^2(1, N = 65) = 46.54, p < .001$ ).

All SLP respondents further declared they considered it useful to consult materials for different heritage languages to support their work with bilingual children, as shown in Table 5 (Q25 - summaries of phonemic inventories; Q26 - overview of main syntactic structures and constructions; Q27 - overview table with a list of the most important prepositions;  $\chi^2(2,3, N = 65) = 54.03-78.40$ , all  $p_s < .001$ ).

Finally, while 76.9% incorporated information about children’s individual language exposure in the diagnostic process, 23.1% of all respondents did not (Q31;  $\chi^2(1, N = 65) = 18.85, p < .001$ ). In their comments, participants who did assess children’s language background for diagnostic purposes, indicated to most frequently refer to children’s AoO and LoE to the societal language followed by information about who spoke which language with the child (e.g., to evaluate how much native-like input the child received in the societal language), and the relative amount of exposure to the societal language (Q32; see Figure 17 for an overview of participants’ response types). Concerning the question of how this information did influence their diagnostic decision (Q33), participant’s responses were less systematic. Most commonly SLPs’ declared that they would interpret children’s German test scores in

relation to their LoE to the societal language in order to be able to determine what to expect in typical language development.



**Figure 17.** Language exposure related information that SLPs incorporate in their diagnostic process. Participants' responses to the open-ended question Q32 were assigned to overarching thematic categories, X-axis displays the absolute count of responses allocated to each category.

### 3.5.4 Effects of SLPs' prior experience

SLPs' prior professional experience was measured in three different ways: (1) through their overall work experience in years (Q1), (2) how frequently they treated bilingual children (Q2; yes/regularly, sometimes, never), and (3) through the specific proportion of bilingual children on their caseloads (Q3). Participants' overall work experience in years was significantly associated with their perceived level of competence when working with bilingual patients (Q28;  $\chi^2(3, N = 65) = 7.705, p = .053, Cramer's V = .344$ ). Of the participants with practical clinical experience below five years, only 41.7% declared to feel confident when working with bilingual children, whereas of the SLPs with clinical work experience of more than five years, 72.7-84.6% felt confident when working with bilingual patients. Furthermore, overall work experience in years was associated with participants' knowledge of the LITMUS Cost Action IS0804 test-battery (Q30;  $\chi^2(3, N = 65) = 8.275, p < .05, Cramer's V = .357$ ). While none of the SLPs with more than ten years of clinical work experience were familiar with this test-battery, 8.3% and 22.2% respectively of the SLPs with below five and five to 10 years of work experience were familiar with the LITMUS test-battery.

More specifically, concerning SLPs' specific experience in working with bilingual children, significant associations were found between Q2 ("Have you every provided therapy

to bilingual children with developmental language disorder“ – “yes/regularly, sometimes, never”) and SLPs’ evaluation of the applicability of heritage language assessment tools (Q15;  $X^2(6, N = 65) = 13.185, p < .05, \text{Somers's } D = .436$ ). While none of the SLPs who sometimes/never provided therapy for bilingual children thought that the use of test/screening tools in children’s heritage language was feasible, 27.4% of the SLPs who regularly assessed and treated bilingual patients considered the use of heritage language assessment tools feasible. Further, there was another significant association with SLPs’ perceived level of competence in providing services for bilingual children (Q28;  $X^2(2, N = 65) = 8.212, p < .05, \text{Cramer's } V = .355$ ). None of the participants who sometimes or never, but 75.8% of the participants who regularly provided therapy for bilingual children reported to feel competent when working with this population.

Moreover, SLPs with fewer bilingual children on their caseloads more often considered specific norm data irrelevant for the assessment of bilinguals, whereas SLPs with a high percentage of bilingual patients on their caseloads ( $> 50\%$ ) more often assigned high relevance to the availability of specific norm data for this population (93.4%-100% of them found it to be (highly) relevant; Q17).

### 3.5.5 Effects of SLPs’ Personal Language Background

SLPs’ ability to speak more than one language at native level (Q4) was found to be associated with their approach to comparing children’s language performance between their heritage and the societal language (Q13;  $X^2(3, N = 64) = 13.744, p < .05, \text{Somers's } D = .463$ ). Therapists who were able to speak more than one language themselves, more often declared to directly observe children’s language behavior in both of their languages. Further, a marginal significant association was found with their use of diagnostic materials specifically designed for bilingual children (Q19;  $X^2(3, N = 65) = 7.573, p = .056, \text{Cramer's } V = .341$ ). Bilingual SLPs more often tended to use specific diagnostic materials with their bilingual patients and they also tended to be more confident in providing services for bilingual children compared to their monolingual colleagues (Q28;  $X^2(1, N = 65) = 3.763, p = .052, \text{Phi} = .241$ ). Finally, bilingual SLPs assigned higher relevance to distinguishing between actual language impairment (i.e., DLD) and language difficulties due to limited exposure to the societal language than the monolingual participants (Q34;  $X^2(3, N = 65) = 8.382, p < .05, \text{Cramer's } V = .359$ ). Table 10 provides an overview of all significant associations between SLPs’ language background and their approaches towards the assessment of bilingual children’s language performance.



**Table 10.** SLPs' Approaches towards Assessing Bilingual Children's Language Performance in Relation to their Personal Language Background

		<b>(Q4) Besides German, do you speak any other language(s) at native level?</b>	
		<b>Yes</b>	<b>No</b>
Q13 (A) were your comparisons based on information provided by the parents or (B) were you able to directly observe children's behavior in both languages (possibly in the presence of parents)?	A	0.0%	33.3%
	A more than B	30.8%	45.1%
	B more than A	23.1%	11.8%
	B	46.2%	9.8%
	never	21.4%	31.4%
Q19 Do you use special diagnostic material/tools for bilingual children with DLD?	rarely	0.0%	23.5%
	sometimes	35.7%	29.4%
	yes	42.9%	15.7%
Q28 Do you feel competent when assessing/treating bilingual children for DLD?	no	7.1%	33.3%
	yes	29.9%	66.7%
Q34 How relevant do you think it is to distinguish between bilingual children with language difficulties in German due to insufficient exposure to the societal language and bilingual children with DLD?	completely irrelevant	0.0%	2.0%
	irrelevant	7.1%	2.0%
	relevant	7.1%	47.1%
	highly relevant	85.7%	49.0%

### 3.5.6 Effects of SLPs' Attitudes Towards Children's Heritage Language

I found several significant associations between SLPs' diagnostic practice and their attitudes towards the heritage language use at home (Q11; "Do you think that (A) it would be better for bilingual children with DLD to speak only one language (both at home and outside of the family environment), or that (B) it would be better for all of a child's communication partners to interact with the child in the language they know best?"). These effects are reported in Table 11. A significant association was found between Q11 and Q10, i.e., SLPs' belief that DLD manifests independently of children's language background (i.e., monolingual vs. bilingual language status) or rather that bilingual language acquisition may exacerbate DLD ( $\chi^2(3, N = 65) = 13.105, p < .05, Cramer's V = .449$ ). This effect is probably best interpreted in the reversed direction: The therapists who clearly believed that DLD is independent of bilingualism significantly more often recommended that each family member should speak the language they know best with the child (55.0%), as opposed to the SLPs who believed that bilingualism has an effect on the manifestation of DLD in children, of which only 20.0% recommended the use of the heritage language at home.

Q11 also produced a significant effect on the value SLPs assigned to the comparison of bilingual children's linguistic performance between their two languages (Q12;  $\chi^2(3, N = 65) =$

13.077,  $p < .05$ , *Cramer's V* = .449). SLPs who were supportive of the use of the heritage language at home also thought that it is always useful to compare children's performance between their two languages (70.0%). By contrast, of the SLPs who were in favor of limiting children's language input to the societal language only 20.0% tended to think that it is always useful to compare children's performance between the societal and the heritage language.

Furthermore, Q11 produced a significant effect on Q31 ( $\chi^2(1, N = 65) = 9.887, p < .05, \Phi = .390$ ): SLPs who favored the use of the heritage language significantly more often incorporated information about bilingual children's language exposure in the diagnostic process (81.7%) than SLPs who were not supportive of the use of the heritage language at home (20.0%).

Finally, a significant association between SLPs' attitude towards heritage language use at home and the relevance they assigned to the differentiation between actual language impairment (i.e., DLD) and language difficulties in the societal language due to limited exposure was found (Q34;  $\chi^2(3, N = 65) = 12.443, p < .05, \text{Cramer's } V = .438$ ).

**Table 11.** SLPs' attitudes towards heritage language use in relation to their approaches in SLP service provision

		<b>Q11 Do you think that...</b>	
		<b>...it would be better for bilingual children with DLD to speak only one language (both at home and outside the family environment).</b>	<b>...it would be better for all of a child's communication partners to interact in the language they know best.</b>
Q10 Do you think that (A) DLD is independent of the second language, or that (B) a disorder can be exacerbated by second language acquisition??	A	20.0%	55.0%
	A more than B	0.0%	21.7%
	B more than A	40.0%	20.0%
	B	40.0%	3.3%
Q12 In the context of SLP for bilingual children with DLD, do you think that it is useful to compare children's language performance in the heritage and societal language?	never	20.0%	0.0%
	rarely	0.0%	3.3%
	sometimes	40.0%	26.7%
	yes	40.0%	70.0%
Q31 Do you incorporate information about children's language exposure (i.e., how often a child hears and/or speaks each of their respective languages into the diagnostic process when you assess bilingual children's need for Speech and Language Pathology services?	no	80.0%	18.3%
	yes	20.0%	81.7%
Q34 How relevant do you think it is to distinguish between bilingual children with language difficulties in	completely irrelevant	20.0%	0.0%
	irrelevant	0.0%	3.3%

German due to insufficient exposure to the societal language and bilingual children with DLD?	relevant	40%	38.3%
	highly relevant	40%	58.3%

### 3.6 Discussion

The results of the survey partly confirmed the research hypotheses. The data showed that the large majority of SLPs does acknowledge the importance of taking into account aspects of children’s heritage language(s) and culture(s) when treating them. However, there is a chance that social acceptability may have influenced SLPs’ responses to a certain extent. Nevertheless, a rather surprising was the finding that, while most SLPs did state to know about assessment tools specifically designed for bilingual children, the majority of respondents did not apply them on a regular basis in their clinical practice. Hypothesis 1, namely that SLPs’ experience in working with bilingual children would influence their awareness of and openness towards bilingual approaches was confirmed by the findings. Furthermore, SLPs overall work experience (in years) had an effect on their approaches towards the assessment of DLD in bilingual children (Hypothesis 2), so did the participants’ personal language background (being monolingual vs. bilingual), which confirmed Hypothesis 3. The hypothesis that SLPs’ knowledge of and even more important the frequency of use of different assessment tools specifically designed for bilingual children would be linked to the application of a wider variety thereof (Hypothesis 4) was confirmed by the data. However, despite being aware of at least some assessment tools, many SLPs do not make use of these materials at all. Leaving room for the speculative assumption that existing tools are either not reasonably applicable or not easily accessible for SLPs in clinical practice. Finally, the hypotheses that (1) SLPs who advocate the notion that DLD is not exacerbated by bilingual language acquisition would be more open towards greater linguistic diversity in children’s everyday language environment (Hypothesis 5) and that in turn (2) a positive attitude towards the use of children’s heritage language at home would be associated with openness to bilingual approaches in SLP service provision (Hypothesis 6) were confirmed. The data further show that many SLPs take into account bilingual children’s case histories in the diagnostic process by asking questions about their language background. Most commonly they refer to AoO and LoE to the societal language and some also inquire about more qualitative aspects of language exposure. However, when interpreting this information, there is no consensus about how to relate children’s language performance to measures of language exposure and most SLPs use this information as a rough estimate about what they should be expecting in children’s societal language performance. Generally, SLPs consider it very important to differentiate between actual language impairment

(i.e., DLD) and weak language performance due to limited exposure to the societal language in bilingual children. Arguably, SLPs seem to be aware of the differential effects of bilingual language acquisition versus the effects of low socio-economic status (SES). This is in line with Roy and Chiat (2013) who proposed that a certain proportion of low SES (i.e., often migrant and thus bilingual) children who, based on their scores in standardized language tests, supposedly present with DLD have “intact language potential”. The authors argue that had these children only been exposed to a more advantageous (i.e., linguistically stimulating) environment, they would have developed language skills within the normal range. That being said, these children still require and should be entitled to adequate language support in order to counteract the effect of accumulated disadvantage. It remains an issue of debate whether in the case of bilingual children a strict discrimination between children in need of SLP services and children in need of language support is necessary, if at all possible.

### 3.6.1. Experience Related Effects

Overall work experience in years, which is most likely linked to the recency of participants’ training and thus also their age, was found to be influential for a subset of questions. SLPs who had accumulated a greater amount of working experience reported to feel more confident when working with bilingual children. However, concerning their familiarity with different assessment tools specifically designed for the bilingual population such as the LITMUS Cost Action IS0804 test-battery, SLPs with less work experience (and thus speculatively younger participants) were more commonly aware of such materials. Nevertheless, even among this group of participants, not more than 20% had heard of the LITMUS tasks. In particular, this highlights the urgent necessity of further development of easily accessible and especially applicable diagnostic tools and a closer collaboration between research and clinical practice. In particular, in-depth involvement of SLPs in this dynamic field of research would be desirable, since they do not only have the theoretical but moreover, especially the practical, application-related expertise that is required for the development of adequate testing tools. Nevertheless, the general measure of SLPs work experience (in years) turned out to have only little influence on participants’ attitudes and approaches towards SLP service provision for bilingual children. I found more specific measures of SLPs experience in treating bilingual children (i.e., frequency and proportion of bilingual children on their caseloads) to be more influential to their approaches towards bilingualism in SLP. Not only did it influence their perception of their competence but also their stance on language assessment in children’s heritage language and the relevance of the availability of bilingual norm data. Therefore, it can be assumed that

specific experience in working with bilingual children fosters linguistically diverse attitudes and approaches in SLPs.

### 3.6.3. SLPs' Language Background Related Effects

SLPs personal language background, that is whether participants were able to speak more than one language at the native level themselves, turned out to also have an impact on their approaches towards assessing bilingual children's language performance. It can be hypothesized that the SLPs' personal language background (being monolingual vs. bilingual) is associated with the habit of reflecting about cross-linguistic phenomena and the potential interaction between different languages, which subsequently would increase SLPs awareness of the importance of examining children's language performances in all of their languages.

### 3.6.4. Discrepancies between Knowledge and Practice

I found widespread awareness of bilingual approaches and bilingually oriented attitudes among German SLPs. The majority is open to learning about and to using new and suitable materials and techniques. However, there seem to be some obstacles that make the actual application of such materials and techniques difficult. I offer a two-fold explanation for this discrepancy between knowledge and practice, namely, that the currently available materials are insufficiently usable because (1) they are only applicable for a small number of children on a therapist's caseload and/or not affordable and (2) they require the therapist's manual evaluation of children's responses, which presumes that the therapist has already acquired a certain amount of knowledge about the respective heritage language, which can be very time-consuming. The gap between SLPs' high knowledge but low usage of diagnostic materials for bilingual children might thus be explained by the high demand on SLPs' resources (time and material) that would be required for the adequate assessment of bilingual children. Stankova and colleagues (2020) suggested that the discrepancy between clinical practice and SLPs' common beliefs/attitudes towards bilingualism reflects the fact that most of the countries sampled in previous and current research have been historically monolingual and that on a global level there is, thus, a lag in the clear appreciation of the need for professional linguistic competence in key minority languages.

### 3.6.5. Limitations and Future Perspectives

In order to provide a more complete picture of the individual differences in SLPs attitudes and approaches towards assessing and treating bilingual children, additional information could be gained from investigating whether and how well SLPs speak more common minority languages.

Further, participants' demographics (e.g., age and gender) should have been investigated in more detail in order to allow for a more in-depth interpretation of the results.

Data concerning SLPs' use of input-related information in the assessment of bilingual children could have been collected in a more systematic way (rather than using open-ended questions only) to allow the application of quantitative analyses and more generalizable interpretation of the results.

Future research should be concerned with the assessment of the usefulness of currently available diagnostic tools that aim to disentangle DLD from imbalanced bilingual language acquisition and look into the question of why they are not commonly applied in clinical practice. Furthermore, studies should compare language support services and SLP services in terms of their accessibility, costs and cost coverage, and success rate in order to evaluate the best way to cater for bilingual children's needs. Comparisons between countries in which SLP services are part of the education rather than the health care system might provide guidance in terms of how to comprehensively provide adequate treatment for bilingual children.

### 3.7 Conclusions

The data suggest that particularly having experience in working with bilingual children is associated with more bilingually-oriented attitudes and bilingualism-sensitive approaches in SLP service provision. In spite of the SDG 10 "Reduce inequalities" (United Nations, 2021), there is still a risk for residual inequalities in the provision of SLP services to bilingual children that should be tackled immediately. The additional efforts required to adequately address bilingualism-related issues should not lead to more inequality in the provision of SLP services but require substantial infrastructural and systemic changes.

The complexity of adequate language assessment related to bilingualism can only be met through increased resources and multi-professional awareness and knowledge (not only in SLPs but also in pediatricians who are responsible for children's referrals, teachers, and other professionals working with bilingual children). The various mechanisms involved in the parallel development of different languages, especially the role of language input and how it relates to the various expressions of typical language performance, should not be ignored but rather be acknowledged and correctly addressed in a synergic perspective in order to provide adequate assessment and treatment of DLD in bilingual children.

## 4. Bilingual Language Experience and Nonword Repetition<sup>9</sup>

The previous chapters have demonstrated the impact language exposure has on typical language development, how difficult it is to adequately quantify bilingual children's language experience and that there is no systematic approach applied by SLPs in incorporating these measures in their diagnostic process when assessing bilingual children's language abilities. This chapter is concerned with the development and use of a novel Nonword Repetition Task (NWRT), a diagnostic procedure in which children are acoustically presented a nonword (NW) and are asked to immediately repeat the unfamiliar sequence of sounds. NWs are defined as strings of phonemes, which unlike real words do not have meaning in the language of assessment but based on their suprasegmental, syllabic, phonological, and phonotactic properties could potentially be real words in this language.

The adequate identification of DLD in bilingual children has been widely acknowledged as a serious challenge in present-day Speech and Language Pathology (Armon-Lotem, 2012; for more in-depth description of the challenges SLPs face when trying to adequately assess bilingual children's language skills see section 1.1). Generally, previous research has identified NWRTs as promising tools in the clinical differentiation between TD children and children with DLD not only in the monolingual population but also in bilinguals (for an overview see Schwob et al., 2021). Children's NW performance skills have been found to be associated with vocabulary (Hoover & Storkel, 2006; Farabolini et al., 2021) and grammar skills (e.g., Rispens & Been, 2007) and can therefore be considered appropriate indicators of (bilingual) children's language development. The accurate repetition of a NW involves discrimination, encoding and production demands (Ebert, 2014). In particular, the demands of such tasks stem from a lack of lexical representation supporting NW repetition as typically verbal stimuli automatically prime the listener's word knowledge (and thus motoric routines), resulting in more successful (i.e., accurate) repetition of whole phonological sequences; if they are robustly supported by prior lexical knowledge (Cychosz, Erskine, Munson, & Edwards, 2021). Therefore, NWRTs are

---

<sup>9</sup> This chapter is based on (1) Bloder, Eikerling, & Lorusso (submitted to *Clinical Linguistics and Phonetics*). Language (un)specific Nonword Repetition Tasks for bilingual children – construction, implementation & validation of the MuLiMi NWRT; and (2) Eikerling, Bloder, & Lorusso (2022). A Nonword Repetition Task Discriminates Typically Developing Italian-German Bilingual Children from Bilingual Children with Developmental Language Disorder: The Role of Language-Specific and Language-Unspecific Nonwords. *Frontiers in Psychology*, 13, 1-17.

Note that for the purpose of this chapter, only data of a subgroup of participants, namely the bilingual TD children, will be reported in order to explore the effects of language experience on typical bilingual language development and children's nonword repetition performance. Therefore, the nonword selection process deviates slightly from the procedure described in Bloder, Eikerling, & Lorusso. Similarly, the correlation analyses are not identical to the ones performed in Eikerling, Bloder, & Lorusso.

considered increasingly relevant for the identification of DLD in bilingual children as their phonological processing abilities can be assessed directly, irrespectively of their prior vocabulary knowledge that might be reduced compared to their monolingual peers due to variations in language input and exposure. Especially in bilingual children, NWRTs have thus been argued to be advantageous over other diagnostic tasks as children's performance is less subjected to the influence of language experience in the language of assessment (Chiat, 2015) and that therefore children with limited exposure to this language may be less disadvantaged. However, when trying to create a valid NWRT, several NW characteristics have to be taken into account as they can have a significant impact on children's repetition performance.

#### 4.1. Nonword Characteristics and Children's Repetition Performance

As many of the world's languages differ substantially in their phonological, phonotactic, morphological, and suprasegmental properties, the degree to which a particular NW is in line with these characteristics (i.e., its degree of **language specificity**) can determine repetition accuracy in bilingual children (Chiat, 2015). This influence of **language specificity** has been argued to be mediated by children's experience with a given language and depends on whether and how strongly the characteristics of a child's languages diverge or coincide. The following characteristics may contribute to a NW's degree of language specificity:

One factor that is often considered to influence monolingual and bilingual children's NW repetition scores is NW length, more specifically defined in terms of the **number of syllables** that constitute a NW. Most commonly, NW length has ranged from one to five syllables in previous research. With respect to the language-specificity of NW length Summers, Bohman, Gillam, Peña, and Bedore (2010) found that experience-effects may not necessarily be limited to one language and may be visible in children's NW repetition performance regardless of the language a NW reflects. In their study, children with greater experience in Spanish (a highly inflectional, multi-syllabic language) performed better when repeating English four-syllable NWs than bilingual children with more English experience (see also Gibson et al., 2015). Nevertheless, in general, both TD and DLD children's repetition accuracy decreases with increasing NW length (e.g., Bortolini, Arfe, Degasperi, Deevy, & Leonard, 2006; Dispaldro, Leonard, & Deevy, 2013). However, children with DLD have been found to perform even less well than their TD peers across all NW lengths (Boerma et al., 2015).

Furthermore, the degree of **complexity of syllable structures** (simple CV vs. more complex CCV structures) and thus articulatory complexity has been found to impact children's NW repetition performance. Jones and colleagues (2010), for instance, found that children with



DLD had more difficulties than TD children when NW structures were complex. In particular, the presence of consonant clusters has been found to lead to a decrease in repetition accuracy in both mono- and bilingual children with DLD (dos Santos & Ferré, 2018). Italian and German, for instance, differ in terms of their common syllabic structure. While German allows clusters of two to three consonants both in onset and coda position (i.e., before and after the vowel), Italian – although allowing up to three consonants in onset position – only permits a small number of coda consonants and therefore has lower syllable complexity than German.

Archibald and Gathercole (2007) found that **prosodic features** of NWs also had an impact on children's repetition performance. The authors compared monolingual TD and DLD children's repetition performance of NWs (CVC sequences) that were either presented as a list of isolated monosyllabic units or coarticulated as a single sequence of syllables that also incorporated prosodic cues. Although both TD and DLD children scored higher when NW syllables were prosodically organized, the benefit was greater for the TD group. Therefore, the authors suggested that children with DLD are less sensitive to the familiarity of prosodic patterns of their surrounding language. Thus, it can be assumed that prosodic cues are less helpful to children with DLD when storing and reproducing sound sequences and therefore NWRs that incorporate language-specific prosody can be hypothesized to have greater discriminative potential (Roy & Chiat, 2004). Furthermore, it has been found that monolingual children find it difficult to repeat unstressed syllables in word positions that are untypical in their native language (Chiat & Roy, 2007; Sahlén, Reuterskiöld-Wagner, Nettelbladt, & Radeborg, 1999). Therefore, the probability of particular stress patterns in bilingual children's respective languages should also be taken into account. German, for instance, is a trochaic language (i.e., a strong syllable typically precedes the weak one(s)), whereas Italian typically applies the iambic stress pattern (the strong syllable typically follows the weak syllable). Thus, German words are most frequently stressed on the first syllable (Mengel, 2000), whereas in Italian the most frequent case is stress on the penultimate syllable (Delmonte, 1981). In both Italian and German, lexical stress can be phonologically distinctive (e.g. *um'fahren* [to drive around an obstacle] vs. *'umfahren* [to run over something/someone] in German; *compito* [task] vs. *com'pito* [polite] in Italian).

Languages further differ in their respective **segmental repertoires**. Standard Italian, for instance, features seven vowels and 20 consonants (all of which can geminate; Kramer, 2009), compared to standard German that roughly includes 14 vowels and 26 consonants (although different authors have proposed somewhat divergent lists; Wiese, 1996). Furthermore, the two languages differ in terms of the phonetic realization of some contrasts (e.g., VOT, see chapter

2 for a detailed discussion of this matter). The influence of familiarity with language-specific speech sounds has been documented by Sharp and Gathercole (2013), who found differences in children's repetition accuracy of Welsh consonants in bilingual Welsh-English children depended on their experience with the language. Moreover, the ways in which single phonemes combine with each other can also be specific to different languages. For instance, Italian allows the combination of /s/ and any voiceless stop consonants (/p/, /t/, /k/) or the fricative /f/ in word onset (e.g. *sfruttare* [to exploit]), whereas these combinations are not permissible in German. **Phonotactic legality** therefore should also be taken into account when designing a NWRT for bilingual children.

Finally, the influence of **word-likeness**, the degree of resemblance between a NW and a real-word, on children's repetition performance has also been investigated in previous studies. A NW's degree of word-likeness increases if a NW contains a syllable that is a real morpheme (i.e. a whole word or an affix) in the language of assessment (Casalini et al., 2007). And the most word-like NW stimuli, are generated by simply exchanging the vowels of a real word (e.g., *Schukulide* instead of *Schokolade* [chocolate] in German; cf. the German NWRT PhoMo-Kids; Stadie & Schöppe (2013). Gathercole (1995) found that whether a NW was more or less word-like affected TD children's repetition performance but not only TD children, also children with DLD have been found to perform significantly better on NW stimuli with a higher degree of word-likeness (Jones et al., 2010). Considering the influence lexical knowledge has on children's NW repetition performance (Cychosz et al., 2021), especially for bilingual children the effects of word-likeness should be taken into account since children's vocabulary and thus their familiarity with the phonetic and phonemic properties of the target language is dependent on the amount of their language exposure.

Because of these effects of language experience, language-specific NWRTs have often been claimed to be putting bilingual children at a disadvantage because their language experience is spread across two languages that potentially differ with respect to their lexical, phonological, and suprasegmental characteristics. And even though generally children's performance in NWRTs is less affected by their prior language knowledge (compared to other diagnostic procedures), some studies have shown that they are indeed not completely free from language-specific influences and that bilingual children performed worse on nonword stimuli that were based on the phonological characteristics of their lesser known language (e.g. Boerma et al., 2015). It can be assumed that features that are not present in the repertoire of a child's native language(s) will be more difficult to produce, whereas experience with a certain feature will be an advantage in repeating NWs that mimic this feature. While NWRTs that rely on such

language-specific attributes and therefore to a certain extent tap into specific knowledge of a given language have been found to be more challenging for (monolingual) children with language difficulties than for TD children (Archibald & Gathercole, 2007) and thus seem to have greater potential in identifying DLD, they are also more sensitive to effects of language experience in bilingual children which may lead to misdiagnoses of DLD in this population.

In an attempt to approach the diagnostic dilemma that is tightly linked with the adequate identification of DLD in bilingual children, many studies have investigated the usability of NWRs. However, the extant literature provides controversial evidence as to which type of NWs (i.e. language-specific vs. more neutral, i.e. unspecific stimuli) is required to account for children's varying degrees of language experience with the language of assessment in order to successfully differentiate between weak language performance due to underexposure and actual language impairment in bilingual children. The number of conclusions to be drawn from previous research is limited, especially, since findings on whether and how language experience and repetition performance on language-specific NWs are intertwined, are mixed. In this chapter, I explore whether it's more advantageous to assess bilingual children's language skills by using NWRs that take into account the characteristics of both of their native languages or whether it is more beneficial to use NWs that are (relatively) free from such language-specific attributes.

First, I will provide an overview of previous findings regarding the effects of language experience on language-specific versus unspecific NWRs, before I will describe the process of creating a novel Italian-German NWR and examine the effects bilingual Italian-German children's language experience on their repetition performance.

## 4.2 Language-Specific Nonword Repetition Tasks

Table 12 provides an overview of studies that examined TD bilingual children's NW repetition performance on language-specific NWRs. Several of these studies suggest that performance on language-specific NWRs is independent of specific language experience with the characteristics of a particular language. Thordardottir and Juliusdottir (2013), for instance, found that their five-to-seventeen-year-old participants who were all L2 learners of Icelandic and in general did not perform well on standardized Icelandic tests of language knowledge, scored uniformly high on an Icelandic NWR independent of when and for how long they had been exposed to the L2. Moreover, their repetition accuracy for both NWs with familiar Icelandic word endings and stress patterns (i.e., word-like) as well as NWs without inflectional Icelandic word endings and unfamiliar stress patterns (i.e., nonword-like) was well over 90%.

This performance is similar to accuracy achieved by nine-year-old monolingual TD speakers of Icelandic and even more so, considerably better than accuracy observed in Icelandic monolinguals with DLD (Thordardottir, 2008). Using English NW stimuli, Lee and Gorman (2012) did not find any significant differences in overall repetition accuracy among four different groups of children (monolingual English versus bilingual Korean-English, Spanish-English, and Chinese-English children). However, the authors find that at the age of seven their participants had had sufficient experience with English and thus had already established sufficiently strong phonological representations of their L2 in order to perform equally well as their monolingual English peers.

In contrast, Engel de Abreu (2011) found a significant difference between the NW repetition performance of monolingual Luxembourgish children and a group of simultaneous bilingual children, living in Luxembourg and acquiring Luxembourgish and another European language from birth. Despite Luxembourgish being the bilinguals' dominant language, they performed with less accuracy than their monolingual peers on an NWRT that included items that conformed to the phonotactic properties of Luxembourgish. The effect disappeared once children's lexical knowledge was taken into account, suggesting that the difference in repetition performance was moderated by children's language experience and thus language knowledge of Luxembourgish.

Overall, the heterogeneity in findings may be explained by the great variability in methodological approaches that were applied in these studies, such as differences in the construction of the NWRTs, and thus the particular characteristics of the NWs. Further differences in methodologies concern the language of assessment (children's L2 in Thordardottir & Juliusdottir, 2013; Lee & Gorman, 2012; Engel de Abreu, 2011; Farabolini et al., 2021; children's L1 in Lee et al., 2013) and children's distribution of language dominance (only reported in Engel de Abreu, 2011), which could have led to either the over- or under-estimation of the influence of language experience.

**Table 12.** Overview of previous studies using language-specific nonword repetition tasks in one language with TD bilingual populations.

Authors	NWRT	Participants	Results
Engel de Abreu (2011)	NWs that conformed to the phonotactic properties of Luxembourgish	6 y.o. monolingual Luxembourgish vs. simultaneous bilingual children speaking Luxembourgish in	Despite Luxembourgish being their dominant language, bilingual children performed worse than their monolingual peers

		combination with various other languages	
Farabolini et al. (2021)	NWs followed the phonotactic constraints of Italian	3-5 y.o. L2 Italian speakers with various L1s	NW repetition performance correlated with parents' estimates of their children's receptive vocabulary knowledge in Italian but not with measures of language exposure
Lee & Gorman (2012)	NWs followed the phonotactic constraints of English	7 y.o. TD monolingual English vs. bilingual Korean-English, Spanish-English, and Chinese-English children	No significant differences between the four groups were found
Lee, Kim, & Yim (2013)	NWs followed the phonotactic constraints of Korean	4 y.o. TD children; bilingual Korean-English vs. monolingual Korean	Both groups performed equally well
Thordardottir & Juliusdottir (2013)	Icelandic word-like and nonword-like NWs	5-17 y.o. TD L2 learners of Icelandic	High repetition accuracy for both types of NWs

### 4.3 Nonword Repetition in Bilingual Children's Two Languages

Table 13 provides an overview of studies that used two separate language-specific NWRs to address the influence of bilingual children's language exposure on their NW repetition performance in greater detail. Comparing two studies that looked at NW repetition performance in both languages of relatively young bilingual TD children, one found that children's repetition performance for English NWs but not for Spanish NWs was positively correlated with their language input in English (Parra, Hoff, & Core, 2011), whereas the other found no association between their bilingual Spanish-English participants' language experience and their performance on an English and a Spanish NWR (Core, Chaturvedi, & Martinez-Nadramia, 2017). Note that the participants in Core et al. (2017) were older than the participants included in the study by Parra et al. (2012) (mean age 30 vs. 22 months), suggesting that the effects of language experience may be more observable in the early stages of language exposure and that once a certain threshold of experience has been reached, its influence on phonological memory diminishes. Thordardottir and Brandeker (2013) found no correlation between their four-to-five-year-old simultaneous bilingual English-French participants' experience in their two respective languages and their performance on the French NWR but they did find a correlation with the English NWR. While the French task included only NWs with CV syllable structure,

without diphthongs and without consonant clusters, the English NWs had more diverse syllable structures (CV, CCV, and CVC), and included diphthongs and consonant clusters. Arguably rendering the English NWRT more language-specific than its French counterpart, and thus potentially more susceptible to the influence of specific language experience. Along the same lines, Summers and colleagues (2010) pointed out that if two language-specific NWRTs were not necessarily matched on phonological difficulty, any comparison of children's repetition accuracy between the two languages may not be reliable.

Despite the interesting theoretical implication that can be derived from these studies on TD children, it is necessary to actually compare TD and DLD children's NW repetition performance in order to consider the diagnostic accuracy of NWRTs. To this end, also using two language-specific NWRTs (one in Spanish and one in English), Gutiérrez-Clellen and Simon-Cerejido (2010) showed that Spanish-English bilingual TD children scored significantly higher in both tasks than their bilingual peers with DLD. However, children's repetition performance in only one of the two languages (i.e. either only English or only Spanish) resulted in inadequate specificity and sensitivity rates. The authors, thus, suggest that assessment only in children's dominant language may lead to the reduction of diagnostic accuracy and an increase in the risk of underidentification of DLD, and therefore highlight the importance of bilingual approaches (i.e., assessing repetition performance in both languages of a child).

The immense complexity and the entangled interplay of various factors in bilingual language development becomes even more apparent when comparing bilingual TD versus DLD children's NWRT performance with monolingual TD versus DLD children's NW repetition accuracy. Ahufinger and colleagues (2021), for instance, found that both bilingual DLD and TD children showed worse NW repetition performance in two language-specific NWRTs representing their two native languages than their monolingual peers. They argued that this was because reduced language exposure leads to reduced long-term linguistic representations, and thus reduced accuracy on language-specific NWRTs in bilinguals. They further argued that in bilingual children with DLD, language exposure had an even greater influence on their NW repetition performance because these children in particular have "limited linguistic representations within their extant long-term memory" (Ahufinger et al., 2021, p. 24). Moreover, when comparing five-to-six-year-old monolingual Russian/Hebrew and bilingual Russian-Hebrew speaking TD versus DLD children, Armon-Lotem and Meir (2016) found that for Russian (bilinguals' L1) the cut-off between TD and DLD children was higher in the bilingual than in the monolingual group (0.79 vs. 0.71) which was interpreted to reflect

bilinguals' advantage in NW processing due to being used to greater phonetic unfamiliarity in their everyday language environment. Conversely, for Hebrew (bilinguals' L2) the cut-off was lower in the bilingual group than in their monolingual peers (0.71 vs. 0.86) which subsequently was interpreted to reflect bilinguals' limited experience with their L2 (Hebrew). Similarly, Windsor, Kohnert, Lobitz, and Pham (2010)'s results reflect the intertwined effects of specific language experience as well as the effects of the state of children's language system (i.e., DLD vs. TD) on their NW repetition performance.

**Table 13.** Overview of previous studies using language-specific nonword repetition tasks in both languages of bilingual TD and DLD populations.

<b>Authors</b>	<b>NWRT</b>	<b>Participants</b>	<b>Results</b>
Ahufinger et al. (2021)	3 different NWRTs following the phonotactic constraints of each target language (Catalan, Spanish, Portuguese)	5-16 y.o. bilingual Catalan-Spanish speaking TD & DLD vs. 7-11 y.o. monolingual Portuguese speaking TD & DLD children	Bilingual and monolingual children with DLD achieved greater repetition accuracy for NWs with high wordlikeness and high phonotactic frequency compared to low wordlikeness and low phonotactic frequency, while in TD controls this effect was not as extreme
Armon-Lotem & Meir (2016)	Russian-specific as well as a Hebrew-specific NWRT	5-6 y.o. monolingual Russian/Hebrew and bilingual Russian-Hebrew speaking TD vs. DLD children	Both tasks showed good efficacy in discriminating bilingual children with and without DLD, especially if bilingual cut-off points were applied
Core, Chaturvedi, & Martinez-Nadramia (2017)	12 English-specific vs. 12 Spanish-specific NWs	30 m.o. bilingual Spanish-English TD children and an aged-matched monolingual English control group	Children's NW repetition performance was not related to measures of relative amount of language input in either of the two languages No difference between bilingual and monolingual children with respect to their NW repetition accuracy of English-specific NWs
Gutiérrez-Clellen & Simon-Cerejido (2010)	English-specific as well as a Spanish-specific NWRT	3-6 y.o. bilingual Spanish-English TD. vs. DLD children	Bilingual TD children scored significantly higher in both tasks compared to their bilingual peers with DLD

Parra, Hoff & Core (2011)	12 English-specific vs. 12 Spanish-specific NWs	22 m.o. bilingual Spanish-English TD children	Children's repetition performance for English NWs but not for Spanish NWs was positively correlated with their language input
Summers et al. (2010)	English-specific as well as a Spanish-specific NWRT	4-6 y.o. bilingual Spanish-English TD children	Children performed better on repeating NWs in their first language (Spanish) than their second language (English).
Thordardottir & Brandeker (2013)	English-specific as well as a French-specific NWRT	4-5 yo. simultaneous bilingual English-French TD children	High performance in both NWRTs across the entire continuum of bilingual exposure patterns
Windsor et al. (2010)	English-specific as well as a Spanish-specific NWRT	8 y.o. bilingual Spanish-English speaking TD & DLD vs. 8 y.o. monolingual English speaking TD & DLD children	Spanish-English TD children performed better on the Spanish NWRT than their monolingual English peers and conversely monolingual TDs performed better than bilingual TDs on the English NWRT, which was also true for the two DLD groups. Furthermore, bilingual TD children performed better than monolingual DLD children on the English NWRT

Taken together these findings suggest, that even when applying a bilingual approach by assessing NW repetition performance in both a child's languages, children's performance may be dependent of the respective level of specificity of a given NWRT and thus the greater influence of experience associated with higher levels thereof. Moreover, differences in results concerning the influence of language-specific knowledge on children's NW repetition performance may not only stem from differences in NWs used but also be dependent on how language experience was defined and measured across the different studies. Generally, gathering information about children's individual language background through parents' and/or teachers' reports (e.g. through questionnaires) has proven to be a useful method (Pua, Lee, & Rickard Liow, 2017). However, reiterating on what has been described in great detail in section 1.2, the adequate quantification of language experience has been proven to be difficult and there is no best-practice approach (Hoff, 2020). Age at first onset of exposure (e.g. in Thordardottir & Juliusdottir, 2013), current exposure (e.g. in Sharp & Gathercole, 2013), and cumulative exposure (e.g. in Thordardottir, 2017) have previously been used to measure the influence of language experience on NW repetition performance. However, as previously pointed out by



Farabolini and colleagues (2021), up to this point, no study has considered the effects of children's active language use (i.e. the relative amount of children's production in either of their two languages, their language output) on children's accuracy in NW repetition. When comparing children's repetition performance of real words and NWs, Dispaldro and colleagues (2013) found significantly reduced scores in DLD children for NW repetition accuracy compared to real-word repetition accuracy, as – so the authors find – when repeating real words children can rely on well-learned production routines. In fact, these routines may be more beneficial in the repetition of language-specific NWs than in the repetition of unspecific NWs as higher phonological frequency and/or higher word-likeness (i.e., the resemblance between a NW and a real word) have been found to result in higher NW repetition accuracy in children (Munson, Kurtz, & Windsor, 2005).

To summarize, previous results are inconclusive in terms of the specific effects language experience poses on bilingual children's NW repetition performance. Some findings are even contradictory in nature, but discrepancies may be explained by the different methodological approaches applied and/or the heterogeneity in study designs (i.e., participants' age ranges or particular NW characteristics), which may have made some study participants appear to be more or less susceptible to effects in language experience.

#### 4.4 Language-unspecific Nonword Repetition Tasks

Work on NW repetition in the Language Impairment Testing in Multilingual Settings (LITMUS) test battery within the framework of COST Action IS0804 has addressed this issue of language experience related effects in bilingual children in great detail. To minimize the potentially confounding effects of language experience on bilingual children's NW repetition performance, they suggest the use of language-so-called “(quasi) universal” (i.e. and henceforth language-unspecific) NWs that are compatible with a great range of language typologies (Chiat, 2015). These items were sought to be designed to be compatible with a wide array of cross-linguistically diverse constraints on lexical phonology. Additionally, NWRTs constructed in accordance with the LITMUS principles also include a set of language-specific NWs in order to address the empirical question which type of NWs can best balance out the competing demands of having the highest discriminative power (i.e. successfully differentiating between TD and DLD children) but at the same time being least influenced by language experience (i.e. being applicable for monolingual and bilingual children alike) and is thus suitable for a wide array of languages and language combinations.

Despite the fact that within the COST Action framework NWRTs following these principles have been designed for a great variety of different languages (e.g. German, French, English, Lebanese Arabic, Luxembourgish, Serbian, Hebrew, Lithuanian, Slovakian, Turkish among others; for an overview see Chiat, 2015), to my knowledge, only few studies have been published on the efficacy of the task and even fewer have focused on the comparison of children's repetition accuracy between language-specific and unspecific NWs.

Using the French LITMUS NWRT, de Almeida and colleagues (2017) found no effect of language experience and dominance on the repetition performance of bilingual TD and DLD children with either Arabic, European-Portuguese, or Turkish as their L1. Although repetition performance of language-specific compared to unspecific NWs was not analyzed separately, the authors found high specificity and sensitivity rates (around 80%) in their bilingual samples for their NWRT. This further corroborates the notion that even language-specific NWRTs – if at all – only require very little exposure to the language of testing in order to adequately identify DLD in bilingual children. These findings have been further substantiated for French and also replicated with the German version of the LITMUS NWRT (Tuller et al., 2018). Also using the French LITMUS NWRT, that comprises French language-specific and unspecific NWs, but focusing on syllable complexity (i.e. the presence of consonant clusters and word-final consonants), dos Santos and Ferré (2016) did not find a significant difference when comparing monolingual and bilingual children's repetition performance. Based on the assumption that language experience influences bilingual children's repetition performance of language-specific NWs, the authors hypothesized that bilinguals would show more difficulties in repeating the language-specific items than their monolingual peers but that there would not be a difference between the two TD groups for the language-unspecific items. However, monolingual and bilingual children performed equally well independent of the respective type of NW (language-specific vs. unspecific). Using the German LITMUS NWRT with monolingual German and bilingual Arabic-German and Turkish-German TD and DLD children, Abed Ibrahim and Hamann (2017) neither found any significant differences between monolingual TD and bilingual TD nor between monolingual DLD and bilingual DLD children on either of the types of NWs (i.e. language-specific vs. unspecific). In fact, although not statistically significant, bilingual children with DLD slightly outperformed monolingual children with DLD on the language-specific German NWs, suggesting that linguistic diversity through exposure to two languages may have had beneficial effects on bilingual children's executive functions and thus repetition performance. Moreover, children's performance on

language-specific items proved to be more indicative of DLD in both monolingual and bilingual participants.

More globally, when using the Dutch adaptation of the LITMUS NWRT with only language-unspecific NWs, Boerma and Blom (2017) found that combining children's language background information derived from a parental questionnaire with this direct measure of language-unspecific skills was a reliable method to identify DLD in both monolingual and bilingual children. Nevertheless, the evidence in favor of using language-unspecific NWs when assessing bilingual children for DLD is weak and given the limited results on actual differences between language-specific versus unspecific items within the COST Action framework, the empirical question concerning which particular features and which degree of specificity influence children's repetition performance and how this influence varies as a function of different language exposure/dominance scenarios, to my knowledge, has not yet been answered satisfactorily. Given that the a truly language-unspecific NWRT is not a possibility, there is no clear-cut distinction between language-specific and unspecific NWs as the construct of specificity is spread fluidly across a continuum. The language-unspecific (cf. "quasi universal" within the LITMUS framework) items as described by Chiat (2015) are presented with the phonetic properties of the language of assessment, which – as also noted by the authors – may differ from children's L1 and may thus convey a certain degree of language specificity. Additionally, within the several LITMUS NWRTs, language-specific and unspecific items were created similarly with the only exception that for the language-specific NWs a few additional phonemes were used (e.g., Grimm & Hübner, 2017 for the German version of the LITMUS NWRT). So, especially when language-unspecific items were presented with the prosody of the language of assessment, language-specific and language-unspecific items may have been too similar in order to yield significant differences in children's repetition performance.

#### 4.5 Research Questions/Hypotheses

Inspired by the discussion raised within COST Action IS0804, the present study aims to follow up on the question concerning which type of NWs is most informative about children's language ability status. To do this, a novel NWRT that comprises different types of NWs was constructed and implemented. It comprises both (1) language-specific items that take children's language experience/ dominance into account and (2) language-unspecific items that are supposed to assess children's language abilities regardless of their language background. The goal of the study is to examine children's repetition performance as a function of NW

characteristics in relation with their bilingual language exposure situation in order to identify the optimal type of NWs or combination thereof to subsequently improve the clinical assessment of DLD in bilingual children. Therefore, I pose the following research questions and hypotheses:

**RQ1:** Depending on children's dominance status (German-dominant vs. balanced vs. Italian-dominant) does their NW repetition accuracy differ as a function of NWs' language specificity (i.e., conforming to the phonotactic constraints of either of the two languages)?

Hypothesis 1: If language experience modulates children's performance on repeating language-specific NWs (Chiat, 2015), I expect that German-dominant bilinguals will outperform Italian-dominant bilinguals on NWs that conform to the phonotactics of German and vice versa, while balanced bilinguals should show equal repetition accuracy independent of language specificity.

**RQ2:** Do measures of children's current relative language input correlate with their NW repetition accuracy of language-specific versus language-unspecific items?

Hypothesis 2: If language experience, specifically hearing a certain language (i.e., children's current relative language input), modulates children's performance on repeating language-specific NWs (Sharp & Gathercole, 2013), I expect there to be a significant correlation between the proportion of children's German input and their performance on language-specific NWs that conform to the phonotactic constraints of German. Similarly, I expect a correlation between the proportion of children's Italian input and their performance on language-specific NWs that conform to the phonotactic constraints of Italian but no significant correlations between children's language input in either German or Italian and their repetition accuracy for language-unspecific NWs.

**RQ3:** Do measures of children's current relative language output correlate with their NW repetition accuracy of language-specific versus language-unspecific items?

Hypothesis 3: If language experience, specifically the use of a certain language (i.e., children's current relative language output), modulates children's performance on repeating language-specific NWs, as it helps them to practice language-specific motoric routines (Dispaldro et al., 2013; Munson, Kurtz, & Windsor, 2015), I expect a correlation between the proportion of children's German output and their performance on language-specific NWs that conform to the phonotactic constraints of German. Similarly, I expect a correlation between the proportion of children's Italian output and their performance on language-specific NWs that conform to the phonotactic constraints of Italian.

**RQ4:** Does children's lexical knowledge correlate with their NW repetition accuracy of language-specific NWs that correspond with the phonotactic constraints of the target language?

Hypothesis 4: If children’s repetition of language-specific NWs is supported by their prior lexical knowledge (Cychosz et al., 2021; Engel de Abreu, 2011), I expect there to be a correlation between children’s vocabulary size in German or Italian and their NW repetition accuracy of language-specific NWs that follow the phonotactic constraints of the respective language.

## 4.6 Methodology

### 4.6.1 Participants

Twenty-six simultaneous or early-sequential bilingual Italian-German speaking children between the ages of 3;10 to 6;3 years ( $M = 63$  months,  $SD = 9$  months) participated in this study. At the time of their participation in this study, all children were living in Germany. They all had at least one native Italian-speaking parent and were exposed to Italian – even though to varying degrees – on a daily basis. Participants with two Italian-speaking parents had been exposed to the German language for at least two years. While one of the children was attending a monolingual German kindergarten, 25 were enrolled in a bilingual Italian-German kindergarten program. The majority of these children also participated in Study I (see section 2.4.1). The study was approved by the Ethical Committee of the Catholic University Eichstätt-Ingolstadt and all parents signed informed consent according to the Declaration of Helsinki.

### 4.6.2 Assessment tools

To answer the research questions, children’s NW repetition performance, receptive lexical knowledge in Italian and German, and their language experience and current exposure patterns were assessed.

**MuLiMi Nonword Repetition Task.** In order to be able to link children’s NW repetition performance with their language experience in Italian and German respectively and to further be able to directly compare children’s NW repetition performance in the societal compared to their heritage language, a novel set of language-specific (LS) German and Italian NWs, following the exact same principles of construction, were developed for the purpose of this study. First, to make sure that LS NWs complied with the phonotactic constraints of their respective language, language-specific phonemes and consonant clusters were identified (see Orzechowska & Wiese, 2015 for German; and Baroni, 2012 for Italian). While I aimed for maximal language specificity in LS NWs, the goal for language-unspecific (LU) NWs was to minimize the potential effects of varying degrees language experience on children’s repetition

performance. Therefore, in order to create a set of LU NWs, only phonemes common to both languages were selected (differences in the phonetic realization of these phonemes e.g., voicing for stop consonants were considered tolerable). As opposed to the LS items, no consonant clusters were included in the LU NWs. Thus, the LS items could be considered more complex than the LU NWs (see also dos Santos & Ferré, 2016).

As the target group of participants was defined as pre-school children, aged 4;0-5;11 years, I further decided not to include late acquired phonemes in the construction of the NWs (as previously suggested by Mathieu, Lindner, Lomako, & Gagarina, 2016) in order to ensure that children's potential repetition errors would not be due to speech production difficulties considered typical in the course of phonological development (note that the selection of early vs. late acquired phonemes was based on the phonological development trajectories described for monolingual TD children; see Fox, 2000 for German; and Bortolini, 1995 for Italian).

An original pool of 23 German-specific (LSger), 24 Italian-specific (LSit), and 17 LU NWs was created, each comprising two-, three-, up to four syllable lengths. Because of floor effects previously found in preschool children (see Guiberson & Rodríguez, 2015), no five-syllable NWs were included in the MuLiMi NWRT.

LS NWs were recorded by a female native speaker of the respective language. In order for LS NWs to be reflective of language-specific word stress, during the recordings, speakers were asked to produce the LS NWs with the prosodic features that felt most natural to them. By contrast, LU NWs were recorded by native speakers of both languages. During the recordings, speakers were instructed to produce LU items with flat, neutral prosody (carefully avoiding placing word stress on any specific syllable in a NW) so as not to reflect any pattern of lexical prosody that could be typical of a specific language.

Ahufinger and colleagues (2021) demonstrated that relying on native speakers' ratings of self-constructed NWs results in a valuable measure that can increase the diagnostic sensitivity a NWRT. Therefore, a two-step rating procedure was employed in order to select the NWs to comprise the final MuLiMi NWRT (i.e., maximally language-specific LS NWs, compared to as unspecific as possible LU NWs). First, in the stage of the so-called "direct ratings", after auditory presentation, adult monolingual native speakers of either Italian or German ( $n = 5$  to  $7$ ) were asked the repeat (1) all of the LS NWs of their respective native language and (2) all LU items also recorded by the speaker of the same native language. Based on their responses, a repetition score for each NW was computed (1 = repetition was identical to the target; 0 = the participant's production did not match the target) Additionally, adult participants were asked to rate each NW's L1-Alikeness and Pronounceability on a Likert-scale

from 1-5 (1 = difficult to pronounce and not L1-alike respectively; 5 = easy to pronounce and L1-alike respectively). Table 14 provides an overview of the NWs' characteristics as judged by adult native speakers in the direct ratings.

**Table 14.** LS and LU NWs' characteristics as judged by adult native speakers of Italian and German in the direct rating procedure.

	L1 German direct raters <i>n</i> = 5 or 7		L1 Italian direct raters <i>n</i> = 5 or 7	
	LSger NWs <i>n</i> = 23	LU NWs <i>n</i> = 17	LSit NWs <i>n</i> = 24	LUNWs <i>n</i> = 17
Pronounceability 1 (difficult) - 5 (easy)	<i>M</i> = 3.68, <i>SD</i> = .57	<i>M</i> = 3.64, <i>SD</i> = .74	<i>M</i> = 4.00, <i>SD</i> = 0.49	<i>M</i> = 3.69, <i>SD</i> = .74
Repetition accuracy (%)	<i>M</i> = 92.69, <i>SD</i> = 11.21	<i>M</i> = 84.70, <i>SD</i> = 21.83	<i>M</i> = 97.50, <i>SD</i> = 6.76	<i>M</i> = 72.94, <i>SD</i> = 21.14
L1-Alikeness 1 (not L1-alike) - 5 (L1-alike)	<i>M</i> = 3.39, <i>SD</i> = .62	<i>M</i> = 2.49, <i>SD</i> = .73	<i>M</i> = 3.82, <i>SD</i> = .59	<i>M</i> = 1.62, <i>SD</i> = .87

In general, LU items had lower L1-Alikeness ratings than the LS NWs. Further Repetition Accuracy on average was higher for LS compared to LU NWs. Nevertheless, both LS and LU items were judged to be overall relatively easily pronounceable by both Italian and German adult speakers. Based on results of these direct ratings, the number of NWs was narrowed down and a pre-selection was made for further evaluation. For the second round of ratings, only LS NWs with overall high (Repetition-Accuracy, L1-Alikeness, and Pronounceability) scores were selected, whereas for the LU NWs the goal was to include items that were easily pronounceable but as distant from all involved languages (i.e. least L1-alike) as possible (18 LSger; 21 LSit; 16 LU). These remaining NWs were then implemented in an online rating-task using Qualtrics Software (2020) for the purpose of a larger scale evaluation. 27 German and 27 Italian native-speaking adults participated in this online rating procedure. The raters were invited by email and they were recruited among the experimenters' co-workers and acquaintances. Most of them spoke other languages (usually English) besides their L1, but they did not speak the other target language (i.e., Italian raters did not speak German and German raters did not speak Italian). While similar to the direct ratings, participants in the indirect ratings did only judge the LU items recorded by a native speaker of their own language, online raters were also asked to evaluate all of the remaining LS NWs, thus including also the ones that did not follow the

phonotactic constraints of their native language. Similar to the direct ratings, participants were asked to rate each NWs Pronounceability on a Likert-Scale from 1 to 5. Furthermore, in order to verify L1-Alikeness, raters had to allocate each NW to the language it could most likely belong to (for this Language Allocation raters were asked to choose one out of five options, German, Italian, Spanish, English, other). NWs were presented in random order (see Table 15 for an overview of the NWs' characteristics as assessed during the online rating procedure).

**Table 15.** LS and LU NWs' characteristics as judged by adult native speakers of Italian and German in the online rating procedure.

	L1 German online raters <i>n</i> = 27		L1 Italian online raters <i>n</i> = 27	
	LSger NWs <i>n</i> = 18	LU NWs <i>n</i> = 16	LSit NWs <i>n</i> = 21	LU NWs <i>n</i> = 16
Target Language Allocation (% of allocation to raters' native language)	<i>M</i> = 82.12, <i>SD</i> = 13.45	<i>M</i> = 63.23, <i>SD</i> = 13.82	<i>M</i> = 92.78, <i>SD</i> = 6.97	<i>M</i> = 28.88, <i>SD</i> = 17.86
Pronounceability 1 (difficult) - 5 (easy)	<i>M</i> = 3.90, <i>SD</i> = 0.42	<i>M</i> = 4.21, <i>SD</i> = 0.25	<i>M</i> = 4.45, <i>SD</i> = 0.19	<i>M</i> = 4.01, <i>SD</i> = 0.30

Similar to the direct ratings, Pronounceability scores were uniformly high for both LS and LU NWs. Generally, LS NWs were more often allocated to the target language than their LU counterparts. However, overall, LU NWs' Target Language Allocation was more heterogenous with German native speakers more often allocating the LU NWs to their L1 than Italian natives. In order to counterbalance the variability within LU L1-Alikeness/Target Language Allocation scores, it was decided to include those recordings of the native German speaker for LU NWs with high ratings of Italian specificity and vice versa. In order to consolidate the distinctness between LS and LU items, only LS NWs that were allocated to the target language by > 70% of the adult native-speaking raters were included in the final task, whereas only LU NWs that were allocated to both target languages by ≤ 60% of the respective native speaking adult raters were selected. Furthermore, in order to ensure large variation in NWs' Pronounceability while avoiding floor effects, an average rating of at least 2.00 was defined as the cut-off. For the same reasons, LS and LU NWs with Adult Repetition Accuracy > 66.6% (i.e., mispronounced by more than one third of adult native speakers of the respective target language) were eliminated. See Table 16 for an overview of the final selection of NWs.



**Table 16.** Overview of the final selection of NWs including their average Target Language Allocation, Pronounceability, and Repetition Accuracy scores. Target Language Allocation and Pronounceability scores reported here were collected during the online rating procedure ( $n = 27$  German speakers;  $n = 27$  Italian speakers). Repetition Accuracy scores stem from the direct ratings ( $n = 5/7$  German speakers;  $n = 5/7$  Italian speakers). LS items' scores are based on ratings obtained from monolingual speakers of the respective target language (i.e., either German or Italian), whereas LU items' scores were computed by averaging across both groups of participants, German and Italian adult native speakers.

NW subcategory	Number of syllables	Nonword	Target Language Allocation	Pronounceability	Repetition Accuracy
LSger	2	[ˈhʊxtʰəl]	96.30%	4.15	100%
		[ˈnɛ:ɐ̯lax]	88.90%	4.07	100%
		[pʰa:'mɔltʰ]	85.20%	4.37	80%
		[ˈtʰrɔ̯xtʰəl]	96.30%	4.26	80%
	3	[fɛ:g'jaʊxtʰən]	100.00%	4.33	100%
		[ra:və'laɪ]	77.80%	3.78	80%
		[re:so:'lantʰ]	100.00%	4.59	100%
	4	[ramlo:tə'ristʰ]	96.30%	3.67	100%
		[tʰɔlməfo:'kans]	81.50%	3.56	100%
		[ru:və'tʰɔŋən]	77.80%	3.56	100%
LSit	2	[ˈfjɛto]	92.60%	4.70	100%
		[ˈdɛpjo]	96.30%	4.70	100%
		[ˈzbalta]	96.30%	4.63	80%
		[ˈzbe:fo]	96.30%	4.59	100%
	3	[ˈbwɔmali]	88.90%	4.33	100%
		[fɔl'da:na]	96.30%	4.56	100%
		[man'tjo:ro]	88.90%	4.37	100%
	4	[biɛla'na:re]	100.00%	4.52	100%
		[fiŋka'mo:ro]	96.30%	4.33	100%
		[stelo'casto]	100.00%	4.22	100%
LU	2	/lefum/	38.85%	4.26	90%
		/maful/	40.75%	4.40	100%
		/mefal/	27.80%	3.87	100%
		/pimal/	50.00%	4.30	80%
	3	/famelep/	51.85%	4.17	70%
		/fulsamit/	42.60%	3.82	100%
		/melinak/	38.90%	4.29	100%
		/nisala/	55.60%	4.22	80%
	4	/minalefe/	59.30%	4.01	70%

/nufalemik/                      38.90%                      3.84                      70%

Finally, in order to further examine the distinctness between the final selection of LSger and LSit NWs and to verify whether the two categories were, in fact, perceived as language-specific, I compared adult German and adult Italian speaker’s Target Language Allocation. While German speakers allocated LSger NWs in more than 90% of cases to German, none of the LSit NWs were allocated to German. The same pattern was found for Italian speakers Target Language Allocation (see Table 17).

**Table 17.** Adult native speaker’ Target Language Allocation and Pronounceability ratings of LSit and LSger NWs obtained in the online rating procedure. Target Language Allocation refers to the percentage of monolingual speakers of either German or Italian who allocated the NWs to their respective L1.

	L1 German raters <i>n</i> = 27		L1 Italian raters <i>n</i> = 27	
	LSger NWs <i>n</i> = 10	LSit NWs <i>n</i> = 10	LSger NWs <i>n</i> = 10	LSit NWs <i>n</i> = 10
Target Language Allocation (%)	<i>M</i> = 91.49, <i>SD</i> = 7.81	<i>M</i> = .00, <i>SD</i> = .00	<i>M</i> = .00, <i>SD</i> = .00	<i>M</i> = 95.19, <i>SD</i> = 3.92

This final selection of NW recordings was then implemented in the computerized screening platform MuLiMi (Eikerling & Lorusso, 2021). During the assessment, a change in a picture of a space scenery displayed on a computer screen automatically occurs after each NW in order to keep the children entertained and guide their attention onto the task (i.e., to encourage them to listen to and repeat the NWs). In the MuLiMi NWRT, children’s repetition accuracy is judged on the whole-word level. That means that a simple binary judging procedure is applied. Children’s repetition attempts are either scored as correct (and assigned a score of 1; if judged to fully match the target) or incorrect (and assigned a score of 0; if a child’s realization of a NW deviates from the target by one or more phonemes). Variations in children’s accent or in their language-specific phonetic realization of single phonemes (e.g., variations in the realization of the rhotic /r/ realized as either a voiced apical coronal trill [r] vs. an uvular fricative [ʁ] or a voiced uvular trill [R], and the feature VOT in plosives such as /p/, /b/, /t/, /d/, /k/, and /g/) were considered acceptable as long as the phoneme was clearly identifiable. Raw scores obtained thus reflect the number of NWs that were correctly repeated by each child (e.g., as in Chiat &

Polišenská, 2016). Finally, for each child, the percentage of overall repetition accuracy per NW subcategory is calculated.

**Cross-Linguistic Lexical Tasks (CLTs).** In order to evaluate children's lexical knowledge in both Italian and German, the receptive subtests of the CLTs (German version: Rinker & Gagarina, 2014; Italian version: Roch, 2014), which assess children's comprehension abilities of nouns and verbs in the respective language, were used. The CLTs are sought to provide an indication of children's vocabulary size; word comprehension is assessed through a picture choice task. The tasks were created within the framework of Cost Action IS0804 and are currently available in 30 languages (see <https://multilada.pl/en/projects/clt/> for an overview). The German and Italian tasks were developed according to the same principles and criteria using complexity or age of acquisition indices (Haman, Łuniewska, & Pomiechowska, 2015). The CLTs assess what monolingual TD children of a certain language are expected to understand/produce at age five. Both the Italian and the German version of the CLT are available as an App for the Apple operating system iOS and can be administered by examiners without any prior knowledge of the target language. In the receptive subtests of the CLTs, children are presented with an auditory stimulus (i.e., noun or verb depending on the respective subtest) and are asked to choose the corresponding picture out of four possible options presented on the tablet screen. A maximum raw score of 32 can be obtained per subtest. The administration of each subtest takes about ten to fifteen minutes.

**Language Background Questionnaire.** Based on information provided by children's main caregivers, for each child an individual input (i.e. in regard to the language(s) the child hears on a daily basis) and output (i.e. in regard to the language(s) the child speaks on a daily basis) score were calculated for each language. Calculations first estimate the average number of hours each child is exposed to/ actively speaks Italian during a regular week. Then, the number of hours was converted into a percentage in order to express the ratio between children's weekly language input and output in Italian vs. German. Values obtained reflect children's current language experience at the time of their participation in this study (see Section 2.4.3 for a more detailed description of children's relative input and output score calculations)

Based on these parental reports, an additional variable was created expressing linguistic dominance. This compound score of language dominance in which children's input and output scores in both languages are merged was created based on the following equation:

$$\text{Language Dominance} = \frac{((\text{InputIT} - \text{InputGER}) + (\text{OutputIT} - \text{OutputGER}))}{2}$$

This new variable ranged from -1 (German-dominant) to 1 (Italian-dominant) and was used to group children according to their dominance status. Children with a score of -1 to -0.16 were considered German-dominant; -0.15 to 0.15 balanced; and .16 to 1 Italian-dominant.

**Colored Progressive Matrices (CPM).** See section 2.4.2 for a detailed description of the CPM.

#### 4.6.3 Procedure

Each child was tested individually in a quiet room in one of the kindergartens where they were tested by well-trained researchers. Only in the few cases, when this was not possible due to Covid-19 contact restrictions, children were tested in quiet rooms in their homes. The MuLiMi NWRT was carried out on a Lenovo laptop, model YOGA 720-15IKB under the Windows 10 Pro operating system. The online screening platform MuLiMi was accessed via the Mozilla Firefox web browser. Children's responses were scored by a qualified, bilingual Italian-German speaking SLP. In order to verify these judgements, following the procedure described by Bajpai, Bajpai, and Chaturvedi (2015), inter-rater reliability (IRR) was calculated from two further examiners (one native German speaker and one native Italian speaker) who independently scored children's NW repetition attempts. For IRR (based on three raters), Cronbach's Alpha was  $\alpha > .70$  for all NWs ( $M = .88$ ,  $SD = .07$ ). The CLTs were administered using an Apple Ipad 7 but scored manually on the parallel pen-and-paper protocol sheet by the examiner.

#### 4.6.4 Data Analysis

Data were analyzed by means of IBM SPSS Statistics v.26. Considering the relatively small sample size, nonparametric Kruskal Wallis H and Mann-Whitney U tests for independent samples were run to evaluate group differences (i.e., German-dominant vs. balanced vs. Italian-dominant) in children's NW repetition performance. Furthermore, nonparametric Friedman-tests were run to assess within group differences. In order to investigate the association between the various NW properties and children's repetition performance, parametric correlations were run. Further, the association between children's NW repetition performance and their linguistic background was assessed through partial correlation analyses, controlling for age (in months) and nonverbal intelligence (i.e., CPM scores).

## 4.8 Results

### 4.8.1 Description of the Sample

Table 18 provides an overview of all participants according to their dominance status (German-dominant vs. balanced vs. Italian-dominant). Across the three groups, children's age did not differ significantly (Kruskal-Wallis,  $p > .05$ ) neither did their performance on the CPM (Kruskal-Wallis,  $p > .05$ ). Further Kruskal-Wallis H tests showed a statistically significant difference across the three groups (German-dominant vs. balanced vs. Italian-dominant) concerning their receptive German and Italian vocabularies (with the exception of German nouns).

**Table 18.** Descriptive Statistics of the sample grouped in accordance with their dominance status.

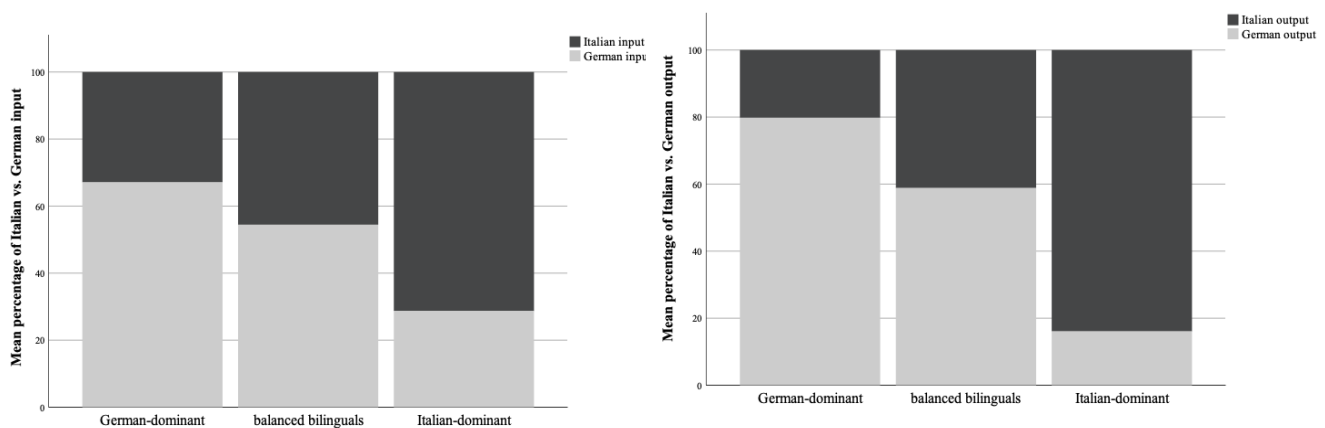
\*Note that for the Italian CLTs (both the Nouns and Verbs subtest)  $n = 13$  due to one participant failing to complete the entire test protocol

	German-dominant $n = 14^*$	Balanced bilinguals $n = 7$	Italian-dominant $n = 5$	Kruskal-Wallis H Statistics
Age (in months)	$M = 61.85,$ $SD = 8.50$	$M = 64.14,$ $SD = 9.51$	$M = 63.20,$ $SD = 9.60$	$H(2) = .720,$ $p > .05$
CPM (raw scores)	$M = 14.69,$ $SD = 2.25$	$M = 16.29,$ $SD = 6.05$	$M = 18.80,$ $SD = 3.90$	$H(2) = 4.542,$ $p > .05$
CLT Italian Nouns (%)	$M = 74.69,$ $SD = 17.76$	$M = 94.47,$ $SD = 5.50$	$M = 97.42,$ $SD = 4.21$	$H(2) = 13.051,$ $p < .05$
CLT Italian Verbs (%)	$M = 49.04,$ $SD = 20.39$	$M = 73.21,$ $SD = 16.32$	$M = 88.75,$ $SD = 7.19$	$H(2) = 12.355,$ $p < .05$
CLT German Nouns (%)	$M = 93.99,$ $SD = 7.16$	$M = 98.21,$ $SD = 1.67$	$M = 70.00,$ $SD = 38.06$	$H(2) = 3.442,$ $p > .05$
CLT German Verbs (%)	$M = 82.69,$ $SD = 14.80$	$M = 83.48,$ $SD = 13.94$	$M = 55.00,$ $SD = 19.71$	$H(2) = 6.997,$ $p < .05$

Mann-Whitney-U-Tests confirmed significant differences for German-dominant versus balanced children in the two Italian subtests (CLT Italian Nouns:  $U = 10.50, Z = -2.791, p = .003$ ; CLT Italian Verbs:  $U = 13.50, Z = -2.546, p = .008$ ) but not in the German subtests ( $p_s > .05$ ), whereas when comparing the Italian-dominant and balanced children, significant differences were found only for the German Verb subtest ( $U = 3.50, Z = -2.282, p = .018$ ). Finally, when comparing German-dominant and Italian-dominant children, significant differences emerged for all, except the German Noun subtest (CLT Italian Nouns:  $U = 3.00, Z$

= -2.920,  $p = .002$ ; CLT Italian Verbs:  $U = 3.50$ ,  $Z = -2.869$ ,  $p = .002$ ; CLT German Verbs:  $U = 8.50$ ,  $Z = -2.461$ ,  $p = .010$ ).

Furthermore, measures of current language input (how often children heard a language) and output (how often children used a language) assessed through parental reports were significantly different across groups (language input:  $H(2) = 12.436$ ,  $p < .05$ ; language output:  $H(2) = 14.248$ ,  $p < .05$ ; see Figure 18). German-dominant children on average received 32.92% of their language input in Italian, balanced bilinguals 45.57%, and Italian-dominant children 71.29%. A similar pattern was found for children’s use of their two languages: German-dominant children on average produced 20.21% of their utterances in Italian, balanced bilinguals 14.63%, and Italian-dominant children 83.94%.



**Figure 18.** Children’s relative amount of current Italian versus German input and output according to their dominance status (German-dominant vs. balanced vs. Italian-dominant)

These measures of language input and output were found to correlate with children’s receptive vocabulary knowledge in Italian and German respectively. More specifically, there was a strong and significant positive correlation between Italian input and children’s Italian noun ( $r = .965$ ,  $p < .001$ ) and verb ( $r = .848$ ,  $p < .001$ ) knowledge. Similarly, children’s Italian output was strongly and significantly correlated with their lexical knowledge (for Italian nouns:  $r = .647$ ,  $p = .001$ ; for Italian verbs:  $r = .825$ ,  $p < .001$ ). Further, children’s Italian input was significantly and negatively correlated with their German verb knowledge ( $r = -.480$ ,  $p < .05$ ) and marginally significant for German noun knowledge ( $r = -.386$ ,  $p = .057$ ). Significant negative correlations were also found between children’s Italian output and their performance on both German CLT subtests (for German nouns:  $r = -.513$ ,  $p < .05$ ; for German verbs:  $r = -.635$ ,  $p = .001$ ). The same, but reverse, pattern was observed for children’s language input and output in German.

#### 4.8.2 Children’s Nonword Repetition Performance

Independent of NW subcategory and children’s dominance status, there was a significant correlation between the number of syllables and children’s NW repetition accuracy ( $r = -.617$ ,  $p < .001$ ). Looking more closely at the association between children’s dominance status (German-dominant vs. balanced vs. Italian-dominant) and their repetition accuracy on the three different NW subcategories (LSger vs. LU vs. LSit), it was found that the number of syllables was significantly correlated with children’s repetition accuracy only in their non-dominant language (for German-dominant children and LSit NWs:  $\rho = -.776$ ,  $p = .008$ ; for Italian-dominant children and LSger NWs:  $\rho = -.686$ ,  $p = .029$ ). By contrast, the presence and number of consonant clusters within a NW was not significantly correlated with children’s repetition accuracy ( $p > .05$ ). Neither were adult Repetition Accuracy nor Pronounceability ratings (obtained in the direct ratings and online ratings respectively;  $p_s > .05$ ).

Concerning the question whether there was an association between children’s language experience (i.e., measures of language input and output) and their performance on LS vs. LU items, no significant correlations were found (all  $p_s > .05$ ). Furthermore, when grouping the children according to their dominance status (German-dominant vs. balanced vs. Italian-dominant) and comparing their NW repetition accuracy across the different MuLiMi NWRT subcategories (LSger vs. LSit vs. LU) no significant differences were found neither between nor within the three groups (see Table 19).

**Table 19.** Overview of children’s NW repetition accuracy according to their dominance status (German-dominant vs. balanced vs. Italian-dominant) across the different MuLiMi NWRT subcategories.

	German-dominant $n = 14$	Balanced $n = 7$	Italian-dominant $n = 5$	Kruskal-Wallis H Statistics
MuLiMi NWRT LSger (%)	$M = 60.80$ , $SD = 19.77$	$M = 55.70$ , $SD = 16.18$	$M = 50.00$ , $SD = 22.36$	$H(2) = 1.616$ , $p > .05$
MuLiMi NWRT LSit (%)	$M = 60.00$ , $SD = 10.00$	$M = 56.51$ , $SD = 9.44$	$M = 50.00$ , $SD = 20.00$	$H(2) = 1.128$ , $p > .05$
MuLiMi NWRT LU (%)	$M = 66.15$ , $SD = 15.57$	$M = 64.29$ , $SD = 17.18$	$M = 62.50$ , $SD = 10.90$	$H(2) = .308$ , $p > .05$
Friedman-Test Statistics	$\chi^2(2) = 2.085$ $p > .05$	$\chi^2(2) = 3.217$ $p > .05$	$\chi^2(2) = .500$ $p > .05$	

There was a significant correlation between children’s receptive German noun knowledge and their repetition accuracy of LSger NWs ( $r = .442$ ,  $p < .05$ ) that remained significant also when controlling for children’s age and their nonverbal Intelligence (i.e., CPM scores), whereas no

significant association was found between children's receptive Italian noun knowledge and their repetition accuracy of LSit NWs ( $p > .05$ ). Further, no significant associations between children's German or Italian verb knowledge and their performance on the respective LS NW subcategory were found ( $p_s > .05$ ).

#### 4.9 Discussion

This chapter was concerned with the effects of children's bilingual language experience and their performance on language-specific versus unspecific NW repetition. Generally, NWRTs have been identified to be valuable tools in the identification of DLD in bilingual children as they do not require extensive knowledge of the structures and regularities of a language and thus do not put children with limited experience with a given language at risk for receiving a misdiagnosis of DLD. Nevertheless, it has been suggested that even NW repetition is not completely free from the influence of language experience (Chiat, 2015). This study contributes to this discussion by examining the children's repetition accuracy on a novel Italian-German NWRT (the MuLiMi NWRT) as a function of their current language experience (i.e., language input and output). The NWs for this MuLiMi NWRT were selected as to be maximally (in the case LS NWs) versus minimally (in the case of LU NWs) language-specific (as assessed by adult native speakers of both Italian and German; cf. Ahufinger et al., 2021). Summarizing the results of the study with respect to the research questions that were initially raised, it can be concluded that children's repetition accuracy in the MuLiMi NWRT did not differ across NW subcategories (LSger vs. LU vs. LSit) according to their dominance status (German-dominant vs. balanced vs. Italian-dominant; RQ1). Furthermore, children's NW repetition performance was neither associated with measures of children's current language input (RQ2) nor language output (RQ3). Interestingly, while there was no association between children's receptive vocabulary knowledge in Italian and their repetition accuracy of LSit NWs, we found a significant correlation between children's receptive German noun knowledge and their performance on the LSger NWs (RQ4). These findings and their implications for bilingual children with DLD are discussed in more detail below.

##### 4.9.1 Nonword Characteristics and Children's Repetition Accuracy

Similar to what has been found in previous studies (e.g., Bortolini, Arfe, Degasperi, Deevy, & Leonard, 2006; Dispaldro, Leonard, & Deevy, 2013) children's repetition accuracy generally decreased with increasing NW length. Interestingly though, when examining this effect as a function of children's dominance status coupled with NW specificity, it was found that it



remained significant only for NWs in children's non-dominant language which suggests an interaction between phonological memory demands and familiarity with the stimuli. Although not significant, children achieved higher repetition accuracy for LU compared to LS NWs. Remember that LS NWs differed from LU NWs in two respects. First, while LS NWs contained language-specific consonant clusters, this feature was not included in the construction of the MuLiMi LU items. Arguably, LS NWs could therefore be considered more complex than LU NWs (cf. dos Santos & Ferré, 2016). However, no association was found between NW complexity (i.e., the presence and number of consonant clusters within a NW) and children's repetition accuracy. Second, while LS NWs were presented with the prosodic features of the respective target language, LU NWs were presented with flat (i.e., equal) syllable stress (cf. Chiat, 2015). Therefore, I would expect that the difference in repetition accuracy between LS and LU NWs to be more pronounced in bilingual children with DLD. First, due to the increase in articulatory demands associated with NW complexity (cf. Jones et al., 2010; dos Santos & Ferré, 2016) in LS over LU NWs. Second, because children with DLD may benefit less from familiarity with phonological word structure (Windsor et al., 2010; Kohnert et al., 2006) and prosodic features (Archibald & Gathercole, 2007) for the storage and repetition of (speech) sound sequences.

#### 4.9.2 Language Experience and Children's Repetition Accuracy

These findings add to the theoretically driven discussion regarding the usefulness of LS versus LU NWs for the identification of DLD in the bilingual population. Since in this study, children's accuracy of NWs, that were considered maximally language-specific by native speakers of the respective target language, did not differ according to children's dominance status (German-dominant vs. balanced vs. Italian-dominant), it might be concluded (1) that the role of language experience in NW repetition is limited even in bilingual children or (2) adult ratings of the NWs' language specificity (a judgment task based on the perceived characteristics of the items) and child repetition scores (based on the accordance between the target nonword and children's realization that is likely to be influenced by the characteristics of the NWs) are not directly related. In other terms, language specificity does not directly relate to accuracy of pronunciation/repetition of the NW. Furthermore, the absence of any influence of language experience on children's NW repetition accuracy may suggest that the sample of bilingual children included in this study had already acquired sufficient experience with both Italian and German in order to perform equally well on the German- and Italian-specific items (cf. Thordardottir & Brandeker, 2013) and that thus potential effects of different levels of language

experience had been minimized in this study. Remember that studies that had reported effects of language experience on children's NW repetition performance were usually conducted with groups of sequential bilinguals (e.g., Gutiérrez-Clellen & Simon-Cereijido, 2010; Summers et al., 2010) or with groups of younger bilingual children (e.g. Core et al., 2017; Parra et al., 2012) suggesting that language experience may have more impact on NW repetition performance during the early stages of language exposure and that once a certain threshold of experience has been reached, its influence diminishes. The current data seem to support this claim for TD children.

However, it has been proposed that children with DLD may have an underlying statistical learning deficit, meaning that they are less sensitive to (statistical) regularities in their (linguistic) input (Lammertink, Boersma, Wijnen, & Rispens, 2017). Effects of language experience and the relative amount of language input may therefore have a greater impact in language impaired children.

Finally, there was a significant correlation between children's receptive German noun knowledge and their repetition accuracy of German-specific NWs, whereas this pattern was not found for Italian-specific NWs. This finding suggests that German-specific NWs may have been more word-like than their Italian counterparts (Gathercole, 1995), a result that should be kept in mind especially when extending the study to children with DLD (Jones et al., 2010).

#### 4.9.1 Limitations and Further Directions

The main limitation of the study, due to limited access to kindergartens and bilingual families during COVID-19-related restrictions, is a relatively small sample size and unequal numbers of children in the three dominance-status groups (German-dominant vs. balanced vs. Italian-dominant). Moreover, the children included in this study do not sufficiently represent the great heterogeneity of language experience scenarios in the bilingual population. Participants were predominantly simultaneous or early sequential bilinguals who most likely had already had sufficient exposure in both languages to perform equally well on German- and Italian-specific NWs. This might be due to the recruitment method with the help of (although not exclusively through) bilingual kindergartens where children received dual language input. It was therefore not possible to examine how later onset of exposure and acquisition of a second language would influence performance on LS versus LU NWs. These issues should be addressed in future studies where a greater heterogeneity of (bilingual) acquisition patterns is represented in the sample, including simultaneous, early as well as later sequential bilinguals. It would also be interesting to include monolingual participants in order to examine how repetition accuracy is

affected by the presentation of NWs with characteristics that violate the phonotactic constraints of their native language.

Binary, whole-word NW scoring (correct vs. incorrect) has been found to be sufficient in discriminating DLD from TD children (Chiat & Polišenská, 2016) but in how far in-depth analyses on the syllable- and/or phoneme-level as well as individual error analyses could be useful to the theoretical discussion concerning the impact of language experience on NW repetition accuracy will need to be addressed in future studies. This becomes increasingly relevant in light of the fact that especially structural simplifications have been identified as good indicators for low versus average and high language competence with less competent talkers showing a higher rate of weak syllable and/or consonant/vowel deletions (Zanobini, Viterbori, & Saraceno, 2012).

Finally, in order to address the question whether diagnostic accuracy can be increased when using LS versus LU NWs, this study needs to be replicated with a sample of language impaired bilingual children.

#### 4.10 Conclusions

The current findings confirm that it is valuable to take the concept of language- specificity (in this study assessed through adult native speaking raters as “L1-Alikeness” and “Target Language Allocation”) into account in the creation of a NWRT that is supposed to be suitable for bilingual children. LU NWs could be especially useful in the assessment of children with minority L1s for whom LS items (and related norms) may be very difficult to obtain. However, complete unspecificity in linguistic stimuli is very difficult to reach and even approach. LU items are more difficult to construct and record, yielding lower repetition accuracy by native speakers, lower pronounceability ratings and more heterogeneous specificity. It is thus challenging but crucial to create comparable sets of LS NWs for particular language combinations.

Despite the independence of simultaneous bilingual children’s repetition accuracy and their dominance status (German-dominant vs. balanced vs. Italian-dominant) and measures of their language input and output, I argue in favor of a combination of LS and LU NWs such as the ones used in the MuLiMi NWRT in order to provide a satisfactory and broadly applicable solution for the identification of DLD especially in light of the great variety of language experience patterns present in the bilingual population (that unfortunately could not be represented in this study).

## 5. General Discussion

One of the overarching goals of this dissertation was to examine the language input situation of typically developing simultaneous bilingual Italian-German children growing up in Germany as heritage speakers of Italian and to relate their language experience to indices of language attunement and phonological development (cf. **chapter 2** of this dissertation). Although simultaneous bilingual children acquire their two languages from birth, the individual circumstances under which these children become bilingual can differ substantially (Carroll, 2017). In Germany, the heritage language Italian can be at risk as compared to Turkish for instance, only a minority of Italian parents seem to be interested in passing on their heritage language to their children (third generation: 39%; compared to Turkish: 82%; Fick et al., 2014). The willingness of parents to speak the heritage language Italian has consequences for the development of Italian in their children as the success of heritage language acquisition is influenced by factors such as the quantity and the quality of language input within the families' homes (Caloi & Torregrossa, 2021). Almost all children (with the exception of one), who participated in the studies presented in **chapter 2** and **chapter 4** respectively, were recruited from a pool of bilinguals with presumably ideal prerequisites for balanced bilingual language development, namely bilingual Italian-German-speaking children attending bilingual Italian-German kindergartens, which were intended to assure a certain extent of Italian language input not only within but also outside of the children's families' homes. However, despite the presence of Italian input in the entire sample of bilingual children (although to widely varying degrees) not all of them showed similar levels of active Italian use, with one child not producing any Italian output at all as assessed through a parental questionnaire. This is in line with De Houwer (2002) suggesting that not all bilingual exposure results in active bilingualism and that growing up in a two-language-context does not necessarily assure bilingualism. Also because children often eradicate the use of one of their two languages once social settings shift which can lead to language attrition (cf. Kaltsa, Maria, & Rothman, 2015) even in relatively young children. Nevertheless, these individual differences in the quantity of heritage language exposure within this group of simultaneous bilingual preschoolers provide an adequate starting point for the examination of the influence of language input on bilingual children's (typical) dual language acquisition.

Understanding the impact of children's language input on language acquisition is of particular interest for the adequate identification of DLD in the bilingual population. Relating the relative quantity of children's language input to their performance in the respective language has been deemed crucial in order to disentangle the effects of limited exposure from those of

actual language impairment (Armon-Lotem et al., 2015). This dissertation thus further aimed to assess the current state of how Speech and Language Pathologists in Germany approach bilingualism in their everyday clinical practice and whether and how they incorporate information and measures of children's language experience in the diagnostic process (cf. **chapter 3** of this dissertation). The results of the questionnaire study indicate that despite Speech and Language Pathologists' knowledge of the importance and impact of children's language experience on their language outcomes, there is no uniform way to account for differences in language exposure in the assessment of DLD in the bilingual population.

To approach this diagnostic dilemma, several scholars have proposed the use of nonword repetition tasks for the identification of DLD in bilingual children as they are considered relatively free or at least freer from the influence of language experience compared to other measures of language performance (Schwob et al., 2021). However, it has been argued that even children's performance on nonword repetition tasks can be subjected to the influence of their individual language experience depending on the nonwords' particular characteristics (Chiat, 2015), although results from previous studies are far from conclusive. Therefore, **chapter 4** of this dissertation has focused on the effect of bilingual Italian-German children's language experience on their performance in a novel nonword repetition task that included items that differed in terms of their perceived language specificity. Results indicated that independent of their pattern of language dominance, the effect of language experience on nonword repetition accuracy is negligible in simultaneous typically developing bilingual preschoolers, potentially because they have already accumulated sufficient experience with both languages (Thordardottir & Brandeker, 2013).

Collectively, it can be said that input quantity (and quality) influence the dual language acquisition process of simultaneous bilingual children and that they thus have to be taken into account when assessing bilingual children for language impairment. Especially the data presented in **chapter 2** highlight that simultaneous bilingual children develop their two languages independently (although crosslinguistic influence is to be expected) but that their two languages do not develop monolingual-like. Therefore, in the identification of DLD in the bilingual population, it is crucial to either use assessment tools that have been developed specifically for bilingual children or to apply bilingual norms to monolingual tests/procedures. In the following sections, I will discuss and link the main results of the chapters 2 to 4 before I finally proceed to elaborate how the present results lead to future research on bilingual language acquisition and most importantly their practical implications in order to advance and promote bilingually oriented practices in clinical settings.

## 5.1 Bilingual Children's Production and Perception of VOT

The primary research questions addressed in chapter 2 were (1) whether bilingual Italian-German-speaking children developed automaticity in speech sound perception in their two languages similarly to their monolingual peers and (2) whether they showed monolingual-like performance of the equivalent phonemes in speech sound production.

The VOT production experiment in chapter 2 showed that in the societal language German, with regards to the long lag-short lag VOT distinction, the bilingual Italian-German children were indistinguishable from their monolingual German peers. However, in the heritage language Italian, they showed a different pattern with regards to the short lag-voicing lead distinction when compared to their monolingual Italian peers. Although even one monolingual Italian child did diverge from the rest of the group and did not show robust signs of voicing lead for Italian voiced word onsets. This finding is in line with previous research suggesting that even in monolingual language development voicing lead may be acquired relatively late (cf. Khatlab, 2002 for Arabic monolinguals; MacLeod, 2016 for French monolinguals). Therefore, it may be assumed that the acquisition of voicing lead poses particular challenges to bilingual children, especially in the heritage language, presuming that heritage speakers of a voicing lead language have fewer opportunities to (1) hear and (2) practice this complex velopharyngeal voicing feature. Nevertheless, chapter 2 showed that bilingual Italian-German children produced language-specific VOT for both of their languages although their VOT productions in Italian were not monolingual-like. Moreover, chapter 2 revealed that children's relative amount of language input in both the societal and the heritage language were associated with target-like voiced VOT (i.e., short lag in German and voicing lead in Italian) but no association was observed between children's input quantity and their production of voiceless stops in the respective language (i.e., long lag in German and short lag in Italian). Similarly, a global measure of parental language proficiency was associated with children's productions of language-specific target-like VOT in both languages. Although not measured directly this finding suggests that the model parents provide for their children influences their speech production (cf. Stoehr, 2018). Furthermore, children's amount of active Italian usage (i.e., the relative amount of Italian output) was significantly associated with their target-like production of Italian voicing lead, which is in line with the theoretical framework of the A-Map model (McAllister Byun et al., 2016), suggesting that children who have increased opportunities to use and thus practice motoric routines of their heritage language show monolingual-(adult)-like productions earlier than their peers with fewer opportunities to actively use their heritage language Italian.

With respect to children's automatic perception of the two language-specific VOT contrasts findings in regard to bilingual children's processing of the German-like Long Lag contrast were surprising, especially in the light of their reliable monolingual-like productions of German long lag. While bilingual participants reliably neurally discriminated between the short lag standard and the German-like long lag deviant in the easy EEG paradigm (i.e., with a greater acoustic difference between the two stimuli), they did not automatically detect the difference between the short lag standard and the German-like long lag deviant in the difficult EEG paradigm (i.e., with a more subtle acoustic difference between the two stimuli). By contrast, monolingual German preschoolers showed robust neural signs of automatic, pre-attentive speech sound discrimination to both the easy and the difficult short lag-long lag distinction. Two particular explanations for this monolingual-bilingual difference were brought forward in chapter 2: One suggestion was that as a group, bilinguals were not (yet) automatic in discriminating German-like VOT. Although within the entire group of bilingual participants, iMMR amplitude was not significantly correlated with measures of their input quantity, visual inspection of sub-groups' brain responses showed that iMMR amplitude as well as polarity differed for high German versus high Italian input bilinguals, with the latter showing signs of more immature processing (as indexed by a small but positive peak within the MMR time window of interest; Shafer et al., 2010). Moreover, even the group of high German input bilingual children did not match the group of monolingual German controls in their iMMR amplitude. Taken together these results suggest that due to their reduced amount of German experience (compared to their monolingual peers) bilingual children had not (yet) established robust SPRs for the German voicing contrast. Alternatively, it has been proposed that bilinguals may not have shown robust signs of neural discrimination for the short lag-difficult long lag distinction because their perceptual boundary had shifted towards higher VOT values as they accept longer VOT values as instances of 'voiced' plosives in German (cf. Stoehr, 2018) and that subsequently they assimilated the two sounds into one category, suggesting that the bilingual children's voicing systems were still developing. This developmental shift towards adult-like perceptual boundaries has been observed in four-to-five-year-old monolingual children but not in a group of age-matched sequentially bilinguals (McCarthy, Mahon, Rosen, & Evans, 2014), suggesting that bilinguals may take longer to establish these adult-like boundaries due to their reduced language experience. Finally, the result that Italian-German bilinguals and German monolinguals did not differ with respect to their perception of Italian-like voicing lead (both in the easy as well in the difficult EEG paradigm) was unexpected. Both groups showed neural signs of automatic, pre-attentive speech sound discrimination. Again,

two alternative explanations were put forward for the lack of a bilingual advantage in chapter 2. First, it may show that bilinguals were not yet automatic in processing their heritage language-specific voicing contrast and that monolingual German children were similarly sensitive to the voicing lead feature as they may have come across instances of it even in their German input (Jessen, 1998). An alternative explanation was that conceptual language cues based on the linguistic German context during the EEG experiment may have overridden the perceptual cues provided in the VOT stimuli (Wig & García-Sierra, 2021). This suggests that during the EEG experiment their phonetic boundary did not flexibly shift between Italian and German (García-Sierra, Diehl, & Champlin, 2009) and that they were processing all stimuli in “German-mode”.

Taken together the findings of chapter 2 highlight the importance of combining speech production and speech perception methods to fully grasp the development of phonological systems in children who grow up bilingually and acquire their languages simultaneously. Future research should thus address the missing puzzle pieces of this study, by (1) incorporating a behavioral perception task equivalent to the stimuli used in the EEG paradigm to assess children’s individual perceptual boundaries, (2) evaluating the influence of linguistic context and conceptual cues on children’s processing abilities, and (3) including a second control group of monolingual Italian children to better interpret bilingual children’s processing abilities in their heritage language.

In chapter 2, I established that the developmental trajectory of perception and production of language-specific phonetic units in monolingual compared to simultaneous and early sequential bilingual children does not align. However, the acquisition of language-specific phonological systems is relevant for children’s future language development as phonetic units can be considered the most basic building blocks of language. The relevance for children to be able to distinguish between VOT categories becomes evident in the light of so-called minimal pairs (i.e., a pair of words that only differ in one phonological element and have distinctive meaning; e.g., big-pig in English). For example, /b/ in English is phonetically realized and perceived as /p/ in speakers of Romance languages (e.g., Italian), and this different pattern in VOT for bilinguals is expected to have a significant impact on the acquisition of words, as well as reading/writing skills (e.g., English “big” versus “pig” where Romance-language listeners will perceive English /b/ as [p] and the orthographic label “p”).



## 5.2 Assessing Bilingual Language Development in Clinical Practice

Already in 1989, Grosjean postulated that “the bilingual is not two monolinguals in one person”, which was also reflected in the results obtained in chapter 2 of this dissertation. In general, the notion that monolingual-like linguistic skills are not the ultimate goal in bilingual language acquisition has nowadays been widely accepted in bilingualism research due to the multifactorial ways in which a person’s language environment and their language proficiency interact (e.g., Rothman, 2007, for heritage language acquisition). It is therefore only logical to argue that in clinical practice, bilingual children’s language performance should not be compared to monolingual language attainment when screening for Developmental Language Disorder. Both, the characteristics of the languages to which they are exposed and the fact that they are bilingual need to be acknowledged when assessing bilingual children’s language skills. Chapter 3 of this dissertation was concerned with how bilingualism is approached in Speech and Language Pathologists’ everyday practice in Germany. The results of the questionnaire study reported in chapter 3 indicate that bilingualism is a highly relevant issue in today’s every-day clinical practice in Germany. Almost all Speech and Language Pathologists who participated in the questionnaire study indicated to be regularly confronted with bilingual children on their caseloads (although to varying degrees). Overall, Speech and Language Pathologists have shown good theoretical knowledge about the relevance of applying bilingually oriented approaches in the assessment of children who acquire more than one language simultaneously or sequentially. Beyond that, experience in specifically working with bilingual children was found to be most influential to clinicians’ approaches to bilingualism and their perceived level of competence when working with the bilingual population. The vast majority of participants was found to be aware of at least some diagnostic materials specifically designed (or normed) for the bilingual population. However, only a small proportion of the sample indicated to actually use them in their every-day practice. Their reluctance to make use of readily available tools in the assessment of their bilingual patients leaves room for the speculation that they are not content with the existing materials. Further research will be needed in order to determine which factors keep clinicians from applying available tools and, moreover, what they would consider valuable features of diagnostic materials for the bilingual population. Overall, German Speech and Languages Pathologists were found to acknowledge the necessity to take their bilingual patients language experience into account when assessing their language skills (in the societal language). Nevertheless, the procedures they follow to measure and incorporate information about children’s individual language exposure patterns in the diagnostic process are highly variable. Similarly, in research, there is little consistency in how

input quantity is optimally measured (see Section 1.2 of this dissertation for more details). The majority of Speech and Language Pathologists rely on parents' estimates of their children's proportional input quantity and use this estimate as a rough reference point to consider whether they should adopt a wait-and-see approach or initiate therapeutic intervention immediately. Generally, German Speech and Language Pathologists appear to be more prone to over-diagnosing DLD in bilingual children than to risk under-identification. Seemingly, the consequences of an over-diagnosis are considered less severe than mistakenly attesting a bilingual child's deviations from (monolingual) language development to their bilingual language experience and unreasonably withhold this child from the support they would need. Although they acknowledge that over-identification of DLD would unnecessarily occupy resources of the therapists themselves, the children and their families, and also the health system.

### 5.3 Approaching the Diagnostic Dilemma

In order to approach the diagnostic dilemma that evolves around the inconsistencies in systematically measuring bilingual children's language experience and relating it to their language performance, it has been proposed that Nonword Repetition Tasks may constitute promising screening tools that depending on the particular characteristics of the nonwords are relatively free from the influence of children's language background (Chiat, 2015). Chapter 4 of this dissertation was thus concerned with how bilingual typically developing children's nonwords repetition accuracy relates to the nonwords' degree of language specificity. The results were straight forward. There was no link between children's repetition performance and nonword category (Italian-specific, German-specific, and language-unspecific respectively) on the novel MuLiMi Nonword Repetition Task. Italian-dominant, German-dominant, and more balanced bilinguals performed equally well across all categories, suggesting that Nonword Repetition Tasks are in fact useful tools for the identification of Developmental Language Disorder in the bilingual population. Nevertheless, it is important to highlight that the children participating in the study presented in chapter 4 had all been exposed to German before the age of 3 years and had been living in Germany for at least two years at the time of their participation. This could have been sufficient to eliminate any effects of language experience (cf. Thordardottir & Brandeker, 2013). Nevertheless, the results presented in chapter 4 provide a cautiously optimistic forecast for the use of existing language-specific Nonword Repetition Tasks with bilingual children. Although it remains an unanswered question whether these tasks would need to provide normed data specifically for bilingual children (note that there was no

monolingual control group tested on the same nonwords in chapter 4). Furthermore, but hypothetically, sequential bilinguals who have just immigrated to Germany and who have never before been exposed to German-like phonotactics may have shown different patterns of repetition accuracy than the group of simultaneous and early sequential bilinguals who participated in this study. A Nonword Repetition Tasks that comprises language-specific as well as unspecific nonwords could be most suitable for the average population of bilingual children as it would account for the great heterogeneity in children's language backgrounds. Finally, and most importantly, the study needs to be replicated with a sample of children with Developmental Language Disorder.

#### 5.4 Practical Implications

The effect of dual language exposure and different experience scenarios on bilingual children's language development is undeniable and so are the barriers to adequately assessing the language skills of bilingual children. The hurdles that SLPs are confronted with when trying to properly diagnose bilingual children with DLD are multifactorial in nature and the limited access to several key supports and resources for overcoming some of these barriers (i.e., easy access to interpreters or assessment tools in languages beyond the societal language) might explain why the vast majority of therapists associates the assessment of bilingual children with greater efforts than the assessment of monolingual children (Friedrich & Knebel, 2017), which further contributes to maintaining the current predominance of monolingual policies and practices in Speech and Language Pathology. Participants' questionnaire responses were indicative of more preparation time being needed when working with bilingual children in order to provide equal-quality services for bilingual and monolingual children. This should be acknowledged and financially compensated by the health care providers. Facilitated access to Speech and Language Pathology training for second-language speakers of the societal language, minority language classes as part of the curriculum, or re-training for people with migration backgrounds who already have a professional background in Speech and Language Pathology-related areas could constitute a first step towards moving away from the predominance of monolingual and monocultural approaches in clinical practice. In research, over the past decade, a lot of efforts have been dedicated to developing adequate assessment tools for bilingual children and their adaptation for various languages including German (e.g., the LITMUS Cost Action IS0804 test-battery) and several studies using the various LITMUS tasks have report promising results (Armon-Lotem, de Jong, & Meir, 2015 for an overview). Unfortunately, the majority of (German) clinicians is not familiar with this test battery and therefore it seems that a major

hurdle in establishing bilingually-oriented routines in Speech and Language Pathology seems to be the transfer from research to clinical practice. Considering the need for further development of knowledge and of ad hoc instruments in the field of bilingualism and its impact on the identification of DLD, the involvement of clinicians in research projects should be encouraged and supported in order to bridge the gap between scientific research and applied practice and to further pave the way for more evidence-based practices in the field of Speech and Language Pathology. The finding that practical experience seems to be the most influential factor on Speech and Language Pathologists' attitudes and approaches towards bilingualism highlights the need to ensure that concrete experience with bilingual patient populations is firmly established in the training curriculum. Finally, adequate reference points for typical bilingual language development that take into account individual patterns of language exposure are necessary to permit appropriate language assessment and the identification of Developmental Language Disorder in bilingual children.

## References

- Abed Ibrahim, L., & Hamann, C. (2017). Bilingual Arabic-German and Turkish-German with and without Specific Language Impairment: Comparing Performance in Sentence and Nonword Repetition Tasks. In *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 1–17). Somerville, MA: Cascadilla Press.
- Ahlo, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*, 151–155.
- Ahmed, A. U., Clarke, E. M., & Adams, C. (2008). Mismatch negativity and frequency representational width in children with specific language impairment. *Developmental Medicine and Child Neurology*, *50*, 938–944.
- Ahufinger, N., Berglund-Barraza, A., Cruz-Santos, A., Ferinu, L., Andreu, L., Sanz-Torrent, M., & Evans, J. L. (2021). Consistency of a Nonword Repetition Task to Discriminate Children with and without Developmental Language Disorder in Catalan – Spanish and European Portuguese Speaking Children. *Children*, *8*(85), 1–30.
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). <https://doi.org/10.1176/appi.books.9780890425787>
- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical Basis for Speech Sound Discrimination. *Language and Speech*, *46*(2–3), 155–182.
- Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment : More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, *14*(5), 919–924.
- 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. H., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Multilingual matters. Croyden, UK: CPI Group (UK Ltd).
- Armon-Lotem, S. & de Jong, J. (2015). Introduction. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 1-22). Bristol, UK: Channel View Publications.
- Armon-Lotem, S., Joffe, S., Abutbul-Oz, H., Altman, C., & Walters, J. (2014). Language exposure, ethnolinguistic identity and attitudes in the acquisition of Hebrew as a second language among bilingual preschool children from Russian- and English-speaking

- backgrounds. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 77–98). Amsterdam: John Benjamins.
- Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders* 51(6), 715–731.
- Armon-Lotem, S. & Meir, N. (2019). The Nature of Exposure and Input in Early Bilingualism. In A. de Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 193-212). Cambridge, UK: Cambridge University Press.
- Arrhenius, B., Gyllenberg, D., Chudal, R., Lehti, V., Sucksdorff, M., Sourander, O., Virtanen, J.-P., Torsti, J., & Sourander, A. (2018). Social risk factors for speech, scholastic and coordination disorders: A nationwide register-based study. *BMC Public Health*, 18(1), 1-10.
- Bajpai, S., Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 20–27.
- Baroni, A. (2012). A Beats-and-Binding Account of Italian Phonotactics. *Padua Working Papers in Linguistics*, 5(3), 45–72.
- Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating What Infants Attune to in Native Speech. *Ecological Psychology*, 28(4), 216–261.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080.
- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Blumenthal, M., Scharff Rethfeldt, W., Salameh, E. K., Muller, C., Vandewalle, E., & Grech, H. (2015). Position Statement on language impairment in multilingual children. Available online [https://www.researchgate.net/profile/Wiebke-Scharff-Rethfeldt/publication/319184514\\_Position\\_Statement\\_on\\_Language\\_Impairment\\_in\\_Multilingual\\_Children/links/5999a61345851564432dbdd5/Position-Statement-on-Language-Impairment-in-Multilingual-Children.pdf](https://www.researchgate.net/profile/Wiebke-Scharff-Rethfeldt/publication/319184514_Position_Statement_on_Language_Impairment_in_Multilingual_Children/links/5999a61345851564432dbdd5/Position-Statement-on-Language-Impairment-in-Multilingual-Children.pdf) (accessed on February 14<sup>th</sup>, 2022).
- Blumenthal, M.; Scharff Rethfeldt, W.; Grech, H.; Letts, C.; Muller, C.; Salameh, E.-K.; Vandewalle, E. (2016). Position Statement on Language Impairment in Multilingual

- Children. In *Proceedings of the 30th World Congress of the International Association of Logopedics and Phoniatrics (IALP)*, Dublin, Ireland, 21–25 August 2016.
- 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, and Hearing Research*, 58(6), 1747–1760.
- Boerma, Tessel, & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, 65, 65–76.
- Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program].
- Boivin, M. J., Kakooza, A. M., Warf, B. C., Davidson, L. L., & Grigorenko, E. L. (2015). Reducing neurodevelopmental disorders and disability through research and interventions. *Nature*, 527, 155–160.
- Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, M., & Mari, R. (2020). Which Measures Better Discriminate Language Minority Bilingual Children With and Without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment. *Journal of Speech, Language, and Hearing Research*, 63, 1898–1915.
- Bortolini, U. (1995). *Prove per la valutazione fonologica del linguaggio infantile*. Padua, Italy: Edit Master.
- Bortolini, U., Arfe, B., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *International Journal of Communication Disorders*, 41(6), 695–712.
- Bosch, L. Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, 46(2–3), 217–243.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of Acoustical Society of America*, 101(4), 2299–2310.
- Brannon, E. M., Wolfe Roussel, L., Meck, W. H., & Woldorff, M. (2004). Timing in the baby brain. *Cognitive Brain Research*, 21, 227–233.
- Bulheller, S. & Häcker, H. (2001). *CPM Raven's Progressive Matrices and Vocabulary Scales - Coloured Progressive Matrices* adapted from Raven, J.C., Raven, J., & Court, J. H., (2003). Frankfurt am Main: Pearson.
- Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic

- representation in bilingual and monolingual infants. *Applied Psycholinguistics*, 28, 455–474.
- Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science*, 21(3), 343–348.
- Caloi, I., & Torregrossa, J. (2021). Home and School Language Practices and Their Effects on Heritage Language Acquisition: A View from Heritage Italians in Germany. *Languages*, 6(50), 1–20.
- Calvo, A., & Bialystok, E. (2014). Independent Effects of Bilingualism and Socioeconomic Status on Language Ability and Executive Functioning. *Cognition*, 130(3), 278–288.
- Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism*, 20(1), 3–16.
- Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., ... Burani, C. (2007). Non-Word Repetition In Children With Specific Language Impairment: A Deficit In Phonological Working Memory Or Long-Term Verbal Knowledge? *Cortex*, 43, 769–776.
- Caselli, M. C., Bello, A., Rinaldi, P., Stefanini, S., & Pasqualetti, P. (2015). Il Primo Vocabolario del Bambino: Gesti, Parole e Frasi. Valori di riferimento fra 8 e 36 mesi delle Forme complete e delle Forme brevi del questionario MacArthur-Bates CDI. Milan, Italy: Franco Angeli.
- 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 and Communication Disorders*, 49(6), 649–671.
- Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2013). The Development of Mismatch Responses to Mandarin Lexical Tones in Early Infancy. *Developmental Neuropsychology*, 38(5), 281–300.
- Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2015). Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: Mismatch responses to vowels and initial consonants. *International Journal of Psychophysiology*, 96(2), 84–94.
- Cheour, M., Ahlo, K., Čeponienė, R., Reinikainen, K., Sainio, K., Pohjavuori, M., ... Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, 29, 217–226.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain.



- Nature Neuroscience*, 1(5), 351–353.
- Cheour, M., Kushnerenko, E., Čeponienė, R., Fellman, V., & Näätänen, R. (2002). Electric Brain Responses Obtained from Newborn Infants to Changes in Duration in Complex Harmonic Tones. *Developmental Neuropsychology*, 22(2), 471–479.
- Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., & Näätänen, R. (2002). Mismatch negativity shows that 3-6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters*, 325, 187–190.
- 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. 125–150). Bristol: Multilingualism 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, 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.
- Chilla, S. (2014). Grundfragen der Diagnostik im Kontext von Mehrsprachigkeit und Synopse diagnostischer Verfahren. In S. Chilla & S. Haberzettl (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen: Mehrsprachigkeit* (pp. 57-72). Munich, Germany: Urban & Fischer Verlag.
- Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond : Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65.
- Core, C., Chaturvedi, S., & Martinez-Nadramia, D. (2017). The Role of Language Experience in Nonword Repetition Tasks in Young Bilingual Spanish-English Speaking Children. In *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 179–185).
- Crescentini, C., Marini, A., & Fabbro, F. (2012). Competence and Language Disorders in Multilinguals. *EL.LE Educazione Linguistica. Language Education*, 1, 81–98.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.
- Crowe, K., Cuervo, S., Guiberson, M., & Washington, K. N. (2021). A Systematic Review of Interventions for Multilingual Preschoolers With Speech and Language Difficulties. *Journal of Speech, Language and Hearing Research*, 1–26.
- Cutler, A., Norris, D., & Williams, J. N. (1987). A Note on the Role of Phonological

- Expectations in Speech Segmentation. *Journal of Memory and Language*, 26, 480–487.
- Cychoz, M., Erskine, M., Munson, B., & Edwards, J. (2021). A lexical advantage in four-year-old children's word repetition. *Journal of Child Language*, 48(1), 31–54.
- D'Souza, C., Kay-Raining Bird, E., & Deacon, H. (2012). Survey of Canadian Speech-Language Pathology Service Delivery to Linguistically Diverse Clients. *Canadian Journal of Speech-Language Pathology and Audiology*, 36(1), 18–39.
- Datta, H., Hestvik, A., Vidal, N., Tessel, C., Hisagi, M., Wróblewski, M., & Shafer, V. L. (2019). Automaticity of speech processing in early bilingual adults and children. *Bilingualism: Language and Cognition*, 23, 429–445.
- Datta, H., Shafer, V. L., Morr, M. L., Kurtzberg, D., & Schwartz, R. G. (2010). Electrophysiological Indices of Discrimination of Long-Duration, Phonetically Similar Vowels in Children With Typical and Atypical Language Development. *Journal of Speech, Language, and Hearing Research*, 53, 757–778.
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L., ... Barthez, M.-A. (2017). Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism*, 7(3-4).
- De Houwer, A. (1990). *The acquisition of two language from birth: A case study*. Cambridge: Cambridge University Press.
- De Houwer, A. (2002). Comparing monolingual and bilingual acquisition. *Alkalmazott Nyelvtudomány*, 2(1), 5–19.
- De Houwer, A. (2018). Input, context and early child bilingualism: implications for clinical practice. In A. Bar-On & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 599–616). Berlin: Walter de Gruyter.
- De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders*, 46(6), 613–627.
- Delmonte, R. (1981). L'accento di parola nella prosodia dell'enunciato dell'Italiano standard. *Studi di Grammatica Italiana Firenze*, 10, 351-394.
- Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, 24, 351–365.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- 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.
- Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-Word and Nonword Repetition in Italian-Speaking Children With Specific Language Impairment: A Study of Diagnostic Accuracy. *Journal of Speech, Language, and Hearing Research*, 56, 323–336.
- dos Santos, C., & Ferré, S. (2016). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, 25(1), 58–71.
- Ebert, K. D. (2014). The Role of Auditory Nonverbal Working Memory in Sentence Repetition for Bilingual Children with Primary Language Impairment. *International Journal of Language and Communication Disorders*, 49(5), 631–636.
- Eikerling, M. & Lorusso, M. L. (2021). Da guckste in die Röhre?! Pilotierung computer-gestützter Screenings für bilinguale Kinder mit der Screening-Plattform MuLiMi. *Forschung Sprache*, 2, 29-37.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306.
- Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, 62, 971–974.
- Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, 19(5), 529–537.
- Eurostat (2021). *Migration and Migrant Population Statistics*. Available online: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration\\_and\\_migrant\\_population\\_statistics#Migrant\\_population: 23 million non-EU citizens living in the EU on 1 January 2020](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration_and_migrant_population_statistics#Migrant_population:_23_million_non-EU_citizens_living_in_the_EU_on_1_January_2020) (accessed on September 20<sup>th</sup>, 2021).
- Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in Bilingual Phonological Acquisition: Evidence from Phonetic Inventories. *International Journal of Bilingual Education and Bilingualism*, 13(1), 1–19.
- Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics*, 26(2), 148–163.
- Farabolini, G., Rinaldi, P., Caselli, C. M., & Cristia, A. (2021). Non-word repetition in bilingual children : the role of language exposure , vocabulary scores and environmental factors. *Speech Language and Hearing*, 1–16.

- Fick, P., Wöhler, T., Diehl, C., & Hinz, T. (2014). Integration gelungen? Die fünf größten Zuwanderungsgruppen in Baden-Württemberg im Generationenvergleich. Konstanz: Universität Konstanz.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech, Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of Acoustical Society of America*, *97*(5), 3125–3134.
- Fox, A. V. (2000). *The Acquisition of Phonology and the Classification of Speech Disorders in German-Speaking Children*. University of Newcastle upon Tyne.
- Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Cognitive Neuroscience and Neuropsychology*, *13*(10), 51–54.
- Friedrich, M., Herold, B., Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex*, *45*, 662–676.
- Friedrich, S., & Knebel, U. Von. (2017). Sprachtherapie mit sukzessiv mehrsprachigen Kindern mit Sprachentwicklungsstörungen: Eine empirische Analyse gegenwärtiger Praxiskonzepte im Bundesland Berlin. *Forschung Sprache*, *1*, 57–77.
- Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*(4), 369–378.
- Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology*, *110*, 1–17.
- Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., ... Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546–557.
- Garraffa, M., Vender, M., Sorace, A., & Guasti, M. T. (2019). Is it possible to differentiate multilingual children and children with Developmental Language Disorder? <https://api.repository.cam.ac.uk/server/api/core/bitstreams/2cabf91b-6c5b-40b5-b640->

[4b5ddcca582c/content](#) (accessed on May 4<sup>th</sup>, 2023).

- 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.
- Gawlitzeck-Maiwald, I., & Tracy, R. (1996). Bilingual bootstrapping. *Linguistics*, 34, 901–926.
- Genesee, F. H. (2009). Early childhood bilingualism: Perils and possibilities. *Journal of Applied Research on Learning*, 2, 1–21.
- 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.
- Giraud, A.-L., & Poeppel, D. (2012). Speech Perception from a Neurophysiological Perspective. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Fay (Eds.), *The human auditory cortex* (pp. 225–260). New York, NY: Springer.
- Goldstein, B., & Swasey Washington, P. A. (2001). An Initial Investigation of Phonological Patterns in Typically Developing 4-Year-Old Spanish-English Bilingual Children. *Language, Speech, and Hearing Services in Schools*, 32, 153–164.
- Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*, 24(11), 2135–2142.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r.” *Neuropsychologia*, 9, 317–323.
- Grandpierre, V., Milloy, V., Sikora, L., Fitzpatrick, E., Thomas, R., & Potter, B. (2018). Barriers and facilitators to cultural competence in rehabilitation services: a scoping review, *BMC health services research*, 18(1), 1–14.
- Grimm, A., & Hübner, J. (2017). Nonword repetition by bilingual learners of German: The role of language-specific complexity. submitted to C. dos Santos & L. de Almeida (Eds.), *Bilingualism and Specific Language Impairment*. Amsterdam: Benjamins.
- Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, 7, 821–841.
- Grosjean, F. (1982). *Life with Two Languages: An Introduction to Bilingualism*. Harvard Univ. Press. Cambridge, MA, USA, 1982.
- Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual Is Not Two Monolinguals in One Person. *Brain and Language*, 36, 3–15.

- Grosjean, F. (2015). *Parler plusieurs langues: Le monde des bilingues*. Albin Michel.
- Grover, V., Shafer, V. L., Campanelli, L., Whalen, D. H., & Levy, E. S. (2021). Perception of American English Consonants /v / and / w / by Hindi Speakers of English. *Journal of Second Language Pronunciation*, 7(3), 370–407.
- Guiberson, M., & Rodríguez, B. L. (2015). Nonword repetition in Spanish-speaking toddlers with and without early language delays. *Folia Phoniatica et Logopaedica*, 67(5), 253–258.
- Gutiérrez-Clellen, V. F. & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, 24(2), 267-288.
- Gutiérrez-Clellen, V. F., & Simon-Cerejido, G. (2010). Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English Speaking Children: Does the Language of Assessment Matter? *Learning Disabilities Research & Practice*, 25(1), 48–58.
- Haman, E., Łuniewska M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for Bilingual Preschool Children. In J. de Jong, S. Armon-Lotem, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 196-240). Bristol: Multilingual Matters.
- Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.
- Hart, B., & Risley, T. R. (2003). The Early Catastrophe The 30 Million Word Gap by Age 3 Recorded vocabulary size Average. *American Educator*, 27, 4–9.
- Harvey, H., Allaway, H., & Jones, S. (2018). The effectiveness of therapies for dual language children with developmental language disorder: a systematic review of interventional studies. *International Journal of Bilingual Education and Bilingualism*, 24(7), 1–22.
- Hayes-Harb, R., & Barrios, S. (2022). Native English speakers and Hindi consonants: From cross-language perception patterns to pronunciation teaching. *Foreign Language Annals.*, 55, 175–197.
- Hemández-Chávez, E., Burt, M., & Dulay, H. (1978). Language dominance and proficiency testing: Some general considerations. *NABE journal*, 3(1), 41-54.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55–88.
- Hoff, E. (2020). Lessons from the study of input effects on bilingual development. *International Journal of Bilingualism*, 24(1), 82–88.
- Hoover, J.R. & Storkel, H.L. (2006). Using nonword repetition in vocabulary assessment.

- Acquiring Knowledge in Speech, Language, and Hearing*, 8, 106-109.
- Huer, M. B. & Saenz, T.I. (2003). Challenges and Strategies for Conducting Survey and Focus Group Research With Culturally Diverse Groups. *American Journal of Speech Language Pathology*, 12, 209–220.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk , processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), F31–F39.
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of Variability in Children’s Language Growth. *Cognitive Psychology*, 61(4), 343–365.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57.
- Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.
- Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-Year Follow-Up of Children With and Without Speech-Language Impairments: Family, Educational, Occupational, and Quality of Life Outcomes. *American Educator*, 19, 51–66.
- 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.
- Jong Kong, E., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40, 725–744.
- Jordaan, H. (2008). Clinical Intervention for Bilingual Children : An International Survey, 97–105.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. London: The MIT Press.
- Kaltsa, M., Maria, I., & Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua*, 164, 266–288.
- Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2021). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition*, 25, 29–41.
- Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish

- children Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, 7(1), 71–88.
- Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. In D. D. Parkinson & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics XIII–XIV. Papers from the thirteenth and the fourteenth annual symposia on Arabic linguistics* (pp. 1–37). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kleber, F. (2018). VOT or quantity : What matters more for the voicing contrast in German regional varieties ? Results from apparent-time analyses. *Journal of Phonetics*, 71, 468–486.
- Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412–1425.
- Kohnert, K. (2008). *Language Disorders in Bilingual Children and Adults*. San Diego, CA: Plural.
- Kohnert, K., Ebert, K. D., & Pham, G. T. (2020). *Language disorders in bilingual children and adults* (3<sup>rd</sup> edition). San Diego, CA: Plural Publishing.
- Kohnert, K., Windsor, J., & Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals? *Learning Disabilities Research & Practice*, 21(1), 19–29.
- Kramer, M. (2009). *The phonology of Italian*. New York, USA: Oxford University Press.
- Kuhl, P.K. (2000). Language, mind, and brain: experience alters perception. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 99–115). London: MIT Press.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, 255, 606–608.
- Kupisch, T., & Lleó, C. (2017). Voice Onset Time in German-Italian Simultaneous Bilinguals: Evidence on Cross-Language Influence and Markedness. In M. S. Yavaş, M. Kehoe, & W. Cardoso (Eds.), *Romance-Germanic Bilingual Phonology* (pp. 79–98). Sheffield, UK: Equinox.
- Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5), 564–582.
- Ladefoged, P. & Disner, S. F. (2012). *Vowels and Consonants*. Oxford, UK: John Wiley & Sons.



- Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical Learning in Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research, 60*, 3474–3486.
- Lee, H. J., Kim, Y. T. K., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology, 15*(4), 375–382.
- Lee, S. A. S., & Gorman, B. K. (2012). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism, 17*(4), 479–495.
- Lehti, V., Gyllenberg, D., Suominen, A., & Sourander, A. (2018). Finnish-born children of immigrants are more likely to be diagnosed with developmental disorders related to speech and language, academic skills and coordination. *Acta Paediatrica, 107*, 1409–1417.
- Lenneberg E.H. (1967). *Biological foundations of language*. New York: Wiley.
- Leonard, L. B. (2000). *Children with Specific Language Impairment*. Cambridge, MA, USA: The MIT Press.
- Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word, 20*, 384–422.
- Liu, H.-M., Chen, Y., & Tsao, F.-M. (2014). Developmental Changes in Mismatch Responses to Mandarin Consonants and Lexical Tones from Early to Middle Childhood. *PLoS One, 9*(4), e95587.
- Lloyd-Smith, A., Einfeldt, M., & Kupisch, T. (2020). Italian-German bilinguals: The effects of heritage language use on accent in early-acquired languages. *International Journal of Bilingualism, 24*(2), 289–304.
- Lüke, C., & Ritterfeld, U. (2011). Mehrsprachige Kinder in sprachtherapeutischer Behandlung: Eine Bestandsaufnahme - Bilingual children in speech therapy : Results of a survey. *Heilpädagogische Forschung, XXXVII*(4), 188–197.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*(4), 369-390.
- Macken, M. A., & Barton, D. (1979). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*, 41–74.
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: a

- phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7, 433–458.
- MacLeod, A. A. N. (2016). Phonetic and phonological perspectives on the acquisition of voice onset time by French-speaking children. *Clinical Linguistics & Phonetics*, 30(8), 584–598.
- MacLeod, A. A. N., Fabiano-Smith, L., Boegner-Pagé, S., & Fontolliet, S. (2013). Simultaneous bilingual language acquisition: The role of parental input on receptive vocabulary development. *Child Language Teaching and Therapy*, 29(1), 131–142.
- Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, 20(1), 1–23.
- Marinova-Todd, S. H., Colozzo, P., Mirenda, P., Stahl, H., Kay-Raining Bird, E., Parkington, K., ... Genesee, F. (2016). Professional practices and opinions about services available to bilingual children with developmental disabilities : An international study. *Journal of Communication Disorders*, 63, 47–62.
- Mathieu, J., Lindner, K., Lomako, J., & Gagarina, N. (2016). „Wo bist du , kleiner Monster?“ Sprachspezifische nonword repetition Tests zur Differenzierung von bilingualen typisch entwickelten Kindern und entsprechenden Risikokindern für USES. *Forschung Sprache*, 1, 5–24.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- McAllister Byun, T., Inkelas, S., & Rose, I. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language*, 92(1), 141–178.
- McCarthy, K. M., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech Perception and Production by Sequential Bilingual Children: A Longitudinal Study of Voice Onset Time Acquisition. *Child Development*, 85(5), 1965–1980.
- Mengel, A. (2000). *Deutscher Wortakzent: Symbole, Signale*. Hamburg, Germany: Libri Books on Demand.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-Day-Olds Prefer Their Native Language. *Infant Behavior and Development*, 16, 495–500.
- Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of Mismatch Negativity in Typically Developing Infants and Preschool Children. *Ear & Hearing*, 23, 118–136.
- Munson, B., Kurtz, B. A., & Windsor, J. (2015). Phonotactic Probability, and Wordlikeness on

- Nonword Repetitions of Children With and Without Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, 48, 1033–1047.
- Näätänen, R., & Alho, K. (1995). Mismatch negativity—a unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80(1–4), 317–337.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590.
- Näätänen, Risto. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21.
- Nakamura, J. (2015). Nonnative Maternal Input: Language Use and Errors in a Thai Mother's Interactions in Japanese with her Child. *Japan Journal of Multilingualism*, 21, 10–26.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191–230.
- Orzechowska, P., & Wiese, R. (2015). Preferences and variation in word-initial phonotactics: A multi-dimensional evaluation of German and Polish. *Folia Linguistica*, 49(2), 439–486.
- Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics*, 31, 3–28.
- 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., Genesee, F., & Rice, M. (2003). French-English Bilingual Children With SLI: How Do They Compare With Their Monolingual Peers? *Journal of Speech, Language and Hearing Research*, 46, 1–15.
- Paradis, J., & Genesee, F. (1996). Syntactic Acquisition in Bilingual Children. Autonomous or Interdependent? *Studies in Second Language Acquisition*, 18, 1–25.
- Paradis, J., Schneider, P., & Sorenson Duncan, T. (2013). Discriminating Children With Language Impairment Among English-Language Learners From Diverse First-Language Backgrounds. *Journal of Speech, Language, and Hearing Research*, 1–11.
- Parra, M., Hoff, E., & Core, C. (2011). Relations among Language Exposure, Phonological Memory, and Language Development in Spanish-English Bilingually- Developing Two-Year-Olds. *Journal of Experimental Child Psychology*, 108(1), 113–125.
- Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech

- perception along the voicing continuum. *Journal of Acoustical Society of America*, 55(2), 328–333.
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041.
- Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d / – /ð/: Evidence for a new developmental pattern. *Journal of Acoustical Society of America*, 109(5), 2190–2201.
- Pua, E. P. K., Lee, M. L. C., & Rickard Liow, S. J. (2017). Screening bilingual preschoolers for language difficulties: Utility of teacher and parent reports. *Journal of Speech, Language, and Hearing Research*, 60(4), 950–968.
- Qualtrics (2020). <https://www.qualtrics.com> Copyright © 2021 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA.
- Restrepo, M. A. (2003). Spanish language skills in bilingual children with specific language impairment. In S. Montrul & F. Ordoñez (Eds.), *Linguistic theory and language development in Hispanic languages. Papers from the 5th Hispanic Linguistics Symposium and the 4th Conference on the Acquisition of Spanish and Portuguese* (pp. 365–374). Summerville, MA: Cascadilla Press.
- Restrepo, M. A., Morgan, G. P., & Thompson, M. S. (2013). The Efficacy of a Vocabulary Intervention for Dual-Language Learners With Language Impairment. *Journal of Speech, Language, and Hearing Research*, 56, 748–765.
- Rinker, T., Alku, P., Brosch, S., & Kiefer, M. (2010). Brain & Language Discrimination of native and non-native vowel contrasts in bilingual Turkish-German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain and Language*, 113(2), 90–95.
- Rinker, T. & Gagarina, N. (2014). *CLT – Crosslinguistic Lexical Task – German Version*. University of Konstanz & ZAS Berlin, Germany.
- Rinker, T., Shafer, V. L., Kiefer, M., Vidal, N., & Yu, Y. H. (2017). T-complex measures in bilingual Spanish- English and Turkish-German children and monolingual peers. *PLoS One*, 12(3), e0171992.
- Rispens, J. & Been, P. (2007). Subject–verb agreement and phonological processing in developmental dyslexia and specific language impairment (SLI): a closer look. *International Journal Language Communication Disorders*. 42(3), 293–305.

- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, 8(2), 162–172.
- Roch, M. (2014). *CLT – Crosslinguistic Lexical Task – Italian Version*. University of Padua, Italy.
- Roseberry-McKibbin, C., Brice, A., & O’Hanlon, L. (2005). Serving English Language Learners in Public School Settings. *Language, Speech, and Hearing Services in Schools*, 36, 48–61.
- Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism*, 11(4), 359–389.
- Roy, P., & Chiat, S. (2013). Teasing Apart Disadvantage from Disorder: the Case of Poor Language. In C. R. Marshall (Ed.), *Current Issue in Developmental Disorders* (pp. 125–150). Hove, East Sussex: Psychology Press.
- Roy, Penny, & Chiat, S. (2004). A prosodically-controlled word and nonword repetition task for 2-4 year olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47(1), 223–234.
- Ruiz-Felter, R., Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2016). Influence of current input-output and age of first exposure on phonological acquisition in early bilingual Spanish-English-speaking kindergarteners. *International Journal of Language and Communication Disorders*, 4, 368–383.
- Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, 47, 172–196.
- Sahlén, B., Reuterskiöld-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Non-word repetition in children with language impairment - pitfalls and possibilities. *International Journal of Communication Disorders*, 34(3), 337–352.
- Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39, 187–224.
- Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Valeria, M., ... Lorusso, M. L. (2021). Developmental Language Disorder: Early Predictors, Age for the Diagnosis, and Diagnostic Tools. A Scoping Review. *Brain Sciences*, 11(654), 1–38.

- Scharf, R. J. Demmer, R. T., Silver, E. J., & Stein, R. E. K. (2013). Nighttime Sleep Duration and Externalizing Behaviors of Preschool Children. *Journal of Developmental Behavioral Pediatrics, 34*(6), 384-391.
- Scharff Rethfeldt, W. (2016). Kultursensible logopädische Versorgung in der Krise – zur Relevanz sozialer Evidenz. *Forum Logopädie, 5*(30), 38–43.
- Scharff Rethfeldt, W. (2017). Logopädische Versorgungssituation mehrsprachiger Kinder mit Sprachentwicklungsstörung. *Forum Logopädie, 4*(31), 24–31.
- Scharff Rethfeldt, W. (2019). Speech and Language Therapy Services for Multilingual Children with Migration Background: A Cross-Sectional Survey in Germany. *Folia Phoniatica et Logopaedica, 71*, 116–126.
- Scharff Rethfeldt, W., McNeilly, L., Abutbul-Oz, H., Blumenthal, M., Garcia de Goulart, B., Hunt, E., Laasonen, M. R., Levey, S., Meir, N., Moonsamy, S., Mophosho, M., Salameh, E.-K., Smolander, S., Taiebine, M., Thordardottir, E. (2020). *Common Questions by SLT-SLP About Bilingual-Multilingual Children and Informed Evidence-Based Answers*, IALP: Birkirkara, Malta.
- Schulz, P. & Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Bern, Switzerland: Hogrefe.
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language and Hearing Research, 64*(9), 3578–3593.
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism, 3*(1), 3–25.
- Serratrice, L. (2019). Becoming Bilingual in Early Childhood. In A. de Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 15-35). Cambridge, UK: Cambridge University Press.
- Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., ... Strange, W. (2021). The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-language learners of English. *Bilingualism: Language and Cognition, 1*–14.
- Shafer, V. L., Yu, Y. H., & Datta, H. (2010). Maturation of Speech Discrimination in 4- to 7-Yr-Old Children as Indexed by Event-Related Potential Mismatch Responses. *Ear & Hearing, 31*(6), 735–745.
- Shafer, V. L., Yu, Y. H., & Garrido-Nag, K. (2012). Neural mismatch indices of vowel

- discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neuroscience Letters*, 526(1), 10–14.
- Sharp, K. M., & Gathercole, V. C. M. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh–English bilingual children. *Child Language Teaching and Therapy*, 29(1), 77–89.
- Snijders, T. M., Kooijman, V., Cutler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*, 1178, 106–113.
- Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry*, 47(8), 759–765.
- Stadie, N. & Schöppe, D. (2013). *PhoMo-Kids. Phonologie Modellorientiert für Kinder vom Vorschulalter bis zum dritten Schuljahr*. Köln: Prolog.
- Stankova, M., Rodríguez-Ortiz, I. R., Matic, A., Levickis, P., Lyons, R., Messarra, C., ... Law, J. (2020). Cultural and Linguistic Practice with Children with Developmental Language Disorder: Findings from an International Practitioner Survey. *Folia Phoniatica et Logopaedica*, 1–13.
- Stoehr, A. (2018). *Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time*. Radboud University Nijmegen.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466.
- Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language and Communication Disorders*, 45(4), 480–493.
- Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, 100, 369–388.
- Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, 108(1), 369–388.
- Szagan, G., Stumper, B., & Schramm, S. A. (2014). *FRAKIS. Fragebogen zur Frühkindlichen Sprachentwicklung*. Frankfurt am Main, Germany: Pearson.
- Tamburelli, M., Sanoudaki, E., Jones, G., & Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: Evidence from Polish-English bilingual children. *Bilingualism: Language and Cognition*, 18(4), 713–725.
- Thordardottir, E. (2008). Language-Specific Effects of Task Demands on the Manifestation of Specific Language Impairment: A Comparison of English and Icelandic. *Journal of*

- Speech, Language and Hearing Research*, 51, 922–938.
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426–445.
- Thordardottir, E. (2017). Are background variables good predictors of need for L2 assistance in school? Effects of age, L1, amount, and timing of exposure on Icelandic language and nonword repetition scores. *International Journal of Bilingual Education and Bilingualism*, 1–23.
- 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, 1–16.
- Thordardottir, E., Cloutier, G., Ménard, S., Pelland-Blais, E., & Rvachew, S. (2015). Monolingual or bilingual intervention for primary language impairment? A randomized control trial. *Journal of Speech, Language, and Hearing Research*, 58(2), 287–300.
- Thordardottir, E. T., & Juliusdottir, A. G. (2013). Icelandic as a second language: a longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism*, 16(4), 411–435.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & Brien, M. O. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *J Speech Lang Hear Res*, 40(6), 1245–1260.
- Triarchi-Herrmann, V. (2009). Zur Förderung und Therapie der Sprache bei Mehrsprachigkeit. *Spektrum Patholinguistik*, 2, 31–50.
- Tsao, F., Liu, H., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study, 75(4), 1067–1084.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., ... Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language and Communication Disorders*, 53(4), 888–904.
- United Nations (2015). *THE 17 GOALS | Sustainable Development*. Available online <https://sdgs.un.org/goals> (accessed on February 14<sup>th</sup>, 2022).
- Unsworth, S. (2013). 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.
- Valdés, G. (1995). The teaching of minority languages as academic subjects: Pedagogical and theoretical challenges. *Modern Language Journal*, 79, 299–328.
- Vaughan, H. & Kurtzberg, D. (1992). Electrophysiologic indices of human brain maturation



- and cognitive development. In M. R. Gunnar & C. A. Nelson (Eds.), *Developmental behavioral neuroscience*. Hillsdale, NJ, England: Lawrence Erlbaum Associates.
- von Suchodoletz, W., & Macharey, G. (2006). Stigmatisierung sprachgestörter Kinder aus Sicht der Eltern. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 55(9), 711–723.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of Acoustical Society of America*, 113(2), 1033–1043.
- Watson, I. (1990). Acquiring the voicing contrast in French: A comparative study of monolingual and bilingual children. In J. N. Green & W. Ayers-Bennet (Eds.), *Variation and change in French: Essays presented to Rebecca Posner on the occasion of her sixtieth birthday* (pp. 37–60). London, UK: Routledge.
- Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, 37, 278–286.
- Wiese, R. (1996). *The phonology of German*. New York, USA: Oxford University Press.
- Wig, N., & García-Sierra, A. (2021). Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception. *Bilingualism: Language and Cognition*, 24(3), 467–480.
- Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology*, 14(3), 292–305.
- Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-Language Nonword Repetition by Bilingual and Monolingual Children. *American Journal of Speech-Language Pathology*, 19(4), 298–310.
- Winkler, I., Kujala, T., Alku, P., & Näätänen, R. (2003). Language context and phonetic change detection. *Cognitive Brain Research*, 17, 833–844.
- Wintruff, Y., Orlando, A., & Gumpert, M. (2011). Diagnostische Praxis bei mehrsprachigen Kindern. Eine Umfrage unter Therapeuten zur Entscheidung über den Therapiebedarf mehrsprachiger Kinder mit sprachlichen Auffälligkeiten. *Forum Logopädie*, 1(25), 6–13.
- Yu, Y. H., Tessel, C., Han, X., Campanelli, L., Vidal, N., Gerometta, J., ... Shafer, V. L. (2019). Neural Indices of Vowel Discrimination in Monolingual and Bilingual Infants and Children. *Ear & Hearing Hearing*, 40(6), 1376–1390.
- Zanobini, M., Viterbori, P., & Saraceno, F. (2012). Phonology and Language Development in Italian. *Journal of Speech, Language, and Hearing Research*, 55, 16–31.
- Zmarich, C., & Bonifacio, S. (2005). Phonetic Inventories in Italian Children aged 18-27

months: a Longitudinal Study. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology* (pp. 30–33). Lisbon, Portugal.

# Appendix A

**Table A1.** Bilingual children's individual language experience, including Age of Onset (AoO) and relative amount of current Italian versus German exposure (input and output).

participant	Age (in months)	AoO Italian	AoO German	German input	Italian input	German output	Italian output	
204942	49	0	2.5	28%	72%	23%	77%	
205492	54	0	0	58%	42%	63%	37%	
207112	71	0	0	51%	49%	51%	49%	
207152	71	0	0	53%	47%	60%	40%	
207122	71	0	0	51%	49%	51%	49%	
2071102	71	0	0	69%	31%	67%	33%	
206862	68	0	0	31%	69%	31%	69%	
2046142	49	0	1	50%	50%	73%	27%	
2046132	49	0	1	50%	50%	73%	27%	
106012	60	0	0	27%	73%	16%	84%	
105222	52	0	0	46%	54%	46%	54%	
104832	49	0	0	41%	59%	52%	48%	
105862	58	0	0	47%	53%	41%	59%	
306612	66	0	3	36%	64%	36%	64%	
404912	49	0	0	87%	13%	92%	8%	
406222	62	0	0	91%	9%	100%	0%	
405352	53	0	0	77%	23%	95%	5%	
406272	62	2	0	79%	21%	97%	3%	
405362	53	0	0	77%	23%	95%	5%	
406242	64	missing data						
106492	64	0	0	49%	51%	49,00%	51%	
407332	73	0	0	68%	32%	93,00%	7%	
104572	47	0	0	63%	37%	77,00%	23%	
105582	55	0	0	65%	35%	70,00%	30%	

**Table A2.** Overview of participants' mean ERP Amplitude in  $\mu\text{V}$  to the different deviants vs. their respective identities and the early iMMR averaged across electrode sites Fz, F3, and F4 within the time window of 120-280 ms after stimulus onset according to group (monolingual German vs. bilingual Italian-German).

			Deviant	Identity	early iMMR
German-like Long Lag	Easy (92 ms VOT)	Monolinguals	$M = 3.73,$ $SD = 3.94$	$M = 4.90,$ $SD = 3.63$	$M = -1.17,$ $SD = 3.99$
		Bilinguals	$M = 2.46,$ $SD = 3.69$	$M = 3.74,$ $SD = 3.89$	$M = -1.28,$ $SD = 3.25$
	Difficult (36 ms VOT)	Monolinguals	$M = 7.46,$ $SD = 3.16$	$M = 8.09,$ $SD = 3.76$	$M = -.63,$ $SD = 3.19$
		Bilinguals	$M = 5.63,$ $SD = 2.96$	$M = 5.33,$ $SD = 3.44$	$M = .30,$ $SD = 2.51$
Italian-like Voicing Lead	Easy (-112 ms VOT)	Monolinguals	$M = 5.99,$ $SD = 4.08$	$M = 7.02,$ $SD = 4.06$	$M = -1.03,$ $SD = 3.42$
		Bilinguals	$M = 4.74,$ $SD = 3.75$	$M = 5.38,$ $SD = 3.32$	$M = -.63,$ $SD = 3.87$
	Difficult (-36 ms VOT)	Monolinguals	$M = 7.84,$ $SD = 2.68$	$M = 8.17,$ $SD = 2.67$	$M = -.33,$ $SD = 3.21$
		Bilinguals	$M = 4.40,$ $SD = 3.10$	$M = 6.00,$ $SD = 2.99$	$M = -1.59,$ $SD = 2.59$

**Table A3.** Overview of participants' mean ERP Amplitude in  $\mu\text{V}$  to the different deviants vs. their respective identities and the late iMMR averaged across electrode sites Fz, F3, and F4 within the time window of 360-520 ms after stimulus onset according to group (monolingual German vs. bilingual Italian-German).

			Deviant	Identity	late iMMR
German-like Long Lag	Easy (92 ms VOT)	Monolinguals	$M = -10.62,$ $SD = 5.30$	$M = -7.26,$ $SD = 6.13$	$M = -3.37,$ $SD = 4.71$
		Bilinguals	$M = -12.55,$ $SD = 6.21$	$M = -7.99,$ $SD = 6.73$	$M = -4.57,$ $SD = 4.76$
	Difficult (36 ms VOT)	Monolinguals	$M = -10.34,$ $SD = 4.60$	$M = -6.22,$ $SD = 5.25$	$M = -4.11,$ $SD = 3.75$
		Bilinguals	$M = -8.33,$ $SD = 6.44$	$M = -7.59,$ $SD = 5.31$	$M = -.74,$ $SD = 3.84$
Italian-like Voicing Lead	Easy (-112 ms VOT)	Monolinguals	$M = -9.96,$ $SD = 4.37$	$M = -8.24,$ $SD = 5.30$	$M = -1.72,$ $SD = 3.96$
		Bilinguals	$M = -10.49,$ $SD = 6.43$	$M = -8.31,$ $SD = 5.57$	$M = -2.18.,$ $SD = 3.94$
	Difficult (-36 ms VOT)	Monolinguals	$M = -6.41,$ $SD = 3.91$	$M = -5.02,$ $SD = 3.59$	$M = -1.39,$ $SD = 5.14$
		Bilinguals	$M = -7.96,$ $SD = 5.37$	$M = -6.15,$ $SD = 5.31$	$M = -1.81,$ $SD = 3.97$

## Appendix B

**Table B1.** Overview of all questions in the questionnaire study presented in chapter 3

Question Number	Question	Answering options
Q1	For how many years have you practiced/ have you been practicing Speech and Language Pathology?	< 5 years 5 - 10 years 10 - 20 years > 20 years
Q2	Have you ever provided therapy to bilingual children with developmental language disorder (DLD)	yes sometimes never
Q3	What percentages of children you diagnose/ treat for DLD are bilingual?	0 - 5% 6 - 2% 26 - 50% 51 - 75% 76 - 95% 96 - 100%
Q4	Besides German, do you speak any other language(s) at native level?	yes no
Q5	If yes (Q4), what language(s) other than German do you speak at native level?	[open question, free text answer]
Q6	Do you offer your SLP services in multiple languages?	yes no
Q7	If yes (Q6), in which language(s) other than German do you offer your SLP services?	[open question, free text answer]
Q8	Do you think that different approaches are needed in the diagnosis and treatment of bilingual children compared to monolingual children?	yes sometimes never
Q9	Do you think that (A) the therapy of DLD of bilingual children should be limited exclusively to the societal language, or that (B) the child's first language should also be taken into account?	A A more than B B more than A B
Q10	Do you think that (A) DLD is independent of the second language, or that (B) a disorder can be exacerbated by second language acquisition?	A A more than B B more than A B
Q11	Do you think that (A) it would be better for bilingual children with DLD to speak only one language (both at home and outside of the family environment), or that (B) it	A B

	would be better for all of a child's communication partners to interact with the child in the language they know best?	
Q12	In the context of speech and language therapy (SLP) for bilingual children with DLD, do you think that it is useful to compare children's language performance in his/ her first and second language?	yes sometimes rarely never
Q13	If yes/ sometimes/ rarely (Q12)... (A) were your comparisons based on information provided by the parents or (B) were you able to directly observe children's behavior in both languages (possibly in the presence of parents)?	A A more than B B more than A B
Q14	When assessing bilingual children's language performance: How informative do you consider the use of tests/screenings in the child's heritage language?	very informative informative little informative not informative
Q15	How feasible (i.e., applicable) do you consider the use of tests/screenings in children's heritage language?	very feasible feasible poorly feasible not at all feasible
Q16	When assessing bilingual children in the societal language German: Do you find it sufficient to refer to monolingual German norm data?	completely sufficient sufficient insufficient completely insufficient
Q17	When assessing bilingual children: How relevant do you find it that tests/screenings provide specific norm data for this population?	highly relevant relevant irrelevant completely irrelevant
Q18	Are you aware of any testing or other diagnostic material that have been developed specifically for bilingual children with DLD?	yes no
Q19	Do you use special diagnostic material/ tools for bilingual children with DLD?	yes sometimes rarely never
Q20	If yes/ sometimes/ rarely (Q19)... What diagnostic material/ tools do you use for bilingual children with DLD?	[open question, free text answer]
Q21	Do you think it would be useful to check whether the phonemes that cause difficulties to the child in the societal language are present/ also affected in the heritage language?	yes sometimes rarely never
Q22	Do you think it would be useful to check whether the syntactic and morphological structures that cause difficulties to the child the societal language are present/ also affected in the heritage language?	yes sometimes

		rarely never
Q23	Do you think it would be useful to check whether words that the child uses semantically/ lexically incorrectly in the societal language are similar or very different in the heritage language?	yes sometimes rarely never
Q24	Do you think it would be useful to check whether words that the child uses semantically/ lexically incorrectly in German (i.e. the societal language) are also used incorrectly in the heritage language?	yes sometimes rarely never
Q25	Do you think it would be useful to have a summary chart of the phoneme inventory of the child's heritage language?	yes sometimes rarely never
Q26	Do you think it would be useful to have a summary table of the main syntactic structures and constructions in the child's heritage language?	yes sometimes rarely never
Q27	Do you think that an overview table with a list of the most important prepositions (translation + usage) in the child's heritage language would be useful?	yes sometimes rarely never
Q28	Do you feel competent when assessing/treating bilingual children for DLD?	yes no
Q29	Do you gather information on the assessment and treatment of bilingual children?	yes no
Q30	Are you familiar with the LITMUS test-battery (Cost Action IS0804)?	yes no
Q31	Do you incorporate information about children's language exposure (i.e., how often a child hears and/or speaks each of their respective languages into the diagnostic process when you assess bilingual children's need for Speech and Language Therapy?	yes no
Q32	If yes (Q31), which particular information do you include?	[open question, free text answer]
Q33	If yes (Q31), how do you include this information in the assessment?	[open question, free text answer]
Q34	How relevant do you think it is to distinguish between bilingual children with language difficulties in German due to insufficient exposure to the societal language and bilingual children with DLD?	highly relevant relevant irrelevant completely irrelevant

---

Q35	Please explain your decision (Q34)	[open question, free text answer]
Q36	In your opinion, what are the possible consequences of overidentifying DLD (i.e., attesting a bilingual child who is actually linguistically typically developing with a language disorder)?	[open question, free text answer]
Q37	In your opinion, what are the consequences of under-identifying DLD (i.e., not recognizing DLD in a bilingual child who actually needs therapy)?	[open question, free text answer]

---