ASSESSING URDU-SPEAKING MULTILINGUAL CHILDREN IN HONG KONG AND PAKISTAN: HERITAGE LANGUAGE ACQUISITION AND DEVELOPMENTAL LANGUAGE DISORDER

HAMDANI, SABOOR ZAFAR

Doctor of Philosophy

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Assessing Urdu-Speaking Multilingual Children in Hong Kong and Pakistan: Heritage Language Acquisition and Developmental Language Disorder

HAMDANI, Saboor Zafar

A thesis submitted in partial fulfilment of

the requirements for the degree of Doctor of Philosophy

August 2022

Certification Of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

__________________ (Signed)

HAMDANI, Saboor Zafar (Name of Student)

Abstract

This dissertation examined the language abilities of children acquiring Urdu as their first language (L1) in Hong Kong and Pakistan, addressing a range of conceptual and methodological issues in multilingual child language acquisition.

Study 1 addressed two key issues in the heritage language acquisition of multilingual ethnic minority (EM) children: the effects of reduced language experience and linguistic vulnerabilities. This study assessed a group of typically developing (TD) EM multilingual children acquiring Urdu as L1 and minority heritage language (The EM group; N=31; mean age = 8;1 [years; months], range = 6;1–10;11), and compared their L1 language abilities with their age- gender- and grade-matched TD peers acquiring Urdu as L1 and majority language in Pakistan (The majority group; N=31; mean age = 8;1; range = 6;1–10;10). In general, the EM children residing abroad acquire their L1 Urdu under reduced input conditions in their host countries. Whereas children from the majority group living in Pakistan acquire their L1 Urdu under relatively abundant input conditions. It is predicted that reduced input might affect certain linguistic abilities more than other linguistic abilities, and reduced input associated with minority language acquisition context might lead to more restricted age-related progress in the first/heritage language (L1) attrition.

Both groups were assessed in a range of outcome measures, encompassing lexical comprehension and production, expressive morphosyntax, narrative comprehension and production, using the Urdu versions of Crosslinguistic Lexical Tasks (CLT, Haman et al., 2015; Hamdani et al., 2020a), and Sentence Repetition task (SRep, Marinis & Armon-Lotem, 2015; Hamdani et al., 2020b), and Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2019a; Hamdani et al., 2020c), respectively.

The results showed that the EM children obtained comparable scores to their majority peers in the domains of narrative comprehension, use of internal state terms, and narrative macrostructure production (story complexity and story structure). On the other hand, the EM children demonstrated significantly poor performance than their majority peers in the lexical, morphosyntactic, and grammaticality (proportion of grammatical C-Units) domains. Although the EM children were similar to their majority peers in terms of exhibiting comparable age-related progress in other language outcome measures, they differed in two narrative microstructure measures indexing syntactic complexity and grammaticality, specifically, mean length of C-units and verb accuracy. In these two measures, the EM children demonstrated restricted age-related progress as compared to their majority counterparts. The overall error analyses also showed that the EM children made significantly more morphosyntactic errors (during narrative production and sentence repetition) and lexical errors than their majority group peers.

The findings brace the notion that the effect of language input conditions varies across different language domains. Lexical and morphosyntactic competence are more susceptible to input conditions. Whereas narrative macrostructure is comparatively more resilient to language-specific experiences/knowledge.

Study 2 addressed a long-standing issue of differentiating between the effects of language experience and genuine language disorder in multilingual children- a challenging issue in identifying Developmental Language Disorder (DLD) in these children. It presents a case study to demonstrate the potential of the CATALISE diagnostic criteria (Bishop et al., 2017) in combination with the assessment tools from the European Cost Action Language Impairment Testing in Multilingual Settings (LITMUS) battery, to identify language difficulties in a multilingual child with suspected DLD via remote online testing.

The participants included one six-year-old (6;8) Urdu-Cantonese EM multilingual child at risk for DLD (S-DLD), and seven age and grade-matched multilingual peers from similar linguistic backgrounds. Their abilities in multiple language areas were assessed via Zoom using Urdu versions of the CLT (Haman et al., 2015; Hamdani et al., 2020a), SRep (Marinis & Armon-Lotem, 2015; Hamdani et al., 2020c), MAIN (Gagarina et al., 2019a; Hamdani et al., 2020c), and CL-NWR (Chiat, 2015; Hamdani et al., 2020d). A nonverbal IQ test (Raven's Progressive Matrices; Raven & Court, 1998) and a parental questionnaire (LITMUS-PABIQ; Tuller, 2015) were also administered.

The findings indicated that the child with S-DLD scored significantly lower in multiple language measures compared to her peers in her best/first language, Urdu. Considering also the presence of negative functional impact (language problems affecting her academic and social skills, day-to-day communication) and poor prognostic features (problems persisting beyond age five, problems in multiple linguistic domains, and low receptive skills), and absence of associated biomedical conditions, the results collectively suggested that this participant could be identified as a child with DLD (Bishop et al., 2017).

Study 3 also addressed the prevailing issue of differentiating between the effects of language experience and genuine language disorder in multilingual children but focused on Nonword repetition (NWR). NWR has been reported as one of the most sensitive measures in discriminating between children with and without language disorders cross-linguistically. Since nonwords are supposedly unfamiliar words to all children acquiring a certain language, NWR is less influenced by prior linguistic knowledge and therefore could be less biased against multilingual children from diverse language backgrounds who might have less experience of the language targeted during the assessment. This is, therefore, the first empirical study examining the clinical utility of NWR in identifying DLD in Urdu-speaking children. Specifically, it examined whether the new Urdu NWR test would disadvantage TD children with reduced experience in the language of assessment (by comparing TD children with and without reduced experience) and whether this NWR test can also discriminate between children with typical development but reduced experience to the language being assessed and children with DLD.

Three groups of children were compared: (i) a group of typically developing (TD) children acquiring Urdu as L1 and majority language in Pakistan (The majority language TD group; N=31; mean age = 8;1, range = 6;1–10;10); (ii) a group of age- gender- and grade-matched TD peers acquiring Urdu as L1 and minority heritage language in Hong Kong who had reduced experience of the language of assessment (The minority language TD group; N=31; mean age = 8;1, range = 6;1–10;11); and (iii) a group of age-matched peers with DLD acquiring Urdu as L1 and majority language in Pakistan who do not have the issue of reduced experience of the language of assessment but suffer from language disorder (The majority language DLD group: N=14; mean age = 8;1, range = 6;1–10;11). NWR performance was assessed by adapting a quasi-universal NWR task from the European COST Action IS0804 LITMUS battery (currently known as the LITMUS Crosslinguistic Nonword Repetition test, CL-NWR) into Urdu (Chiat, 2015; Hamdani et al., 2020d).

Results indicated that the majority language DLD group scored significantly lower than both the majority and minority language TD groups at the segment correct level. Moreover, children from the minority language TD group performed similarly to their majority language TD peers suggesting that the new Urdu NWR test did not disadvantage these minority language TD children acquiring Urdu as L1 under reduced input conditions. The findings provided the first evidence that the Urdu CL-NWR test could yield significant TD/DLD group differences, even for TD children with reduced language experience.

The three studies in this dissertation are significant in many ways. Findings from Study 1 provide beneficial insights for practitioners and parents working with the EM children related to focused intervention by elucidating their linguistic weaknesses. Findings from Study 2 demonstrated the promise of using the CATALISE framework, the LITMUS battery tasks, and remote online testing in identifying DLD and collecting reference data (i.e., data from a smaller sample to guide developmental expectations when normative data from the larger sample is not available) in multilingual children. Findings from Study 3 demonstrated the diagnostic potential of the Urdu CL-NWR test in identifying DLD in Urdu-speaking children. Urdu, although ranked 10th in the world as a major world language (Eberhard et al., 2022), is lacking in assessment tools. Overall, this dissertation creates four new assessment tools to assess linguistic competence in Urdu-speaking children. Based on their flexibility to be adapted into any language and clinical utility to identify DLD in multilingual children, these new Urdu assessment tools might be practical for many researchers.

Publications Arising from This Thesis

Hamdani, S. Z., Chan, A., Kan, R., Chiat, S., Gagarina, N., Haman, E., Łuniewska, M., Polišenská, K., & Armon-Lotem, S. (2024). Identifying Developmental Language Disorder (DLD) in Multilingual Children: A Case Study Tutorial. *International Journal of Speech-Language Pathology*. [accepted]

Hamdani, S. Z., Kan, R., Chan, A. & Gagarina, N. (2020). The Multilingual Assessment Instrument for Narratives (MAIN): Adding Urdu to MAIN. *ZAS Papers in Linguistics, 64,* 257–261. https://doi.org/10.21248/zaspil.64.2020.580

Hamdani, S. Z., Chan, A., Polišenská, K., & Chiat, S. (2024). Non-word repetition as a clinical marker of Developmental Language Disorder in Urdu: evidence from the Urdu LITMUS Crosslinguistic Nonword Repetition test (Urdu CL-NWR) [in preparation]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong.

Hamdani, S. Z., Chan, A., Gagarina, N., Haman, E., Łuniewska, M., & Armon-Lotem, S. (2024). Learning Urdu as a minority heritage language: reduced input effects and linguistic vulnerabilities [in preparation]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong.

Acknowledgments

First of all, I would like to thank the Almighty for providing me with this spectacular opportunity and blessing me with such amazing and supportive people in my life. I would like to extend my heartfelt gratitude to my Ph.D. supervisor, Dr. Angel Chan, who has been an incredible mentor throughout this journey. She has always been kind and supportive. Her encouraging behavior helped me to always strive for better. Apart from being an incredible mentor, she is also a beautiful human being and I have learned so much from her not only professionally but also on a personal level. She will always be my role model and the things I have learned from her will always stay with me.

Next, I would like to thank my parents, siblings (Asna and Faizan), nephew (Shah zain), husband (Arslan), and friends (especially Saira and Vasundhara) for their never-ending support and healing words that have played a significant role in lifting me up and completing this journey successfully. I am also deeply indebted to my research assistants for their assistance and for keeping up with me patiently.

My admiration and respect also go out to the participants and their parents whose participation made this thesis possible. Last, but not the least, I am grateful to my co-supervisors Dr. Ewa Haman and Prof. Natalia Gagarina, all the collaborators and everyone else who has been a part of this journey directly or indirectly.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CLT	Crosslinguistic Lexical Tasks
COST Action	European Cooperation in Science and Technology Action
Cprop	Proportion Complex C-units
DLD	Developmental Language Disorder
EM	Ethnic minority
Gprop	Proportion of Grammatical C-units
IQ	Intelligence Quotient
IST	Internal State Terms
L1	First language
L2	Second language
LD	Language Disorder
LITMUS	Language Impairment Testing in Multilingual Settings
MAIN	Multilingual Assessment Instrument of Narratives
Maj	Majority Group
MLCU	Mean Length of C-units
Max	Maximum
NDW	Number of Different Words
NWR	Nonword Repetition
ORC	Object Relative Clause
SC	Syntactic Complexity
SLT	Speech-Language Therapist
Sprop	Proportion of Subordinate C-units
SRep	Sentence Repetition
SS	Story structure
TD	Typically Developing
TNCU	Total Number of C-units
TNW	Total Number of Words
VA	Verb Accuracy
WC	Word Category

Chapter 1 - Introduction

1. Introduction

The maintenance of heritage language (HL) is central to an individual's well-being (De Houwer, 2015; Yu, 2013) and the preservation of that language at the communal level (Potowski, 2013). However, studies investigating comprehensively the acquisition and maintenance of minority language in ethnic minority (EM) multilingual children are relatively scarce (V. C. M. Gathercole & E. M. Thomas, 2009; Winsler et al., 1999). Given this, the present thesis focuses on the HL (Urdu) acquisition of the Urdu-Cantonese EM multilingual children in Hong Kong. This thesis employs the term "multilingual" based on the definition by the International Expert Panel on Multilingual Children's Speech (2012):

> People who are multilingual, including children acquiring more than one language, are able to comprehend and/or produce two or more languages in oral, manual, or written form with at least a basic level of functional proficiency or use, regardless of the age at which the languages were learned (International Expert Panel on Multilingual Children's Speech, 2012, p. 1).

This chapter begins with the formal introduction of minority heritage language and heritage speakers, followed by sections highlighting this thesis's conceptual context and practical constraints practitioners face when assessing language competence in EM multilingual children. Furthermore, it provides an overview of the present thesis's methodological novelty, conceptual framework, and three studies.

1.1 Minority Heritage Language

The term "heritage", as defined by the American Heritage College Dictionary, is a property that is passed down by the prior generations or anything that is acquired at the time of birth or can be inherited. So, if language is also an acquired property, then all languages used humans that are learned at the time of birth and passed down to the next generations can be claimed as heritage languages. However, the word heritage, when used with languages and speakers, does not refer to its absolute characterization. Instead, it depicts its relative characterization.

The term heritage language was initially employed in Canada during the 1970s, and then in the 1990s, it was used in the United States to denote minority languages (Cummins, 2005). In simple words, heritage language refers to the language spoken at home by the EM groups and is not the dominant societal language (Rothman, 2009). It is also known as mother tongue, minority, first, home, family, ethnic, and indigenous language (Montrul, 2016; Ortega, 2020; Wiley, 2008). This thesis will use the terms heritage, minority, and first (L1) language interchangeably, given the acquisition context that it refers to.

A language is identified as a heritage language based on its social status at the regional level and not how it is characterized in the glossary. In addition, as employed in various fields, including linguistics, language policy, education, and others, heritage language is barely neutral based on its socio-political overtones associated with the distinctions between minority and majority languages (Montrul, 2016). As mentioned earlier, heritage languages are used by the EM groups and might or might not hold a co-official status. Heritage languages are usually relegated within territories or nations for different reasons. The given ethnic group could be a demographic minority or, despite a large population, might have lower political and social status linked to circumstances like colonization or immigration. The majority languages, on the contrary, often hold official status and are used for education, government administration, and even media. In other words, the minority vs. majority distinction is not an intrinsic feature of any language but is established based on its regional context (Montrul, 2016). For example, Urdu is the majority language in Pakistan but holds a minority status in Hong Kong.

1.2 Heritage Speakers

Following the definition of heritage language, the heritage speakers could be broadly defined as multilingual individuals who belong to EM language groups and are exposed to minority heritage language at home and majority language via community (Benmamoun et al., 2013; Montrul, 2018). Montrul (2016) has outlined some defining characteristics of heritage speakers as follows:

- multilingual individuals who grew up in a multilingual setting and have some proficiency in the two languages
- speakers whose first language or any of the first languages used at home holds minority status in the local context (the heritage language)
- speakers whose heritage language is usually the weaker language
- multilinguals are often dominant in the majority/societal language (although there are also balanced heritage speakers)
- speakers whose level of proficiency in heritage language ranges from mainly receptive (minimal) to fully fluent (native-like)
- speakers whose level of proficiency in the majority/societal language could be native-like (depending on their education level)

From the above-defining characteristics, it can be concluded that even though young heritage speakers belong to various cultures, linguistics backgrounds, socioeconomic status (SES), and education levels and are also exposed to different registers and varieties of their heritage language, they all share one common feature that with growing age they often become dominant in the majority/societal languages. On the other hand, their competency in the heritage language widely varies from having only receptive skills to advanced expressive skills.

Competency in the heritage language is mainly associated with an individual's language learning experience. This experience includes factors like the age when the majority language was acquired, degree of heritage language exposure/input and usage during the language learning phase, how it is used at home and the community level, access to schooling in heritage language, etc. Since heritage speakers are mostly raised under circumstances of subtractive multilingualism, the core features of their heritage language and their linguistic competence might be comparatively different than the linguistic abilities of their age-matched counterparts from their home countries or even parents (Montrul, 2010). Thus, understanding and explaining these wide discrepancies in acquisition outcomes is essential for academic, clinical, and research purposes.

1.3 Conceptual Context: Effects of Reduced Input and Developmental Language Disorder (DLD)

The sociolinguistic status (minority vs. majority) of the two languages acquired by the multilinguals is associated with the overall language input conditions for each language (Paradis, 2010). The context of heritage language development based on its minority status is generally characterized by the reduced amount of input (Montrul, 2016), which could lead to vulnerabilities in certain linguistic domains (e.g., lexical: Bohnacker et al., 2016; morphosyntactic: Haman et al., 2017; expressive narrative microstructure competence: Rodina, 2017). This continuously diminishing heritage language input and use often leads to incomplete heritage language acquisition (Anderson, 2001) or attrition of the indigenous linguistic system, which entails the loss of language abilities in a multilingual situation (Benmamoun et al., 2013; Montrul, 2016; Polinsky 2006; Scontras et al., 2015). It is possible that specific language abilities, for instance, the acquisition of a grammatical structure, first reach complete mastery but then "erode" over time due to several years of disuse or limited input (Scontras et al. 2015).

Moreover, the age at which language exposure to different languages begins is speculated to be highly correlated to the acquisition of heritage language/L1, L2, and even heritage language attrition in multilinguals. The idea that language acquisition should occur in childhood to attain native-like mastery has existed for many years and also draws support from human development (Lenneberg, 1967) and neurological studies (Penfield, 1953). The biological underpinning of language and its pertinence to the critical period hypothesis was put forward by Penfield (1953) and Lenneberg (1967). This hypothesis proposes that language acquisition is sensitive to specific periods in development, beyond which success in achieving native-like competence is no longer possible. Studies have frequently reported positive age effects in typical monolingual language acquisition, but this is not always true in multilingual heritage language development (Gagarina & Klassert, 2018). Studies have reported gains and decline/fossilization (attrition) with age in the L1 acquisition (Meir & Jansen, 2021).

Young children in such instances are likely to lose their heritage language abilities more quickly than individuals who immigrated to another country in their later life and whose heritage language was acquired substantially before relocation (Hulsen, 2000). Studies such as Montrul (2008) and Flores (2012) have shown that heritage language attrition could emerge early in younger children, dramatically impacting language-specific skills like grammar. In other words, the stretch of attrition and acute language loss, depending on different factors, could be more prominent in younger children than in individuals who immigrated after puberty (Scontras et al., 2015). The L1 attrition in a minority context may also lead to restricted progress across age (compared to L1 acquisition in the majority context, which would show more prominent progress across age).

Furthermore, linguistic vulnerabilities associated with reduced input could resemble linguistic vulnerabilities related to Developmental Language Disorder (DLD; Blom et al., 2019). DLD impacts two children in every class of 30. It mainly affects the expressive language skills in children, in the absence of biomedical conditions such as intellectual disability, hearing loss, and developmental disorders such as Autism Spectrum Disorder (ASD) (Norbury et al. 2016). It emerges in childhood and can persist into adulthood if undiagnosed and untreated.

Differentiating the effects of reduced input and the impact of DLD might be more challenging in a heritage minority language context among EM multilingual children. There can be apparent overlaps between the language profiles of typically developing (TD) multilingual children and monolingual children with DLD (Armon-Lotem & de Jong 2015). For instance, DLD may be characterized by restricted vocabulary competence and difficulties in understanding and producing complex sentence structures, which are also commonly attested in the language development of TD multilinguals who have reduced experience of the target language (Bedore & Peña, 2008). Given these overlaps and less awareness about DLD in the general public, there could be higher chances of over- or under- identification of DLD in multilingual children (Armon-Lotem & de Jong 2015).

Studies have also shown that multilingual children who have reduced experience with a language often appear to lag behind their monolingual peers in multiple linguistic domains when being assessed in one language (e.g., Bonifacci et al., 2018; Haman et al., 2017), but when all languages of the multilinguals are taken into account, many of them are found to perform on par with their monolingual peers (Core et al., 2013). Therefore, these TD children with normal language learning potential must be distinguished from children who are not fully proficient in their first and second languages due to a genuine disability in learning any language (i.e., children having language disorder).

1.4 Difficulty in Disentangling Language Experience Versus Language Disorder

Multilingual children from ethnic minorities are expected to acquire both (or all) languages under reduced input conditions, even their L1 (also their heritage language). This might lead to restricted language competence in one or both languages (especially L2), even for TD children with a normal language learning potential. Thus, disentangling the effects of multilingualism and genuine language disorder becomes a challenge for practitioners working with EM children around the globe. The ensuing sections provide a brief account of issues and challenges faced by practitioners working with these EM children.

1.4.1 Practical Constraints. Competency in a language (especially the best language) of a multilingual child is a crucial consideration when making diagnostic decisions regarding the presence or absence of language disorders in multilingual children (Bishop et al., 2017; Peña et al., 2020). However, the speech-language therapists (SLT) and educators working with these EM multilingual children in Hong Kong often do not know the children's home language (which is also the L1 and often the best language). They can mainly assess these children in Cantonese (L2) only, which is generally not their best language. For younger children who entered school not too long ago, their Cantonese (school language) is expected to be still weak.

Therefore, the under-identification of language impairments is a common dilemma when assessing younger multilingual children (Bedore & Peña, 2008). The younger multilingual children (compared to the older multilingual children) have a greater risk of being under-identified for language disorders, as some practitioners often adopt a wait-and-see approach. They are under the impression that due to the young age these children might take longer to develop multilingual competence.

The lack of appropriate resources for assessing the best language of these multilingual EM children is another impediment. Proper assessment tools to evaluate the speech and language abilities of these multilingual children’s heritage language (likely also the best for younger children) are usually lacking or absent. On the other hand, if available, the existing assessment tools/measures are usually normed for predominantly monolingual children. These monolingual norms should not be generalized to multilingual learners (Armon-Lotem, 2018; Bedore & Peña, 2008). This whole situation exacerbates the diagnostic challenges, which could lead to over- and under-identification of language disorders and also affect the provision of support for EM multilingual children with and without language impairments in Hong Kong and around the globe.

1.5 Methodological Novelty of this Thesis

To address the challenges in assessing the language abilities of multilingual children and to improve the differentiation of multilingual children with and without language impairments, there has been an extensive research consortium funded by the European COST (Cooperation in Science and Technology) Action IS0804 (2009–2013), which has developed an assessment battery titled “Language Impairment Testing in Multilingual Settings” (LITMUS- battery). Their primary aim is to refine the assessment of multilingual children with and without language impairment to extricate the effects of DLD and multilingualism. This research consortium has developed multiple assessment tools for multilingual children, some of which have formed the conceptual and methodological basis for different studies of this thesis.

Different assessment measures targeting multiple linguistic domains from this LITMUS battery were adapted into Urdu to ensure that the best language of the young Urdu-Cantonese EM multilingual children could be adequately assessed. The details regarding each tool are presented in the relevant chapters.

Table 1.1 below lists the number of language versions available for each tool that has been adapted into Urdu for this project, with information extracted from its respective website and provided by the key members of the LITMUS committee: Crosslinguistic Lexical Tasks (CLT: Haman & Łuniewska); Multilingual Assessment Instrument of Narratives (MAIN: Gagarina); Crosslinguistic Nonword Repetition Test (CL-NWR: Chiat & Polišenská), Sentence

Repetition (SRep: Armon-Lotem). This table shows how widely these assessment tools are being used around the globe.

Table 1.1 *The Current Crosslinguistic Coverage of the LITMUS Assessment Tools*

Assessment tools	Total language versions	Available language versions	Links to official webpages
Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2019a)	94	Afrikaans, Albanian, Akan, Lebanese Arabic, Palestinian Arabic, Syrian Arabic, Iraqi Arabic, Arabic (Modern Standard), Arabic (Saudi), Azeri, Bagri, Basque, Bengali, Bosnian, Bulgarian, Cantonese-Chinese, Catalan, ChabaCano, Chuvash, Croatian, Czech, Danish, Dutch, English, Estonian, Farsi, Finnish, French, Frisian (West), Georgian, German, Gilaki, Gondi, Greek (Cypriot), Greek, Halbi, Hebrew, Hindi, Hindi for Halbi/Gondi-Hindi multilinguals, Hmong, Hungarian, Icelandic, Irish (Gaeilge), IsiXhosa, Italian, Kam, Kannada, Kazakh, Konkani, Korean, Kurmanji (Kurdish), Lari, Lithuanian, Luganda, Luxembourgish, Malayalam, Maltese, Mandarin-Chinese, Marathi, Megrelian, Mizo, Montenegrin, Nepali, Norwegian, Odia, Polish, Portuguese (Brazilian), Russian, Sahngo, Sámi (North), Scottish Gaelic, Serbian, Shughni, Slovak, Spanish, Swedish, Tagalog, Tajik, Tamil, Tati, Telugu, Tibetan, Torwali, Tshivenda, Turkish, Turkish for Turkish-Swedish multilinguals (Sweden), Ukrainian, Urdu (Pakistan, India), Uyghur, Uzbek, Vietnamese, Yakut, Zulu	https://main.leibniz-zas.de/

Assessment tools	Total language versions	Available language versions	Links to official webpages
Sentence Repetition task (SRep, Marinis & Armon-Lotem, 2015)	30	Albanian, Lebanese Arabic, Palestinian Arabic, Saudi Arabic, Syrian Arabic, Cantonese, Catalan, Croatian, Dutch, English, Farsi, French, Gaelic, German, Greek Cypriot, Standard Greek, Hebrew, Irish, Italian, Lithuanian, Malay, Maltese (created before the LITMUS), Norwegian, Polish, Portuguese, Russian, Spanish, Turkish, Urdu, Welsh	https://www.litmus-srep.info/
Crosslinguistic lexical task (LITMUS-CLT, Haman, Łuniewska, & Pomiechowska, 2015)[1]	38	Afrikaans, Arabic (Lebanese), Cantonese, Catalan, Czech, Dutch, English (American, British, Malaysian, South African), Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Irish, IsiXhosa, Italian, Lithuanian, Luxembourgish, Malay, Maltese, Mandarin, Norwegian, Persian, Polish (revised), Russian, Scottish Gaelic, Serbian 1, Serbian 2, Slovak, Spanish, Swedish, Turkish, Urdu, Western Armenian	https://multilada.pl/en/projects/clt/
Crosslinguistic Nonword Repetition Test (CL-NWR, Chiat, 2015)	19	Versions have been created in: Arabic (for Arabic-Swedish multilinguals), Austrian German, Cantonese, Dutch, English (British, Canadian, Irish, South African), Finnish, French (Canadian), German, Greek, Malay, Maltese, Mandarin, Slovak, Swedish, Tamil, Urdu A Language-Neutral version of the CL-NWR is available for universal use, with the	https://www.bi-sli.org/cl-nonword-repetition

[1] Currently, it is also possible to test children with the CLT using a free mobile app, available from AppStore: https://apps.apple.com/us/app/child-lexicon-clt/id1620554657 and Google Play: https://play.google.com/store/search?q=child%20lexicon%20clt&c=apps

Assessment tools	Total language versions	Available language versions	Links to official webpages
		pronunciation of consonants and vowels as neutral as possible between languages	

1.6 Conceptual Framework

1.6.1 The UK CATALISE Diagnostic Criteria (Bishop et al., 2017). In addition to adapting different tools from the LITMUS battery to Urdu, this Ph.D. project also applied Bishop and colleagues' UK CATALISE diagnostic criteria (2017) as a conceptual basis for identifying children with DLD (see Chapter 3). The following flow chart in Figure 1.1 highlights the significant factors considered during the diagnostic/identification process:

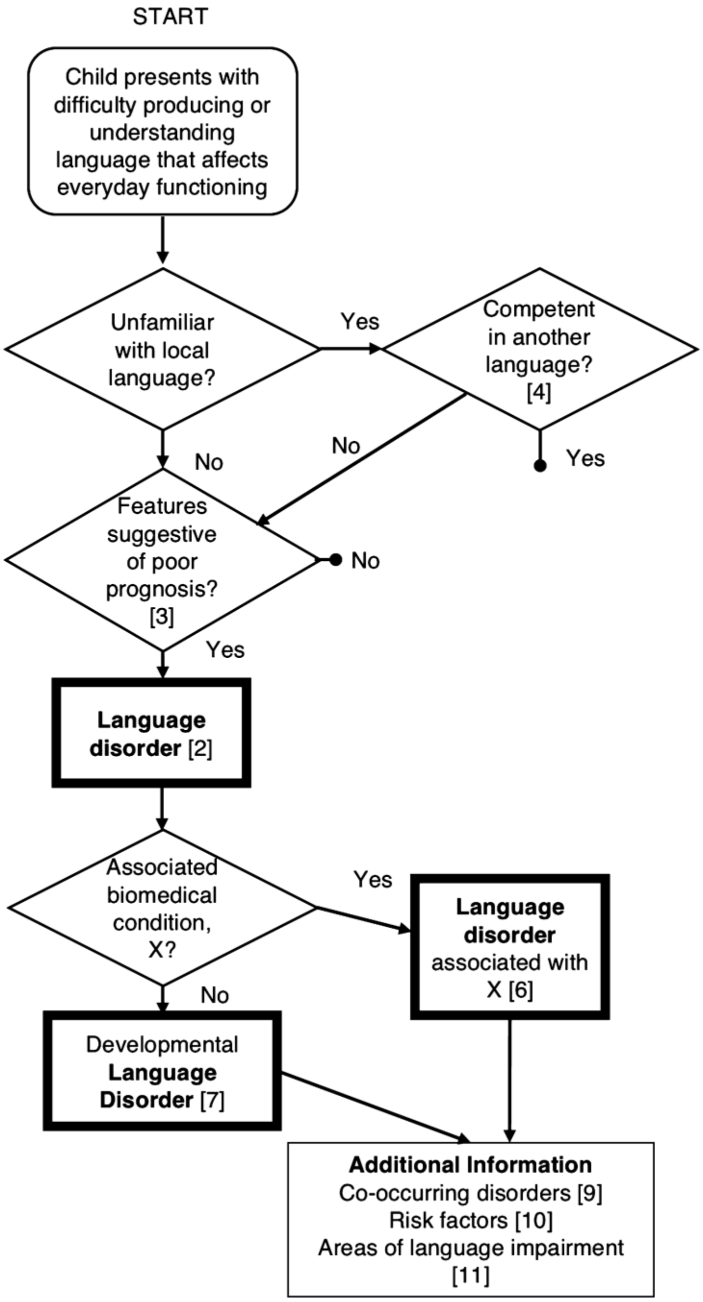

Figure 1.1 Flow Chart Demonstrating Routes to Diagnosis of Language Disorder (Bishop et al., 2017)

Based on the CATALISE criteria, children should not be just labeled as individuals with language disorder (LD)/DLD until or unless the status of their competency in both languages, especially the best language, in the case of multilingual children, is ascertained. There are higher chances that young EM multilingual children might be noticeably weak in the local/societal language but perform according to developmental expectations when assessed in their heritage language (often the best language in the early years). So, for more accurate identification of LD/DLD in multilingual EM children, it is informative to assess these children in their heritage language (e.g., Urdu in the current target group), especially in younger children whose L2 (e.g., Cantonese in the current target group) may often be too weak to even follow the instructions well in dynamic assessments conducted in L2. Moreover, negative prognostic features are a significant consideration in the CATALISE criteria because these features suggest the likelihood of persistent language difficulties, severe and persistent enough that the child would unlikely grow out of them without the intervention and support from specialists.

The child's age, for instance, is one feature relevant to the prognosis under consideration highlighted in Bishop et al. (2017). Specifically, if the language difficulties are still present at the age of 5 years, then it is likely that the problems will persist (Stothard et al., 1998). Another prognostic feature concerns the range of linguistic domains that are affected. If multiple areas of language functioning are involved, it is also a poor prognostic feature, and the problems are likely to continue into school age (Bishop & Edmundson, 1987). Poor receptive language skills (Beitchman et al., 1996; Clark et al., 2007) and relatively low nonverbal ability (Johnson et al., 2010; Rice & Hoffman, 2015) have also been highlighted as poor prognostic features.

2. The Present Thesis

This thesis focuses on the heritage language development of the Urdu-Cantonese EM multilingual children in Hong Kong. A considerable number of Urdu speakers reside in Hong Kong. The estimated number of Urdu speakers in Hong Kong, mainly from Pakistan and other areas of the world, has risen from 18,094 in 2016 to 24,385 in 2021, constituting 24 percent of the total South Asian people residing in Hong Kong (Census and Statistics Department, Government of the Hong Kong SAR, 2021). The majority of them prefer using their heritage language (Poon, 2010) instead of the official languages of Hong Kong or any other Chinese language. As a result, many children from this EM group are exposed more to their heritage language/L1 (Urdu) than their school or societal language (Cantonese), especially in their early years.

Like other EM children worldwide, these Urdu-Cantonese EM multilingual children acquire both languages under reduced input conditions. This is because their L1 is not supported or reinforced in their school and society. Similarly, their L2 is not supported or reinforced in their homes. Moreover, as these multilingual children grow older and spend more time in the outside world (school and society), they might also experience proportionally less Urdu and more Cantonese. The older children, compared to the younger ones, may experience more incomplete acquisition and language attrition in their L1 due to reduced input.

Considering the role of input conditions in the language acquisition of multilingual EM children, this unique acquisition scenario could bear richly on conceptual issues such as: to what extent would reduced input affect the heritage language/L1 acquisition of these ethnic minority children? And are there certain domains of language that are more or less affected by reduced input?

The other dimension of this Ph.D. project is related to identifying DLD in these EM children. Researchers and professional organizations have emphasized the importance of assessing multilingual children in all their languages (e.g., American Speech-Language-Hearing Association, 2004; Armon-Lotem & de Jong, 2015; Bishop et al., 2017). However, in Hong Kong, conducting language assessments in both/all languages of multilingual children is taxing for clinicians due to similar reasons also highlighted by Boerma and Blom (2017). These reasons involve scarcity of suitable assessment tools, paucity of appropriate multilingual norms, time restrictions, and lack of access to experienced multilingual SLTs and interpreters. In this vein, one primary aim of this project is to mitigate the effects of these long-standing issues by providing clinicians with appropriate assessment tools adapted to L1 of these EM

Urdu-Cantonese multilingual children. In addition, this project also aims to provide reference data (i.e., data from a smaller sample to guide developmental expectations when normative data from the larger sample is not available) encompassing multiple linguistic domains to collect information related to these children's developmental expectations. The rationale is to help clinicians minimize the misdiagnosis of these children.

This research project was approved by the ethics committee at the Hong Kong Polytechnic University. All the parents/caretakers of the participants have given informed written consent to participate in this project. A brief description of each study is presented below:

- **Study 1** documents the linguistic vulnerabilities associated with reduced input in multilingual children acquiring Urdu as their L1 and minority language. It examines a group of EM multilingual children in Hong Kong acquiring Urdu as heritage language and compares their lexical, morphosyntactic, and narrative abilities in their heritage language with their age- gender- and grade-matched multilingual peers from Pakistan. The EM multilingual children acquire L1 Urdu as the minority language in their host country under reduced input. In contrast, the control group develops L1 Urdu as the majority language in their country of origin with more abundant input. The varying linguistic profiles of minority vs. majority groups acquiring the same language in different acquisition contexts are presented in Chapter 2. It also highlights the areas of linguistic strengths and weaknesses in EM multilingual children.
- **Study 2** demonstrates the feasibility of identifying DLD in Urdu-speaking children. It is a case study featuring one six-year-old Urdu-Cantonese multilingual EM child suspected of DLD with reference data for comparisons from seven age-and-grade-matched TD multilinguals. This study demonstrates the feasibility of identifying multilingual children with DLD using the CATALISE diagnostic criteria (Bishop et al., 2017) with support from

the Urdu versions of the different LITMUS assessment tools using remote online testing (Chapter 3).

- **Study 3** examines the clinical utility of the new Urdu Nonword Repetition (NWR) Test in differentiating children with and without DLD. It prioritizes NWR because it is relatively less affected by reduced input than other linguistic domains like vocabulary and grammar (e.g., Roy & Chiat, 2013; Thordardottir & Brandeker, 2013). Three groups of children, including L1 Urdu (minority language) TD EM multilingual children, age-matched L1 Urdu (majority language) TD peers, and age-matched L1 Urdu (majority language) children with DLD were assessed to examine the potential of this new Urdu CL-NWR test in differentiating multilingual TD children with limited language experience and those with genuine language impairment (Chapter 4)

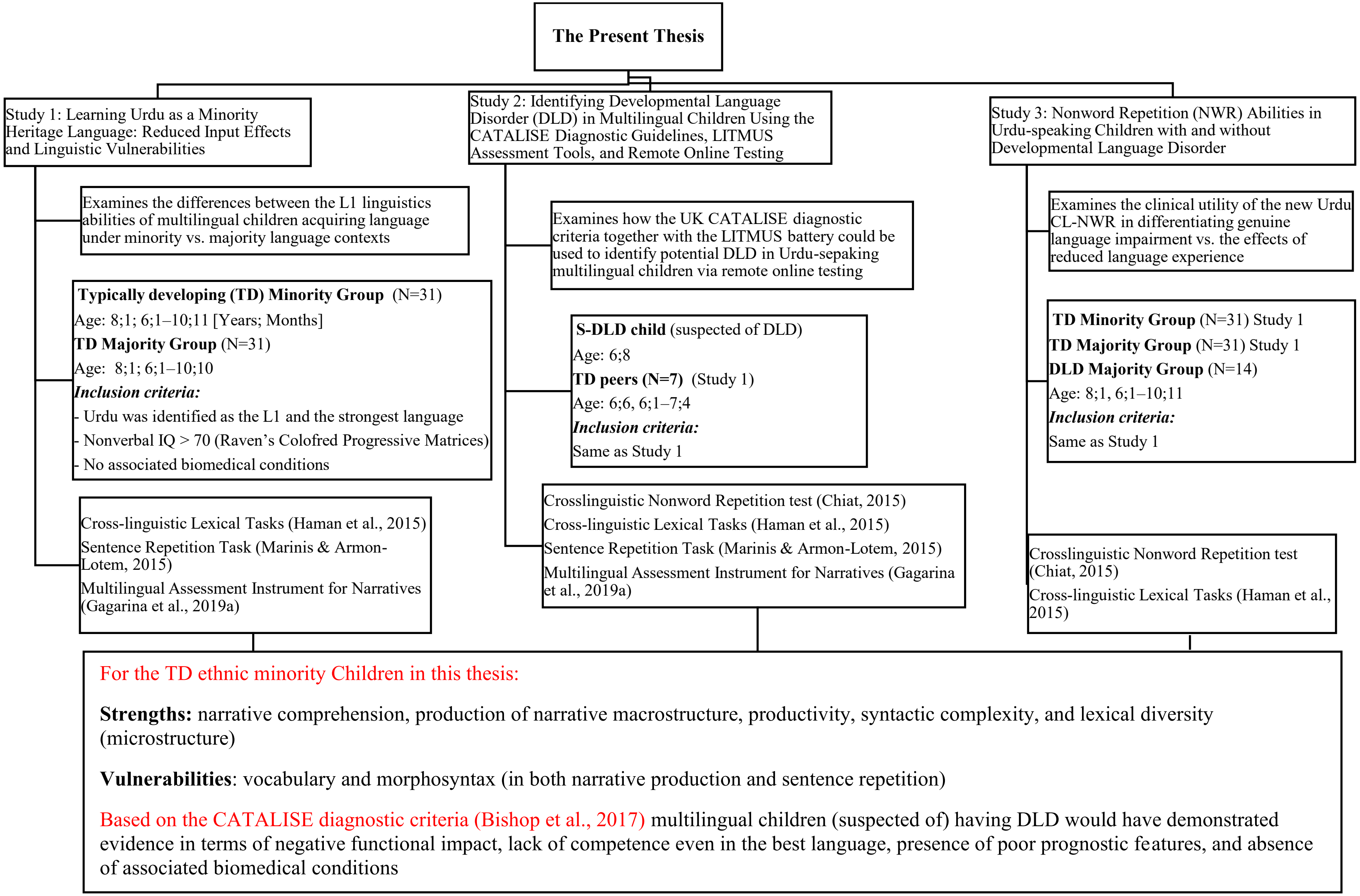

Figure 1.2 Flowchart Presenting Study Design of this Thesis

Chapter 2 - Learning Urdu as a Minority Heritage Language: Reduced Input Effects and Linguistic Vulnerabilities

2.1 Introduction

The number of multilingual children is increasing strikingly worldwide, contributing to the linguistic diversity in the classrooms. This dramatic and rapid increase in the linguistic heterogeneity in classrooms and societies raises the question that to what lengths the conditions of ethnic minority (EM) multilingual children, which inevitably include restricted input to both minority heritage and majority societal languages in comparison to their monolingual or multilingual peers acquiring the same language in a majority context, affects their developmental trajectory in both languages.

Different studies have listed several factors that shape the overall development of both (or all) languages in multilingual children. The list includes the age of acquisition of two/multiple languages (sequential or simultaneous multilinguals), the amount of language input and its use on a daily basis in different contexts, and the socio-political status of languages (minority vs. majority) (Montrul, 2022; Paradis, 2023). Based on these factors, it is suggested that EM multilingual children usually receive limited and divided exposure to both (or all) languages (Bosma & Blom, 2019). Therefore, they often acquire both (or all) languages at a rate slower than their peers acquiring the same language in a majority context (Pearson et al., 1993; Uccelli & Páez, 2007) and usually reach school age with low levels of linguistic abilities in their different languages (Hoff, 2020).

For some multilingual children, one of their languages becomes a minority language (henceforth, L1), which is not spoken frequently outside the home and bears low political, cultural, and educational status in the wider community (Montrul, 2022). It is also recognized as a language that is spoken at home or else is easily accessible by young children but is not prevalent in the larger community (Rothman, 2009). With time, as the EM multilingual children grow older and their exposure to the wider community increases, the L1 becomes weaker. Therefore, EM children exhibit different outcomes and acquisition patterns across language domains compared to their peers acquiring the same language as a majority language. Subsequently, the competence in the majority language keeps on growing to the extent where it congregates or nearly congregates with the level of monolingual peers in the host country, especially once the formal schooling in this language starts (Montrul 2016; Paradis et al., 2021).

Recently, there has been a sharp uptick in the number of studies focusing on language development in multilingual children, and most of these studies have examined the majority language acquisition (e.g., Chondrogianni & Marinis, 2012; Gutiérrez-Clellen et al., 2008; Hoff et al., 2012; Paradis, 2009; Verhoeven et al., 2011). This could be because competence in the majority language is imperative for academic success (Strand et al., 2015; Vettori et al., 2022) and securing jobs in the host countries (Guven & Islam, 2015; Shields & Price, 2002).

Even though maintenance of L1 is also considered central to an individual's well-being (De Houwer, 2015; Yu, 2013) and language preservation at the communal level (Potowski, 2013), studies thoroughly examining the acquisition and maintenance of minority language in the EM multilingual children are rather exiguous (V. C. M. Gathercole & E. M. Thomas, 2009; Winsler et al., 1999). So far, the studies examining L1 development have mostly concentrated on grammatical competence (e.g., Polinsky, 2008; V. C. M. Gathercole & E. M. Thomas, 2009; Hoff et al., 2018; Rothman, 2009, etc.) and intermittently on other linguistic domains, including lexical (e.g., Haman et al., 2017; Montrul 2010) and narrative (e.g., Haman et al., 2017; Rodina, 2017) abilities. Furthermore, only a limited number of L1-based studies focused on the language development process in multilingual children (e.g., Montrul, 2008; Haman et al., 2017; Polinsky, 2011), but instead on its outcomes in adulthood (e.g., Rothman, 2009).

In this vein, the present study is an attempt to address this significant gap in the field by examining performance across multiple linguistic domains in the same group of the typically developing (TD) EM multilingual children in Hong Kong acquiring L1 (Urdu) under reduced input conditions. Their performance was also compared with a control group from Pakistan (Urdu is Pakistan's national and official language) comprising multilingual majority peers acquiring Urdu in relatively abundant conditions. The aim was to identify the areas of linguistic strengths and vulnerabilities in these EM multilingual children by comparing their language abilities with control peers from Pakistan.

This chapter first presents cross-linguistic literature on L1 competence in EM multilingual children. The subsequent sections then elaborate on methodology, results, and discussion. The chapter ends with the conclusion of the findings and their implications.

2.1.1 Acquiring L1 as a Minority Heritage Language

The evidence from studies examining the majority language (henceforth, L2) development suggests that multilingual children lag behind their monolingual peers in various language domains, and sometimes their performance is often comparable to the monolinguals with language impairments (Ebert & Kohnert, 2016; Kohnert et al., 2006). On the other hand, the studies focusing on the L1 of EM bilingual children offer inconclusive findings. Some studies (e.g., Fabiano-Smith & Barlow, 2010; Haman et al., 2017; Rodina, 2017; Thordardottir & Brandeker, 2013; Uccelli & Páez, 2007) suggest a performance gap between multilinguals (acquiring L1 in minority context) and their monolingual counterparts (acquiring L1 in majority context). Whereas others conclude that children from these groups achieve their milestones for phonological (Fabiano-Smith & Barlow, 2010), lexical (Hoff et al., 2012; Pearson et al., 1993), morphosyntactic (De Houwer, 2005; Paradis, 2009), and narrative macrostructure development (Paradis & Kirova, 2014; Kunnari et al., 2016) roughly around the same age.

This demonstrates that there are both differences and similarities between multilingual and monolingual children's developmental trajectories. Studies have shown that multilingual development is idiosyncratic, and therefore, multilinguals should not be benchmarked against monolingual norms (Armon-Lotem, 2018; V. C. M. Gathercole, 2013). Following this, the present study also attempts to yield some baseline/reference data for Urdu-speaking multilingual children. The intention is to inform the professionals regarding the developmental expectations of these TD EM children.

The forthcoming sections will review the literature presenting findings from three different language domains including lexicon, morphosyntax, and narrative. The impact of age, language status, amount of language input, and some other domain-specific factors will be discussed briefly.

2.1.2 Lexical Abilities

Vocabulary is among the first language faculties to develop, and its growth continues throughout one's life (MacLeod et al., 2019). Children mainly learn words from their social environment. Since language exposure in multilingual children is generally considered to be divided between two/multiple systems, they, in comparison to their monolingual peers, are said to have smaller lexicons (in one language). This often paints an incorrect picture, leading to misdiagnosis of Developmental Language Disorder (DLD) in these children. Thus, a precise lexical assessment taking into account both languages and especially L1 (usually the dominant language) in the young EM multilingual children is critical to segregate TD children with balanced language development (similar lexicon in both languages, could be smaller in one language than monolinguals but in line with multilingual peers), unbalanced language development (markedly smaller lexicon in one language, but concurs with multilingual patterns in the other one) and suppressed lexical abilities in both languages (signifying peril for DLD) (Simonsen & Haman, 2017).

As far as age is concerned, one keeps learning new words from infancy until old age. Various studies have suggested vocabulary gains with age in multilingual children's L2 (e.g., Gagarina et al., 2014; V. C. M. Gathercole et al., 2013). On the contrary, the association between lexical abilities in L1 and age is unclear. Studies have reported inconsistent findings, where some found clear lexical gains (e.g., Bohnacker et al., 2021) with age and others did not (e.g., Ganuza & Hedman, 2019; Öztekin, 2019). For instance, in a similar study by Gagarina et al. (2014), the children who showed vocabulary gains in their L2 with age displayed a flat growth curve or very slight gains in their expressive lexical abilities between ages 4 to 7 years in their L1. In line with this finding, V. C. M. Gathercole and colleagues (2013) also found that Welsh-English children, compared to their L2, exhibited less age-related progress in their L1 expressive vocabulary.

Similar findings of no vocabulary gains with age in L1 were also reported by Bohnacker et al. (2016) and Lindgren and Bohnacker (2020). The possible explanation for this ambiguous association between age and L1 lexical development could be ascribed to the strong influence of another factor: the amount of input in the home language (Bohnacker et al., 2016). This implies that when EM multilingual children start preschool, the exposure to L2/societal language increases and alters the proportion of L1, which could stagnate their lexical abilities in L1 and even annihilate the association between age and L1 lexical development (Boeschoten & Verhoeven, 1986; V. C. M. Gathercole & E. M. Thomas, 2009; Hoff et al., 2014; Sheng et al., 2011).

Language input conditions associated with language status (minority vs. majority) are suggested to affect differences between multilingual and monolingual lexical abilities in a target language. In a study, Haman and colleagues (2017) compared the receptive and expressive lexical abilities of the Polish-English multilingual children acquiring L1 (Polish) in the United Kingdom as a minority language and their Polish-speaking monolinguals learning L1 in a majority context. Findings showed that multilingual children obtained significantly lower receptive and expressive lexical scores than their monolingual counterparts. Some other studies (e.g., Pearson et al., 1997; Miękisz et al., 2017) also reported a similar pattern of findings.

The amount of language input is also reported to affect lexical performance in children. In a study based on Welsh-English-speaking children, V. C. M. Gathercole and colleagues (2013) and V. C. M. Gathercole and E. M. Thomas (2009) examined the receptive lexical skills of Welsh-English children in their L1 (Welsh). Parents were divided into three groups based on their language(s) use with children (who speak mostly/only Welsh, who use both, and who speak mainly in English). The children from the group whose parents mainly used Welsh outperformed all other groups. Similar results were also found for expressive lexical skills by Bohnacker et al. (2016), Dijkstra et al. (2016), and Prevoo et al. (2014), who assessed expressive lexical skills in the L1 of Turkish/German-Swedish, Turkish-Dutch, and Frisian-Dutch children, respectively. Their findings suggest that children's language input conditions might directly relate to vocabulary development in EM multilingual children.

Moreover, the association between maternal education and L1 lexical skills has been explored in different studies. The literature presents inconclusive findings. Paradis and colleagues (2020) and Prevoo et al. (2014) found a significant and positive correlation between maternal education and lexical skills in L1. Some studies have reported a non-significant association between maternal education and L1 vocabulary skills in children (e.g., Friend et al., 2022; Hamann et al., 2020).

Comprehension vs. Production. Comprehension of lexical items is recognized as a "purer" measure because it examines the lexical knowledge without the impediments of lexical retrieval and pronunciation problems (Clark, 2009). Children often struggle with these impediments during production tasks, resulting in poorer production scores than comprehension (Goldfield, 2000). A child's poor performance in production tasks could be connected with lexical access challenges and not with the attenuation of semantic knowledge, advocating the importance of assessing comprehension in addition to production (Bialystok et al., 2010). Hence, evaluating both modalities becomes even more critical in multilingual children as the gap between their receptive and expressive vocabulary is usually more pronounced (Altman et al., 2017).

Usually, multilingual children underperform their monolingual counterparts at both receptive and expressive levels when assessed in either their L1 (Haman et al., 2017) or L2 (Altman et al., 2017). However, some studies have demonstrated group effects of varying magnitudes within the two modalities. For example, Gross and colleagues (2014) reported that Spanish-English sequential and simultaneous multilingual children aged 5 to 7 years scored significantly lower than their monolingual peers in receptive and expressive lexical tasks in English. When scores from both L1 and L2 were considered, the difference between simultaneous multilinguals and their monolingual peers became non-significant for receptive but not for productive vocabulary. On the other hand, the difference between sequential multilinguals and monolinguals remained significant in both modalities. This gap in scores based on modalities underlines the significance of examining lexical abilities at both levels.

Nouns vs. Verbs. In addition to modalities (comprehension vs. production), different findings for word categories are also reported. Nouns, in general, are suggested to be easier to acquire than verbs. The size of this noun vs. verb gap varies across languages. For example, this gap was more pronounced in German and English than in Turkish and Korean (Gopnik & Choi, 1995; Kauschke et al., 2007). Haman and colleagues (2017), in a macroscale study assessing monolingual children aged 3 to 6;11 years in 17 languages, found that children who were speakers of Catalan, English (British), English (South African), Finnish, German, Hebrew, Italian, Lithuanian, Luxembourgish, Polish, Serbian, Slovak, Turkish (13 out of 17 languages) performed significantly worse in verbs than nouns. On the contrary, this word category effect was insignificant for Afrikaans, Norwegian, and Swedish-speaking children. In a different study, Lindgren and Bohnacker (2020) also compared the noun vs. verb performance in both L1 (German) and L2 (Swedish) languages of German-Swedish 4-to-6-year-olds. For both languages, performance in nouns was better than in verbs.

Better performance in nouns is generally accredited to nouns being "easier," as they are deemed to be either less dependent on context or are conceptually/semantically less complex or abstract and more meaningful, imageable, and unambiguous than verbs (Altman et al., 2017). Furthermore, this better performance in nouns could also be explained in terms of task effect. In picture-based tasks, displaying objects (nouns) via static pictures is more straightforward than actions/events (Lindgren & Bohnacker, 2020).

2.1.3 Morphosyntactic Abilities

Compared to the development of other linguistic domains such as vocabulary and narratives, multilingual children have been reported to struggle more with grammar development (e.g., V. C. M. Gathercole, 2007; Paradis & Kirova, 2014; Schwartz et al., 2015). It also implies that this linguistic domain is also the most sensitive to language exposure/input (Paradis, 2016) and more prone to incomplete acquisition (Polinsky, 2016; Montrul, 2008) or even attrition (Flores, 2012; Montrul, 2008) when the exposure to L2 societal language increases and that to L1 decreases with age.

There are mixed results from studies examining morphosyntactic development in multilingual children. Some studies (e.g., Conboy & Thal, 2006; De Houwer, 2005; Genesee & Nicoladis, 2007; Meir et al., 2016) concluded that multilingual and monolingual children acquire some syntactic structures around similar ages. In contrast to this claim, several studies have suggested that multilinguals diverge from monolingual patterns (Antonijevic-Elliott et al., 2020; Hoff et al., 2012; Komeili & Marshall, 2013).

Studies assessing L1 morphosyntactic abilities revealed significantly lower performance by EM multilingual children than their monolingual counterparts (Thordardottir & Brandeker, 2013). Studies so far have pointed out different areas of L1 grammar, including overuse of fixed word order patterns (Isurin & Ivanova-Sullivan, 2008), agreement morphology (Bolonyai, 2007; V. C. M. Gathercole & E. M. Thomas, 2009; Polinsky, 2008), or interpretation and application of long-distance binding (Kim et al., 2009) to be challenging even for older multilingual children.

Acquiring L1 as a minority heritage language has been found to negatively impact morphosyntactic abilities in EM multilingual children. In a study, Haman and colleagues (2017) assessed the grammatical skills in the L1 Polish of Polish-English multilingual children compared to their monolingual counterparts. The findings showed that multilingual children performed significantly worse than monolinguals. In a different study, Meir et al. (2016) also found significant differences in the multilingual-monolingual preschool children's performance on a grammatical task in L1 (Russian) of Russian-Hebrew speaking children, and the monolinguals outperformed the multilinguals. Both studies concluded that the worse performance of multilinguals in L1 might be highly related to constrained input conditions in the target minority heritage language.

Some studies have also explored the influence of the quantity of language input on EM multilingual children's grammatical development. The results indicated the amount of altered or impoverished input as one of the leading causes of poor performance on L1 morphosyntactic tasks in EM multilingual children (Benmamoun et al., 2013; V. C. M. Gathercole & E. M. Thomas, 2009; Hoff et al., 2018; Rothman, 2009). Lastly, studies investigating the impact of maternal education on the morphosyntactic competence of young minority heritage speakers have reported inconsistent findings. For example, Paradis and colleagues (2020) stated positive significant associations, while Antonijevic-Elliott et al. (2020) and Hamann and colleagues (2020) reported a non-significant correlation between maternal education and EM multilingual children's L1 morphosyntactic abilities.

2.1.4 Narrative Abilities

Narratives are of interest to both clinicians and researchers. A narrative, in the words of Bruner (1986), is an oral sequence of imaginary or real events. Narrative competence, the capability to comprehend and produce stories, is an intricate skill set that requires encoding and understanding information and then formatting that acquired information in a methodical mental representation (Levelt, 1989). As narratives involve incorporating different linguistic, cognitive, and social abilities (Liles, 1993), they can provide robust information about an individual's language development in an ecologically valid manner (Botting, 2002). Moreover, narratives are also claimed to be a less biased approach for measuring language competence compared to other norm-referenced tests as their structural features are shared across languages (Paradis, 2010), making it an invaluable tool for measuring language abilities in different acquisition contexts.

Narrative Comprehension. Narrative comprehension is considered a precondition for constructing well-formed narratives (Shapiro & Hudson,1991; Trabasso & Rodkin, 1994). If children cannot interpret plotlines, cause-effect associations, thoughts, intentions, and emotions of protagonists, they will not be able to relay this information to the listeners during narrative production (Burris & Brown, 2014; Stein & Glenn, 1979). Furthermore, narrative comprehension entails interconnected cognitive abilities that permit a child to construct a coherent narrative representation. As these abilities also play a crucial part in reading and text interpretation, they are usually considered rudimentary for acquiring reading skills (Lynch et al., 2008; Paris & Paris, 2003).

Previous studies have explored the impact of various factors on children's narrative comprehension abilities. In their work, Bohnacker and Gagarina (2020) reported the age effects by compiling the results of different studies, including monolingual and multilingual children from different language backgrounds aged 2;10 to 9;9 [years; months]. Authors reported a precipitous and more substantial development in comprehension abilities between ages 3;0 to 5;6 after which the increase in comprehension scores started to taper off. The results suggested that the response accuracy increased from 20 to 60–70% between ages 3;0 to 5;6. There was also a marginal increase between ages 5;6 to 6;6 (70–90%). Significant improvements at the group level were not evident after age 6;6 (or 7). The mean rate of progress for children aged between 5;0, 6;0, and 9;0 was mainly between 70 and 90%. It was concluded that if one considers 60–70% as the criterion for "acquired," then narrative comprehension can be regarded as acquired mainly around age 5;0, but it is still not close to the ceiling yet.

In a different study, Rodina (2017) examined the effect of language status and compared the L1 narrative comprehension competence of the EM Norwegian-Russian-speaking children (mean age = 4;6) and their Norwegian-speaking monolingual counterparts (mean age = 4;5). Findings illustrated that children acquiring the same language in majority vs. minority contexts exhibited comparable performances during narrative comprehension tasks. In another study, Roch and Hržica (2020) explored the influence of the quantity of language input and maternal education (taken as a proxy of socioeconomic status) on narrative comprehension skills.

Findings showed that narrative comprehension was immune to the amount of language input and maternal education effects in both L1 and L2.

Narrative Production. This section presents findings related to the production of macrostructure, microstructure, and internal state terms (ISTs).

Macro-and-Microstructure Abilities. Narrative productions can be measured at both macro-and-microstructural levels. Macrostructure analysis is the central feature of narrative competence and is related to the higher-order structural organization of narratives (Heilmann et al., 2010; Trabasso & Nickels, 1992). Depending on the framework, there are different methods of conducting macrostructure analyses, like the six-level framework by Applebee (1978), high point analysis (Labov, 1972; Peterson & McCabe, 1983), and episodic analysis (Stein & Glenn, 1979). The episodic analysis/story grammar model (Stein & Glenn, 1979) is the most frequently used method, and it evaluates narratives based on the inclusion of different story components (e.g., setting, initiating event, internal response, internal plan, attempt, consequence, and reaction). Studies have used different terminologies for macrostructure analyses like event content, story content, structural complexity, and story structure. The current study uses the terms story structure and story complexity.

Story structure (SS) follows episodic analyses by evaluating the occurrence of story grammar components. As the events and intentions represented by these story grammar components include causal and temporal associations, incorporating more of these grammar components contributes to the richness and coherence of a narrative. Hence, it has both quantitative (number of story grammar elements included) and qualitative (in terms of coherence and richness of narrative content) dimensions (Chan et al., 2023).

Structural complexity (SC) refers to the production of complete (or full) episodes, resulting in a well-formed narrative (Gagarina et al., 2015; Stein & Glenn, 1979; Westby, 2012). A complete or full episode should include a goal, attempt, and outcome sequence. A goal is what the protagonist(s) intend to achieve in a narrative and is operationally defined as the protagonist's reaction/response to an initiating event. There could be numerous protagonists with various goals in a narrative. The attempt is the protagonist's effort to obtain that goal, and the outcome reflects whether the protagonist could obtain the goal. There are different kinds of narrative sequences or episodic structures of varying macrostructural complexity. Based on Westby's (2012) decision tree, there are three main types of episodic structures: 1) attempt-outcome sequences (action/reaction sequence, AO); 2) goal-attempt/outcome sequences (these are the incomplete episodes including a goal and an attempt or an outcome, GA/GO); 3) full/complete episodes (includes goal, attempt, and outcome, GAO).

Microstructural analyses, on the other hand, involve more language-specific analyses of internal linguistic structures like morphosyntax, lexical items, and the use of connectives in forming a coherent story in a target language (Gagarina et al., 2016). It covers a broad array of linguistic features, like productivity (a general measure of narrative length, e.g., the total number of communication units or words), lexical diversity (e.g., number of different words), grammaticality (e.g., the proportion of grammatical/ungrammatical utterances, verb accuracy), syntactic complexity (e.g., mean length of word or communication units, the proportion of complex utterances), discourse cohesion, and exclusively multilingual phenomena like code interference and code-switching. As microstructure components are suggested to be language-specific, it is thus predictable that some would differ cross-linguistically (Gagarina et al., 2016; Rodina, 2017; Simon-Cereijido & Gutiérrez-Clellen, 2009).

Different studies (e.g., Squires et al., 2014; Uccelli & Paez, 2007) have suggested that narrative abilities develop between preschool and school age. These longitudinal studies have shown that multilingual children from kindergarten (KG) and first grade like their monolingual peers (acquiring the target language in a majority context), follow similar developmental trajectories for macrostructure. Gagarina (2016) also investigated age-related development in the narrative macrostructure. The Russian-German multilinguals in this study were divided into three age groups, and the results revealed significant developmental growth between preschool (mean age = 3;7) and first grade (mean age = 7;0) in both languages. No significant improvements were found between first and third grade (mean age = 9;2). The same patterns of gains with age were noted between these three groups for story length as well.

There are also studies (e.g., Bamberg & Damrad-Frye, 1991; Trabasso & Rodkin, 1994) that have concluded that the narrative macrostructural competence continues to develop during primary school years, and even at the age of 9;0, narratives are not fully adult-like. Moreover, Bishop and Donlan (2005) and Van den Broek (1997) reported an increase in the number of key story elements in the oral narratives of children with age, with the ability to convey a protagonist's internal states or reactions fully attained around the age of 10. These contrasting findings indicate that narrative macrostructure competence could vary greatly across languages and children.

On the other hand, the age-related progress in some aspects of microstructure such as lexical diversity was reported in both languages of Spanish-English children attending KG (mean age = 5;11) and third grade (mean age = 9;1) by Miller et al. (2006). This is in line with Blankenstijn and Scheper's (2003) finding suggesting that the development of microstructure in TD children is protracted and is still ongoing after age 10.

Different studies have also compared the macro vs. microstructural performance of EM multilingual children and their monolingual peers acquiring the same language in a majority context. Narrative macrostructure in comparison to microstructure is concluded to be more language-independent and less sensitive to language input conditions. Hence, multilingual children, compared to their monolingual peers, are not disadvantaged during the production of macrostructure (Kunnari et al., 2016; Haman et al., 2017). Studies examining performance in L1 have also reported similar findings. For example, Rodina (2017) investigated the L1 narrative production abilities in preschool Norwegian-Russian multilinguals and their Norwegian monolingual peers. Children from both minority (EM multilinguals) and majority (monolinguals) groups obtained comparable macrostructure scores. However, the EM multilingual group exhibited significantly poor performance for some microstructure measures such as narrative length and grammaticality.

In a longitudinal study, Gámez and González (2019) assessed the narrative skills of Spanish-English multilinguals (mean age = 5;6) in comparison to their Spanish-speaking monolingual peers (mean age = 5;7) at two-time points in a retelling task. The findings for the L1 (Spanish) revealed that the difference in their story structure scores (macrostructure) was statistically non-significant. On the other hand, the two groups differed significantly in their microstructural skills (grammaticality), and the monolinguals outperformed the multilinguals at both time points.

This discrepancy in macro vs. microstructural abilities of multilingual children has been explained in terms of cross-linguistic transfer of knowledge. Studies (e.g., Pearson 2002, Rodina, 2017; Squires et al., 2014; Uccelli & Paez 2007) that focused on both languages of multilinguals have shown that cross-linguistic transfer was found only at the macrostructural level and not at the microstructure level. This transfer of knowledge allows EM multilinguals to use story grammar knowledge acquired in their L1 to their L2 or vice versa, putting them at an advantage, so that they are able to perform at par with their monolingual majority peers.

The impact of the quantity of language input has also been investigated for both macro- and- microstructural abilities. Several studies (e.g., Fiestas & Peña, 2004; Haman et al., 2017; Gagarina, 2016) have suggested that the amount of language input might not be crucial in developing macrostructure competence. So, regardless of input quantity, narrative structures are invariant across multilinguals' two languages, and multilingual children could produce comparable story structures in both/all languages. While microstructure skills, on the contrary, are claimed to be vulnerable to input quantity, as children with limited language input were found to obtain significantly low scores across various microstructure domains (Bonifacci et al., 2018; Hipfner-Boucher et al., 2014).

Maternal education is also reported to effect macro-and-microstructural abilities differently. Silva and colleagues (2014) reported a non-significant effect on macrostructure skills. Whereas Dollaghan et al. (1999) reported a positive influence of maternal education on some microstructural elements including syntactic complexity, lexical diversity, and narrative length. Jia and Paradis (2015) also reported a positive association between higher maternal education and narrative retelling in L1.

Internal State Terms (ISTs). Examining the use and types of ISTs is another dimension considered during narrative assessment. Studies have used various terminologies for ISTs, including mental state terms (Altman et al., 2016), internal states (Miller & Aloise, 1989), mental state language (Bartsch & Wellman, 1995), evaluations and inferences (Burns et al., 2012). This study uses the term ISTs.

ISTs explicitly refer to the internal states of protagonists such as their thoughts and perceptions (Labov, 1972; 2013). The use of ISTs entails the comprehension of the order of events and the interpretation of the character's emotions and intentions. Forming mental representations of story characters for a listener demands the narrators to address characters' internal states related to what they think, want, say, or intend (Burns et al., 2012). The use of ISTs is important for gaining coherence in stories. A well-formed narrative with an explicit beginning, middle, and end might lack coherence if evaluative devices like ISTs are not included (Berman & Katzenberger, 2004).

The use of ISTs is suggested to develop with age. Some studies (e.g., Bohnacker et al., 2016; Roch et al., 2016;) that examined the use of ISTs in multilingual children aged between 5;0 and 7;0 found significant age effects, and that the older children used more ISTs than the younger ones. Manhardt and Rescorla (2002) also suggested that children start using the ISTs of protagonists in their narratives around the age of 4, which continues to develop until they are 9.

Different studies have also compared the use of ISTs in multilingual and monolingual children (acquiring the same language in minority vs. majority contexts). In a study, Tsimpli and colleagues (2016) along with other measures, assessed the use of ISTs in Greek-speaking multilingual (mean age = 9;1) and monolingual (mean age = 9;0) children in a retelling condition. The multilingual children made significantly higher use of emotional and mental (+ToM related) ISTs in their narrative productions than their monolingual counterparts. In contrast, the use of physiological, linguistic, and perceptual (-ToM related) ISTs, on the other hand, was comparable in the two language groups. Some studies, for example, Boerma and colleagues (2016), Bonifacci et al. (2018), Chen and Yan (2011), and Kunnari et al. (2016) also reported no differences in the use of ISTs in children acquiring target language in reduced vs. abundant conditions.

Narrative Telling vs. Retelling. Narrative samples can be obtained using two elicitation modes, telling or retelling. Different studies have tried to compare the narrative competence in multilingual children across two elicitation modes. For example, Kunnari et al. (2016) conducted a study to assess the narrative abilities of Finnish-Swedish multilinguals, and they found that the multilinguals achieved higher scores during the retelling condition in their L1

(Finnish) but for their L2 (Swedish). There were no differences in the narrative scores obtained via retelling or telling. Roch and colleagues (2016) examined narrative competence in Italian-English multilinguals, and they found that irrespective of language, the performance for comprehension and production was better during the retelling condition. Lastly, Maviş et al. (2016) and Otwinowska and colleagues (2018) also reported differences in the comprehension and production scores obtained from retelling or telling. The performance in retelling was reported to be better.

All the above-mentioned studies employed the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2019a), and following its design, they used Cat/Dog story for retelling and Baby Birds/Baby Goats for telling (MAIN's design has been detailed under the Methods section). Although Cat/Dog and Baby Birds/Baby Goats stories are structurally comparable, it still needs to be verified whether they are equally challenging for the children regardless of elicitation condition (Lindgren, 2018). The present study also used MAIN, and the differences in the narrative abilities based on elicitation modes were examined in a more controlled fashion, that is, each child participant depending on the counterbalance order, produced two stories in both retelling and telling conditions (i.e., with and without support from a model story). This allowed to inspect that: 1) whether there are any differences in narrative outcomes based on elicitation mode type; and 2) if there are differences in scores, is it due to the type of elicitation mode, or it is that children found Cat/Dog relatively easier than Baby Birds/Baby Goats content.

2.2 Present Study

A dramatic increase in the rate of multilingualism worldwide directs our attention towards the substantial need for appropriate assessment tools and norms suitable to assess multilingual children in all languages. This paucity of suitable assessment tools and appropriate multilingual norms adds to the long-standing challenging issue of disentangling the effects of bilingualism from language impairment (Armon-Lotem, 2018). To deal with this ordeal, ample information regarding the unique attributes of multilingual learners' language development must be obtained through comprehensive assessments that target different facets of their language acquisition and highlight their linguistic strengths and vulnerabilities.

Although the literature shows that a good number of studies have been devoted to multilingual language acquisition, many aspects related to this topic still remain unclear. Many studies have focused on the L2 acquisition of multilinguals, and only a few have examined L1 performance (e.g., Haman et al., 2017; Thordardottir & Brandeker, 2013). Furthermore, just a small number of research studies have examined multiple language measures in the same group of participants (e.g., Haman et al., 2017; Uccelli & Páez, 2007). Hence, there is a definite need for studies that will provide a comprehensive overview of differences in the linguistic performances of multilingual children from different language acquisition contexts.

The present study is an attempt to address this gap by assessing Urdu-Cantonese EM multilingual children in their L1 (Urdu) in multiple linguistic domains. In order to identify EM multilingual children's areas of linguistic strengths and vulnerabilities, their performance was also be evaluated against age gender, and grade-matched Urdu-speaking majority peers (Maj) from Pakistan (acquiring Urdu under relatively abundant input conditions) at both expressive and receptive levels. Studies focusing on multiple linguistic domains of Urdu-speaking children whether acquiring it under limited (as a minority/heritage language) or even abundant (as a majority language) conditions are scarce. To my knowledge, this study is the first one that aims to examine lexical, morphological, and narrative abilities in the same groups of multilingual children learning L1 Urdu in different (minority vs. majority) acquisition contexts. In addition to factors like language status (minority vs. majority), the present study aims to study the effects of some other participant-related factors including age, language input, and maternal education. The impact of some domain-specific factors will also be explored.

To assess the multiple linguistic abilities of the Urdu-Cantonese EM children and their Urdu-speaking peers from Pakistan, different assessment tools from the LITMUS battery developed under the European COST Action were adapted into Urdu. This contributed a set of appropriate assessment tools for use by the professionals working with Urdu-speaking children.

2.2.1 Research Question

This study addresses the following research questions:

- How does reduced input affect competence in different linguistic domains when children acquire the first/heritage language (L1) in a minority context?:
    - effects of language status (minority vs. majority)
    - effects of input quantity
    - a range of outcome measures (lexical, morphosyntactic, and narrative competence)

2.2.2 General Predictions

The following predictions were made based on the literature:

- Reduced input might affect certain linguistic abilities more than other linguistic abilities.
- Reduced input associated with minority language acquisition context might lead to more restricted age-related progress in the first/heritage language (L1 attrition).

2.2.3 Domain-specific Predictions

Lexical Abilities

- The performance in comprehension might be better than production (Altman et al., 2017; Gross et al, 2014), and the EM children may perform significantly worse than their control peers on both tasks.

- Irrespective of language status, the performance in nouns might be better than in verbs (Haman et al., 2017; Lindgren & Bohnacker, 2020).

Morphosyntactic Abilities

- As morphosyntactic abilities are considered to be more language-specific and sensitive to language input, the EM children might perform worse than their control group (Antonijevic-Elliott et al., 2020; Haman et al, 2017; Thordardottir & Brandeker, 2013).

Narrative Abilities

- Both EM and Maj groups might obtain comparable scores in narrative comprehension tasks (Bohnacker, 2016; Rodina, 2017).
- Based on the language-independent nature of the narrative macrostructure, there might not be significant differences in the scores of the two language groups (Gagarina, 2016; Haman et al., 2017; Kunnari et al., 2016).
- As narrative microstructure abilities are claimed to be more language-specific, there might be differences in the performance of the two language groups (Hipfner-Boucher et al., 2014; Pearson, 2002; Rodina, 2017).
- There might be no significant differences in the use of ISTs in children acquiring target language under minority vs. majority contexts (Kunnari et al., 2016).
- Based on the lack of support from the model story during telling tasks, the performance during retelling might be better than telling (Maviş et al., 2016; Otwinowska et al., 2018; Roch et al., 2016).

2.3 Methods

2.3.1 Participants

A total of 62 primary school children were recruited. Table 2.1 presents the demographic characteristics of the study participants. Out of 62, 31 were Urdu-Cantonese EM multilingual children from Hong Kong acquiring L1 Urdu under reduced input conditions (mean age = 8;1, range = 6;1–10;11), and 31 were their age- gender- and grade- matched Urdu-speaking peers from Pakistan (Maj; control group) acquiring Urdu under relatively more abundant input condition (mean age = 8;1, range = 6;1–10;10). The control group was included to test different predictions concerning the disparity in performance between the two language groups. These children were further divided into older and younger age groups. Informed written parental consent was obtained before testing.

An adapted version of the Parents of Bilingual Children Questionnaire (LITMUS-PaBiQ; Tuller, 2015) was used to obtain information related to the child participants' L1 input, usage and frequency of exposure to different languages on a daily basis, developmental milestones, presence/absence of any significant biomedical conditions, speech, language, hearing difficulties, and family history of speech and/or language impairment, parental education and working status, child and parents' receptive and expressive competence for different languages. There were also some questions related to language(s) exposure (e.g., the age at which regular/intensive exposure to L2 started and estimated time and percentage that children spend in different language environments). Urdu was indicated as the best/strongest language of children in both groups.

In addition to a categorical variable, language status (EM vs. Maj), Urdu input/exposure (mean of both school and non-school days) was also included in the statistical models as a continuous variable to obtain insights regarding the effects of language exposure on language performance across multiple linguistic domains. Results are reported for this predictor variable separately in addition to language status. For the amount of Urdu input, mothers were asked to provide the estimated number of Urdu input hours in school and non-school days. Independent samples T-test showed that there were significant differences in the amount of Urdu input received by both groups (t (246) = 15.27, p = < .0001). The Maj group (mean = 101 hours/week;

Range = 54–111) received significantly more Urdu input than the EM group (mean = 66 hours/week; Range = 25–106), indicating that the EM group indeed received an overall restricted L1 input as compared to their control peers from Pakistan.

Table 2.1 *Demographic Characteristics of the Participants*

	Ethnic Minority (N = 31) (Mean; Range)		Majority/Controls (N = 31) (Mean; Range)		Group Difference
Age Groups [Years; Months]	Younger (N=15)	Older (N=16)	Younger (N=15)	Older (N=16)	
	6;9; 6;1–7;7	9;2;8;1–10;11	6;9; 6;1– 7;7	9;2; 8;1–10;10	
Urdu Input [Hours/week]	66; 25–106		101; 54–111		t (246) = 15.27, p = < .0001)

2.3.2 Study Design

In addition to language status, age, and amount of Urdu input, maternal education was also included in the statistical models as a control variable. Maternal education is implied to be a strong environmental factor as it indexes higher quality and quantity of language input/exposure to children as well as cultural capital and overall family environment (Prevoo et al., 2014). Thus far, research on the impact of maternal education on various language domains has yielded contradictory results in the literature, particularly with regard to multilingual children's L1.

In the current study, maternal education, the mothers were asked to provide the number of total education years. Independent samples T-test was conducted to compare the maternal education level in both groups with different language statuses (EM vs. Maj). The results showed that levels of maternal education (total years of education) in the EM group (mean =

11.04; Range = 5–17) and Maj group (mean = 11.64; Range = 0–16) did not differ significantly (t (246) = 1.23, p = .219).

Next, the impact of some domain-specific factors, namely modality (comprehension vs. production) and word category (nouns vs. verbs) for lexicon, sentence structure for morphosyntax, and elicitation mode (retelling vs. telling) for narrative was also investigated.

2.3.3 Material, Tasks, and Procedures

Remote Online Testing. All the testing was moved online due to the COVID-19 pandemic. The digital versions for all assessment measures prepared using Microsoft PowerPoint were used. The general principle was to follow the conventional face-to-face testing procedures to the maximum extent possible while making particular necessary adaptations so that the tests could be administered online in a controlled manner to maximize the quality of the online testing environment/conditions.

The 'Zoom' software was used to conduct online testing. Screens were shared using its 'Share Screen' function. Some essential adaptations were done carefully to suit online testing. For example, the participants are asked to point to the correct pictures during the face-to-face administration of noun and verb comprehension sub-tasks in the Cross-linguistic Lexical Tasks. The picture numbers were added in the digital versions of this test to make the remote online procedure simple and easy for both the experimenter and the experimentee. The participants were then asked to tell the numbers of the pictures instead of pointing.

All participants used the same headsets with a microphone to control the audio quality. These headsets were provided free of charge and sent via mail before the testing sessions. The parents/caregivers were guided to turn on the highest volume level of the standard headsets and ensure that the speech sounds were clear on their ends. Some practice trials were also administered before testing, where the participants were asked to repeat some easy syllable strings to obtain some objective evidence that they could hear the audio stimuli clearly. The experimenter also used the same microphone for all testing sessions.

To participate in the online sessions, the participants were required to use a computer, laptop, or tablet (cell phones were not allowed) in a quiet environment with stable Wi-Fi. They were asked to switch on their cameras during the testing sessions to confirm that the participants were only accompanied by a parent/caregiver. The parents/caregivers were requested not to interfere, disturb, or train the participants during the testing session.

Assessment Measures. Different assessment measures developed under the European COST Action's LITMUS battery were adapted into Urdu to assess multiple language domains. The information on these assessment measures, adaptations, and administration is detailed in the forthcoming sections.

Cross-linguistic Lexical Tasks (CLT; Haman et al., 2015). The CLT includes four sub-tasks that focus on the comprehension and production of nouns and verbs. It specifies the level of receptive and expressive vocabulary size (Haman et al., 2015). So far, the CLT has been adapted into 34 languages. The Urdu adaptation was completed using the adaptation guidelines. The selection of target stimuli was based on the following steps:

Phase 1 was the naming task. Words suitable for the Urdu-speaking children, especially for the ones living in Pakistan, were selected by involving 25 native Urdu-speaking adults from Pakistan. This task was designed using the OpenSesame software (version 3.0; Mathôt et al., 2012). These adult native speakers were presented with a total of 303 pictures (166 nouns and 137 verbs). They were asked to answer four questions (criteria provided by Haman et al., 2015) for each picture: 1) to report whether this picture represents any word in Urdu; 2) to write the first word that comes to their mind after looking at the picture; 3) to assess whether the picture is the correct representation of the object/action or not; and 4) to rank the general style of the picture (colors, shapes, drawing). Different labels for each picture were received through this naming phase. Based on the frequency of the labels obtained during this phase, 204 nouns and 160 verbs (words) were shortlisted.

In phase 2, 25 adults (different from phase 1) who were native Urdu speakers were asked to state the estimated age of acquisition (AoA) for each word (labels from the naming task). After compiling the results from this phase, the final selection of target stimuli for producing and comprehending nouns and verbs was done by eliminating the least frequent, most complex, and loan words from the list.

Nouns and verbs were assigned to one of the three broad semantic categories to ensure semantic variety. Nouns were classified into inanimate natural kinds (e.g., cloud, leaf), animate natural kinds (e.g., tiger, frog), and artifacts (e.g., lamp, drum). Verbs were categorized into unintentional actions or states (e.g., to rain, to boil), actions performed by a human (e.g., to paint, to laugh), and actions performed by animals (e.g., to sting, to bark). Any loan words were avoided during the selection of targets.

As prescribed by Haman and colleagues (2015), certain principles were followed while selecting items for both tasks. For the comprehension tasks, all four pictures in a set (1 target and 3 distractors) belonged to the same semantic domain with a similar level of difficulty (four levels calculated based on the values of complexity index (CI) & value of the AoA for each candidate word). The pictures that could cause perceptual (e.g., words with similar picture styles like leaf and feather) and/or phonological (e.g., doll-dog) distractions were also avoided within a set.

In the final phase, 32 nouns and 32 verbs were chosen for all four tasks (30 target and 2 training items per task). Three distractors for each target item were selected for the comprehension tasks, and each participant was presented with four pictures (1 target and 3 distractors). This was done to minimize the perceptual load and to avoid the possibility of random choice (Haman et al., 2015).

For the production tasks, the stimuli were selected from the distractors in the comprehension tasks. The targets for comprehension and production were matched in terms of semantic domain, AoA, and CI to ensure a similar difficulty level in both tasks. Figures E to H (Supplementary Materials) present the Urdu CLT comprehension and production tasks for nouns and verbs.

Sentence Repetition Task (SRep; Marinis & Armon-Lotem, 2015). The SRep task was used to assess morphosyntactic abilities. The SRep tasks have been used widely to evaluate the acquisition of specific grammatical structures in children and adults in both their L1 and L2. The studies have provided ample evidence that the SRep is not just parroting or does not simply rely on memory capacity. In fact, it represents an individual's linguistic knowledge, highly dependent on morphosyntax and lexical phonology and less on prosody and semantics (Polišenská et al., 2015; Frizelle & Fletcher, 2014). Studies have suggested that children find it challenging to repeat unfamiliar structures (Devescovi & Caselli, 2007) and that an overlap exists between the SRep errors and errors committed in spontaneous settings (Riches, 2012). Hence, it can be concluded that the SRep is a valuable tool for assessing morphosyntactic abilities that could not otherwise be examined via spontaneous language samples (Seefi-Gabriel et al., 2010). It also facilitates classifying impaired vs. typical acquisition of grammatical structures in a target language (Taha et al., 2021).

The Urdu-SRep task has 35 test items based on 10 sentence structures that are graded in syntactic complexity and length depending on blocks: blocks 1 and 2 include items with the same number of words and syllables; block 3 contains items with a slightly more extended number of words and syllables. The sentence length is controlled within each block (see Table B in Supplementary Materials).

Sentence structures have different categories. There are control items/conditions containing a range of syntactically simple structures included as control structures. These include monoclausal and biclausal sentences with subordination and coordination. These sentences are matched to the syntactically complex sentences in terms of length. Second, there are Language Independent Structures (LI), including a range of syntactically complex structures that are claimed to be challenging for children with DLD across languages based on the cross-linguistic literature. Lastly, there are Language Specific Structures (LS), including a range of structures that are likely to be challenging for children with DLD in a specific language. The Urdu version included:

- Control structures/conditions: monoclausal sentences (simple Subject-Object-Verb [SOV] sentences) and biclausal sentences with coordination, subordination, and complement/adjunct
- LI structures: object wh-questions, object relative clauses, and conditionals
- LS structures: SOV with negation and 1 auxiliary/modal, SOV with 2 auxiliaries/1 auxiliary and 1 modal, SOV with negation, 2 auxiliaries/1 auxiliary and 1 modal

Based on the cross-linguistic literature on DLD regarding the vulnerabilities in inflectional morphology, the typological characteristics of Urdu, and clinical experiences of Urdu-speaking SLTs, the items targeting nominal (postpositions/case markers) and verbal (gender and number marking on verbs) morphology were also incorporated in the Urdu-SRep in the simple monoclausal SOV sentence items.

Followings were controlled in the Urdu SRep version: simple SOV sentences (where the syntactic complexity is minimum) were used to allow independent testing of nominal and verbal morphology (similar to the Russian and Hebrew SRep tasks; Meir et al., 2015); the same gender and number marking were used for the nouns within a sentence; nominal expression was also controlled meaning that almost same proper and common nouns were used across sentences to minimize lexical specificity effect; and the number of noun phrases (2) was also kept constant across sentences.

The SRep task was embedded into a PowerPoint using a game-like format where a bear was hunting for a treasure, and as children repeated an utterance, the bear took a step forward. After the last sentence, the bear entered a cave and found the treasure. The gamification aspect helped in keeping children motivated during the activity (Marinis & Armon-Lotem, 2015). Three versions with different randomized orders of sentence items were developed, and each participant was randomly assigned to any one version. Figure I (Supplementary Materials) presents what an SRep task looks like.

Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2019). The MAIN (Gagarina et al., 2012) was used for narrative assessment and analyses. The revised English version (Gagarina et al., 2019a) was adapted into Urdu. The MAIN's key characteristics and strengths are highlighted as follows: it assesses the narrative competence (comprehension and production) of monolingual and multilingual children aged 3 to 12 years speaking different languages. Presently, it has been adapted into 92 languages. It includes four parallel cross-culturally balanced picture stories. Every story is based on six pictures. Every story has three episodes (two pictures for each episode). Each story includes three episodes consisting of the GAO sequence: a Goal (i.e., what a character wants), an Attempt to attain this goal, and the Outcome.

Narratives could be elicited using both storytelling and retelling. Every session begins with some warm-up discussion and is then followed by storytelling or retelling. Figures A to D (Supplementary Materials) present the story sequences used for Urdu-speaking children.

The comprehension section in the MAIN provides an overview of the child's understanding of the story. The 10 comprehension questions included in the MAIN are all inferential in nature. There are no literal or factual questions (e.g., Which color was the boy's shirt? – Blue; Where was the child sitting? – Chair). In contrast to literal or factual questions, inferential questions focus on comprehension of story aspects that are not explicitly depicted or mentioned and should be inferred/induced. For example, the MAIN includes inferential questions like, How does the cat feel? (Bohnacker & Gagarina, 2020). Table 2.2 presents examples of comprehension questions asked from the Dog story. Out of 10, three questions are based on an understanding of goals, six questions elicit ISTs that are related to initiating events or reactions, and the last question is to assess if the child can infer the meaning of the story as a whole.

Table 2.2 *The MAIN Comprehension Questions (Gagarina et al., 2019a)*

Questions	Target Components	Examples from Dog Story
D1	Goal	Why does the dog leap forward/jump?
D2	IST as reaction	How does the dog feel?
D3	IST rationale	Why do you think that the dog is feeling angry/ disappointed/ hurt etc.?
D4	Goal	Why does the boy jump/ leap upwards?
D5	IST as reaction	How does the boy feel?
D6	IST rationale	Why do you think that the boy is feeling good/ happy etc.?
D7	Goal	Why does the dog grab the sausages?
D8	ToM (IST as a reaction)	Imagine that the boy sees the dog. How does the boy feel?
D9	ToM (IST rationale)	Why do you think that the boy feels bad/ angry/ mad etc.?
D10	Overall plotline question	Will the boy be friends with the dog? Why?

Note. IST = Internal State Terms; ToM = Theory of Mind.

The MAIN assesses narrative production at the macrostructural level by evaluating: 1) story structure (SS); and 2) story complexity (SC). SS measures the total number of story grammar components/elements produced. These story grammar elements include setting, initiating events, goals, attempts, outcomes, and reactions. Table 2.3 shows different SS components/elements assessed in a MAIN story.

Table 2.3 *Story Structure Components from MAIN*

Component	Description (Lindgren, 2018)
Setting	Time and place of the events
Internal State as Initiating Event	What does the character perceive/feel that sets the story events in motion?
Goal	What does the character want?
Attempt	What does the character do (to reach the goal)?
Outcome	What is the result? What happens?
Internal State as Reaction	What are the feelings of the character (in response to the outcome)?

SC, conversely, informs about an individual's ability to generate well-formed episodes in a narrative by combining the main episodic components, Goal-Attempt-Outcome (G-A-O). This is based on Westby's (2005; 2012) binary decision tree. It specifies the level of an individual's SC. Different levels are presented in Table 2.4.

Table 2.4 *Different Levels of Story Complexity (Westby 2005; 2012)*

Type of sequence	Components
Action/reaction sequence	Attempt + Outcome (AO)
Incomplete episode	Goal + Attempt (GA)
	Goal + Outcome (GO)
Complete/full episode	Goal + Attempt + Outcome (GAO)

Next, the MAIN can also be used to assess discursive competence at the microstructural level. Microstructure measures focus on assessing competency in different dimensions, including productivity or narrative length, lexis, syntactic complexity, grammaticality, discourse cohesion, and others. The measures to evaluate different microstructural domains and their scoring methods are presented in Section 2.3.3 of this chapter.

The MAIN also measures the use of ISTs, which as mentioned earlier, refer to terms that describe a protagonist's internal states, mainly their feelings and mental states, such as thoughts, intentions, reactions, and emotions (Gagarina et al., 2015). Table A (Supplementary Materials) presents the types of ISTs included in the MAIN manual with some examples from English and Urdu.

Nonverbal Intelligence (Raven's Progressive Matrices; Raven & Court, 1998). Nonverbal intelligence was assessed using the Raven's Colored Progressive Matrices. This was also converted into a PowerPoint, and instructions were translated into Urdu (children's strongest language) to ensure that children gained a complete understanding of the task expectation. The children with a standard score of 70 and above were considered in the normal range.

Administration. As mentioned above, all assessment tasks were converted into a PowerPoint format. The instructions were mainly pre-recorded based on the administration guidelines provided for each assessment measure. The testing session began with greetings and some warm-up questions. For most of the children, testing was completed in one day. Each testing session was conducted using the following task presentation order:

1- Nonword repetition task (discussed in Chapter 3)

2- SRep task- the children were asked to listen to the target stimuli and repeat them

3- MAIN- each participant first retold the story (the selection of stories was based on the counterbalancing order in the manual) after listening to a model story, then answered the standard comprehension questions

4- CLT- the order of the task was counterbalanced between the participants. Half of the task, which could be comprehension or production of both nouns or verbs depending on the counterbalancing order, was completed in between the MAIN retelling and telling tasks

5- MAIN- the child was asked to now tell the story (without support from a model story)

6- CLT- the remaining half of CLT was completed after the MAIN storytelling

7- Raven's Colored Progressive Matrices- administered on the same or separate day depending on the child's status in terms of concentration and cooperation

This order of task presentation was chosen based on the following considerations: Tasks 1 (NWR) and 2 (SRep) were in a game-like format and served as a good base for building rapport with children. In addition, these two repetition tasks also provided a good warmup for making these children talk before a more expressive and linguistically demanding task like story retelling. The children in this study were required first to retell (with support from the model story) and then tell (without any support from the model story) the same stories. Therefore, a break was required between Task 3 (story retelling) and Task 5 (telling). For this purpose, and also following the CLT authors' prescribed method of administration (i.e., Two tasks for the same part of speech (comprehension and production) should not go as consecutive sessions. A break after each session or after at most two sessions was required), Task 4 (i.e., between story retelling and telling) and Task 6 (after story telling) were decided to be either CLT production or comprehension sub-tasks depending on the counterbalancing order. Lastly, Raven's Colored Progressive Matrices (Task 7) was chosen to be at the end, thinking that if the session gets too long for the children, it could be relegated to the new test session.

The Ph.D. student, an Urdu-speaking SLT, collected all the data.

2.3.4 Transcription, Data Coding, and Scoring

All the data were transcribed, coded, and scored by the Ph.D. student and a part-time research assistant, (both were Urdu-speaking SLTs), using the standard guidelines provided by the original authors of each assessment measure. The scoring criteria and the inter-rater reliability are detailed in the following sections.

Scoring.

1. Lexical Abilities. For the comprehension tasks, responses were marked as correct or incorrect. The responses were scored using a binary scoring scheme, and each child could obtain a total of 60 scores (30 nouns, 30 verbs).

In the production tasks, the responses were marked as correct using Haman and colleagues' (2015) categorization of responses provided for the CLT. The items were marked as correct or incorrect using the 0-1 scheme. In addition to the target word, some other alternate responses were also considered semantically correct if: 1) the synonyms or near-synonyms were suitable concerning the target picture (e.g., children used synonyms for wakeup جاگنا/اٹھنا); 2) the synonyms or near-synonyms were acceptable concerning potential input models (e.g., popping for bursting پھاڑنا/پھٹنا); 3) some mispronunciations were recognizable; 4) unexpected inflection were used (e.g., the picture shows one sock and children used plural); and 5) derivations that otherwise did not change the semantic content concerning the target stimuli (e.g., for spoon چمچی/چمچا چمچ) were used. One score was assigned for each correct response.

The responses were marked as incorrect if: 1) the derivations changed the class of the word, like marrying replaced by marriage; 2) the answers excluded the target word during the generic explanation of the picture; 3) there were semantic, associative, phonological, or perceptual confusions; or 4) hyponyms or hypernyms were used. Examples of incorrect answers are provided under the error analyses Section 2.3.5.

2. Morphosyntactic Abilities. All the utterances were transcribed verbatim, and the 0–1 (binary scoring) and 0–3 (error scoring) schemes were employed as described in Marinis and Armon-Lotem (2015). For the 0–1 scheme (binary scoring), the child received a score of one if the target stimuli were produced entirely correctly. If not, then a zero score was assigned. For the 0–3 (error scoring) scheme, the child received three scores if the utterance was produced verbatim, two scores if there was one mistake, one if there were two or three mistakes, and zero if there were more than three mistakes. Another scoring level included a 0–1 score for the correct target structure. Children obtained one score if they produced the correct target structure despite other errors (e.g., lexical errors) and zero if the target structure was omitted or substituted. Allowances for self-corrections, consistent phonological processes, and contractions were provided.

3. Narrative Abilities. Narrative data were scored for both comprehension and production. Narrative productions were segmented into communication units (C-units). Following Loban's (1976) convention, each independent clause with its modifiers was considered a C-unit. All the irrelevant comments, unintelligible words, and mazes were excluded.

Narrative Comprehension. Following the scoring guidelines in Gagarina et al. (2019), one score per correct answer and zero score for each incorrect answer was provided. A child could receive a total of 10 scores for each story.

Narrative Macrostructure.

- *Story structure (SS).* The guidelines from the revised MAIN manual (Gagarina et al., 2019a) for scoring for SS (the maximum score was 17) were used. Each story included a setting (2 scores), three complete episodes (GAO; 9 scores), three internal states as initiating events (3 scores), and three internal states as reactions (3 scores).
- *Story Complexity (SC).* Since there is currently no gold standard in terms of the scoring approach one should use. This study has chosen three scoring schemes (see Table C in the Supplementary Materials) that have been used in the cross-linguistic studies whose score weighting assignment aligns with the level of story structure complexity based on the binary decision tree by Westby (2005) and represents three levels of granularity of distinguishing between different levels of structural complexity in macrostructure. Among them, the scoring scheme by Gagarina et al. (2019) has the finest level of differentiation (6 levels), the scoring scheme by Sheng et al. (2020) has the coarsest level of differentiation (2 levels), and the scoring scheme by Maviş et al. (2016) is in the middle (4 levels). Furthermore, the scoring schemes by Gagarina et al. (2019) and Sheng et al. (2020) have both successfully captured significant DLD/TD group differences in story complexity. The total number of complete Goal, Attempt, and Outcome (GAO) sequence(s)/complete episodes were also counted. See Table D in Supplementary Materials for narrative macrostructure scoring details.

Narrative Microstructure.

- *Productivity and Narrative Length.* The total number of C-units (TNC) and the total number of words (TNW) without mazes, code-switching, and unintelligible words were used to measure productivity and narrative length.
- *Lexical Diversity.* The number of different words (NDW) was used to assess lexical diversity.
- *Grammaticality Accuracy.* The proportion of grammatical C-units (Gprop) was used as one of the grammaticality accuracy measures. All the C-units in a story were tagged as grammatical or ungrammatical. Then, Gprop was computed by dividing the total number of grammatical C-units by the total number of C-units. Moreover, verb accuracy (VA), the number of correct verbs divided by total verbs, was also used to measure grammaticality/morphosyntactic accuracy.
- *Syntactic complexity.* Different measures were used to assess syntactic complexity in each story. To calculate the Proportion of complex C-units (Cprop), each C-unit in a story was marked as complex or simple, and then the total complex C-units were divided by the total C-units. Next, the Mean Length of Communication Units (MLCU) was also used to measure syntactic complexity. It was computed by dividing the total number of words without mazes, code-switching, and unintelligible words by the total number of C-units. Further, the proportion of subordinating C-units (Sprop) was also measured by dividing the total number of subordinating clauses (where the child used any subordinating conjunction(s)) over the total number of C-units.

Internal State Terms (ISTs). Both types and tokens of ISTs in each story were counted. Tokens are the number of individual words in a narrative sample (i.e., also counting the repetitions of individual words). In contrast, type refers to the number of unique word forms (i.e., repetitions of an individual word will only be counted once).

2.3.5 Error Analyses

Errors from the CLTs. Classification of errors was broadly based on the CLT's error categorization (Haman et al. 2015) and Altman et al.'s (2017) work. Table 2.5 presents different types of lexical errors.

Table 2.5 *Types of Lexical Errors*

Error Type	Description	Example
hypernyms	superordinate terms	bird for sparrow
hyponyms	subordinate terms	apple for fruit
semantic confusion	a different word from the same semantic category	dog for cat
associative confusion	words associated thematically with the target word	dog for bone
perceptual confusion	name of the activity/object that is perceptually like the target item	plate for button
phonological confusion	Name of an activity/object produced close in pronunciation to the target word	Mouth for mouse
definition	description of the picture without the target word	the man wears it around his neck for tie
wrong word class	derivational errors across class	noun for verb
other	when an error did not fit any of the above-mentioned categories	truck for apple
na	when the children said they did not know the answer	

The number of errors for each child was counted and then assigned to any of the categories mentioned above. Like SRep, the number of target items for each child was also fixed in CLT, so the raw frequencies for each type of error were used in the statistical analyses.

Errors from the SRep and MAIN tasks. Error Analyses focused on morphosyntactic (gender and number marking, case and case marker/postposition, word order, and incomplete utterances) and lexical errors. First, this section briefly presents the typological characteristics of Urdu. The errors were categorized based on these typological characteristics.

Urdu is a morphologically rich language. Nouns in Urdu have an inherent gender and are inflected for gender, number, and case (Humayoun et al., 2007; Rizvi, 2007). Verbs in Urdu are also inflected for gender, number, tense, aspect, and mood (Humayoun et al., 2007; Rizvi, 2007). Urdu uses affixation in the form of morphemes that are lexically attached to the word for number and gender marking (Syed, 2013). The total occurrences of gender and number of errors were counted and are presented separately.

The errors for case, postpositions/case markers were also counted and are presented under one category. There are three cases for nouns in Urdu, namely nominative, vocative, and oblique (Schmidt, 1999). Urdu nouns usually are listed in their nominative forms in dictionaries. If the nouns are followed by postpositions, they appear in the oblique form. The form of nouns that are used to refer to animate nouns or humans or address any person is known as vocative. Postpositions/case markers, on the other hand, highlight the case of a noun phrase and aid in determining the grammatical function of a noun in a sentence (Rizvi, 2007). Unlike many languages where case markers/postpositions are lexically dependent and morphologically attached to the word, they are lexically independent and syntactically attached in Urdu (Syed, 2013).

The errors based on incomplete utterances and wrong word order were also counted. Urdu mainly follows subject-object-verb (SOV) word order, but variations are also accepted. Based on its robust case marking system that elucidates the object and subject nouns in a sentence, Urdu can exercise the free word order phenomenon. In other words, the order of subject, object, and verbs in Urdu phrases is relatively free, but the most acceptable form is SOV (Butt, 2006; Rizvi, 2007). While counting the errors based on word order, only the utterances that totally disrupted the grammaticality of an utterance were considered. Lastly, the errors at the word or content level, termed lexical errors, were also counted. These include the use of wrong verbs or wrong word forms, etc.

All the morphosyntactic errors emerging from both SRep and narrative tasks were identified, coded, and then categorized based on the locus of error (gender marking, number marking, case or case marker/postposition word order, and incomplete utterances, etc.). The most frequently occurring error types are presented in Table 2.6.

For narrative microstructure errors (narrative task), the procedure by Altman et al. (2016) was followed. The relative frequency of errors was calculated separately for each child. To compute the relative frequency, the raw frequencies of each error type were divided by the TNC by each child. Calculating errors using this method allows for comparing individual differences despite the variations in story length. These relative frequencies were used in the statistical analyses.

In the SRep task, the raw frequencies for each type of error were employed in the statistical analyses as the number of target items for each child was fixed. Table 2.6 presents the Urdu examples of errors under each error type with an English translation in the red font.

Table 2.6 *Error Types with Examples Identified During Narrative Microstructure and SRep Analyses*

Error Types	Example in Urdu	English Translation
Number Marking	بچے کھیل رہا ہے۔	Children is playing.
Gender Marking	بلی بیٹھا ہے۔	Cat [fem.] is sitting [masc.].

Error Types	Example in Urdu	English Translation
Case or Postposition	گیند پانی پر گر گئی۔	The ball fell on the water.
Incomplete Utterances	وہ لڑکا دیکھ ---	That boy see ----
Word Order	ماں لینے پرندوں کی گئی کھانا۔	The mother went birds bring food.
Lexical	وہ خوشی ہوا۔	He became happily.

2.3.6 Inter-rater Reliability

For the inter-rater reliability, 13 percent of the data were transcribed by a part-time research assistant who is an Urdu-speaking SLT and was double-checked by the Ph.D. student. The initial percentage of agreement between the two raters for different tasks was as follows: For the CLT, all the responses in the four sub-tasks (N-C, V-C, N-P, V-P) were scored independently by the two raters (total 1024 items). There was disagreement over three items (V-C task). The percentage of agreement between the two raters was 99.7% (1021 over 1024 items). For the SRep, all the utterances (35 items) from the 8 participants were scored independently by the two raters (total 280 items). The percentage of agreement between the two raters was 99.6% (279 over 280 items). For the MAIN, the narrative samples were scored for both narrative comprehension (2 stories) and narrative macrostructure production (4 stories per child). The percentage of agreement between the two raters for the narrative comprehension questions was 100% (160 over 160 items). The percentage of agreement between the two raters across different macrostructure measures were as follows: SS = 99.6% (542 over 544 items), SC-M = 99.6% (287 over 288 items), SC-Sh = 100%, (96 over 96 items), SC-Ga = 99.8% (575 over 576 items). There were very few minor disagreements due to occasional omissions and typos, which were quickly resolved after double-checking. All the scorings for each measure were completed by one rater and then double-checked by the second rater to ensure that all the utterances and items were scored correctly, not left unscored or scored twice. These disagreements were due to occasional omissions and typos that were resolved to 100 % agreement after inter-rater checks.

2.3.7 Statistical Analyses

The statistical analyses were completed by the Ph.D. student using R (version 4.2.1, R Core Development Team, 2022). Mixed-effects modeling was used as it allows examining the variables of interest (fixed effects) and their possible interactions while also considering variability within and across participants and items simultaneously (random effects and random slopes). Performance on different outcome measures (following binomial or poison distribution, e.g., SS) was predicted using the Generalized Linear Mixed-effects Models (GLMM). For the outcome variables following the Gaussian distribution (e.g., MLCU), the Linear Mixed-effects Models (LMM) were employed. The lme4 package (Bates et al., 2015) was used to run GLMM and LMM. The post-hoc results informing about the pairwise comparisons in case of significant interactions were computed using the emmeans package (Lenth, 2022) and the multcomp package (Hothorn, 2022). The findings were considered significant with p values less than 0.05.

Furthermore, simple bivariate correlations using Spearman's correlation coefficient were computed between the amount of Urdu input (number of hours of Urdu use at home and in school) and the various narrative, morphosyntactic, and lexical measures to obtain an insight into the nature of language exposure on language performance.

2.4 Results

This section presents the findings based on the linguistic abilities of the EM children and their Maj control peers from different language domains. Results for research questions are presented in turn under each language domain in the following sections.

2.4.1 Lexical Abilities

Lexical knowledge was assessed for two different word categories (nouns and verbs) using two different modalities (comprehension and production) via the CLT. The maximal model included total scores (number of correct responses for all four tasks) as an outcome variable, language status (sum-coded; EM = "1"; Maj = "-1"), age (mean-centered), amount of Urdu input (mean-centered), maternal education (mean-centered), modality (sum-coded; Comprehension = "1"; Production = "-1"), word category (WC), and their interactions as fixed effects. Participants were added as random effects with modality as a random slope (Barr et al. 2013). WC was removed as a random slope to fit the model.

Table 2.7 shows that there was a main effect of language status, and as predicted the Maj group outperformed the EM group. Next, the impact of age and the amount of Urdu input was non-significant, indicating that the performance on CLT tasks was not significantly different in older and younger groups. There was a significant effect of modality, and as anticipated, the performance in comprehension tasks was significantly better than in production tasks. There was also a significant effect of WC, and as expected the performance in nouns was better than in verbs. A negative significant effect of maternal education showed that the children whose mothers had relatively lower education had better lexical knowledge in the target language.

There were two sets of significant interactions. The first significant interaction was between language status and WC, suggesting that the effect of WC was not uniform across the two language groups. The interpretation of this interaction is further assisted by Figure 2.1 and the post-hoc analysis (emmeans pairwise comparison), indicating that although both language groups performed better in nouns, this difference in scores based on WC was highly significant for the control group (p = < .0001) in comparison to the EM group (p = .043).

The second significant interaction was between age and modality, implying that the effect of age was not uniform across modalities. Figure 2.2, together with post-hoc analysis (emmeans pairwise comparison), registers that there was a significant age-related (older vs. younger) gap for the comprehension tasks (*p* = .0001) and not for the production tasks (*p* = .059). This shows that production tasks were relatively difficult for even older children.

Table 2.7 *GLMM Analysis Summary for Fixed Effects Predicting CLT Scores (Lexical Abilities; Max. Score 120)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.17	.089	13.1	<.0001***
Language Status (EM vs. Maj)	-.759	.129	-5.87	<.0001***
Age	.007	.004	1.67	.094
Amount of Urdu Input	.003	.002	1.47	.141
Maternal Education	-.027	.010	-2.65	.007**
Modality (Comprehension vs. Production)	1.63	.142	11.4	<.0001***
Word Category (Noun vs. Verb)	-.591	.096	-6.12	<.0001***
Language Status: Age	-.004	.005	-.806	.420
Language Status: Modality	.025	1.57	.160	.873
Language Status: Word Category	.366	.120	3.04	.002**
Age: Modality	.010	.005	1.99	.045*
Age: Word Category	.007	.003	1.92	.054
Modality: Word Category	.205	.132	1.55	.120

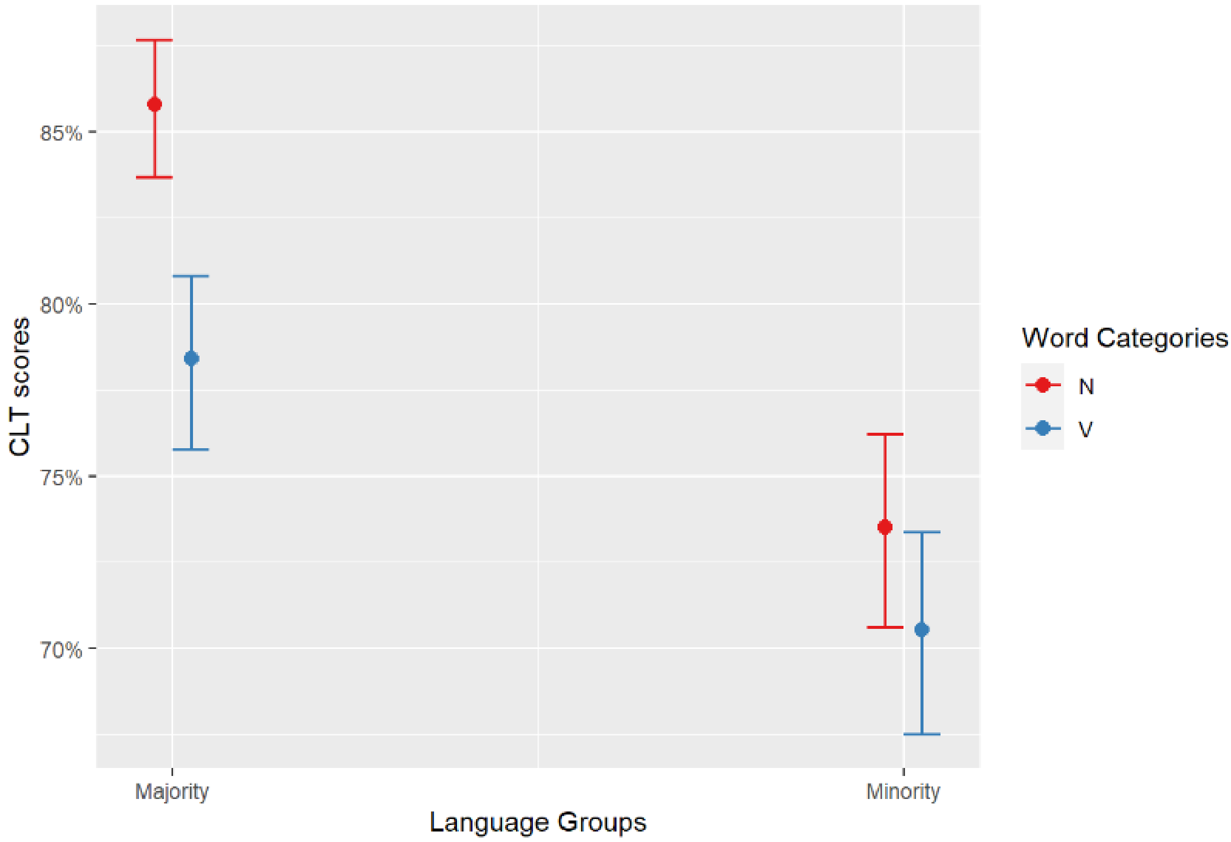

Figure 2.1 Children’s CLT (lexical) scores in Each Language Group by Word Category

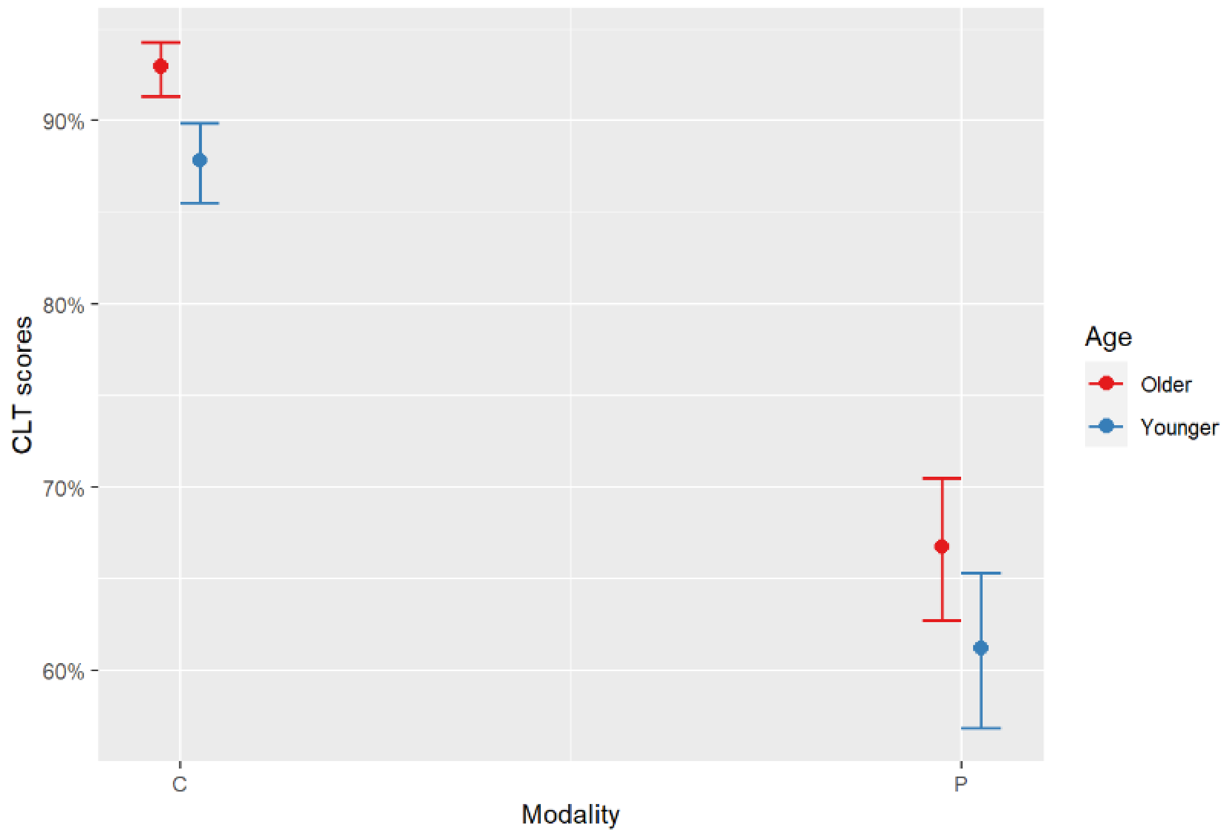

Figure 2.2 Children's CLT (lexical) scores in Each Modality by Age Groups

Error Analyses. This section will present the distribution of errors in verb and noun productions. Table 2.8 shows the overall distribution of errors in verbs. Frequencies suggest that these children made a relatively higher number of associative confusion (e.g., homework for write), semantic confusion (e.g., cut for peel), and other (where the error could not be categorized under any error type, e.g., run for drive) errors. Hypernym errors (e.g., clean for vacuum) were the least common.

For noun production, children produced semantic confusion errors frequently (e.g., apple for pear) followed by na responses (where the children responded that they didn't know the answer). Hypernym errors, like verbs, during noun production (e.g., fruit for orange) were committed least frequently.

Table 2.8 *Distribution of Errors by Word Category*

	Verbs Errors (*N*= 485)	Nouns Errors (*N*= 279)
Error Types	*N (%)* of Errors	*N (%)* of Errors
Hypernym	20 (4)	7 (3)
Semantic confusion	134 (28)	135 (48)
Associative confusion	136 (28)	10 (4)
Perceptual confusion	27 (6)	18 (6)
Other	131 (27)	33 (12)
na	37 (8)	76 (27)

Table 2.9 presents the effect of different factors on the production of lexical errors. The full model included language status (sum-coded), age (mean-centered), error type (sum contrast coding), and their interactions as fixed effects, and participants as random effects with WC as a random slope (Barr et al., 2013). It shows that language status had a significant effect, and EM children committed a significantly higher number of lexical errors than their Maj peers. The effect of age was non-significant. There was a significant effect of error type, which can be visualized in Figure 2.3, showing the percentage of the lexical error distribution. It registers children committing more semantic errors, whereas hypernym errors were the least prevalent.

There were two sets of significant interactions. The first significant interaction was between language status and error type, suggesting that the effect of language status was not uniform across different error types. The other interaction was between age and error type, showing that the effect of age varied between error types. The interaction between language status and age was non-significant.

Table 2.9 *GLMM Analysis Summary for Fixed Effects Predicting the Production of Lexical Error*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	-.312	.056	-5.48	<.0001***

Fixed Effects	*B*	*SE*	*z*	*p*
Language Status (EM vs. Maj)	.157	.055	2.82	.004**
Age	-.006	.003	-1.80	.071
Error_type1 (Hypernym)	-1.22	.169	-7.19	<.0001***
Error_type2 (Semantic)	1.08	.073	14.73	<.0001***
Error_type3 (Associative)	.483	.086	5.60	<.0001***
Error_type4 (Perceptual)	-.778	.143	-5.42	<.0001***
Error_type5 (Other)	.506	.088	5.75	<.0001***
Language Status: Age	.003	.002	1.41	.157
Language Status: Error_type1(Hypernym)	.028	.169	.171	.864
Language Status: Error_type2 (Semantic)	-.326	.072	-4.49	<.0001***
Language Status: Error_type3 (Associative)	-.170	.085	-1.99	.046*
Language Status: Error_type4 (Perceptual)	-.245	.134	-1.82	.068
Language Status: Error_type5 (Other)	.076	.084	.906	.364
Age: Error_type1(Hypernym)	.009	.010	.865	.387
Age: Error_type2 (Semantic)	.007	.004	1.65	.098
Age: Error_type3 (Associative)	.007	.005	1.38	.167
Age: Error_type4 (Perceptual)	-.019	.009	-2.17	.030*
Age: Error_type5 (Other)	-.018	.005	-3.37	.000***

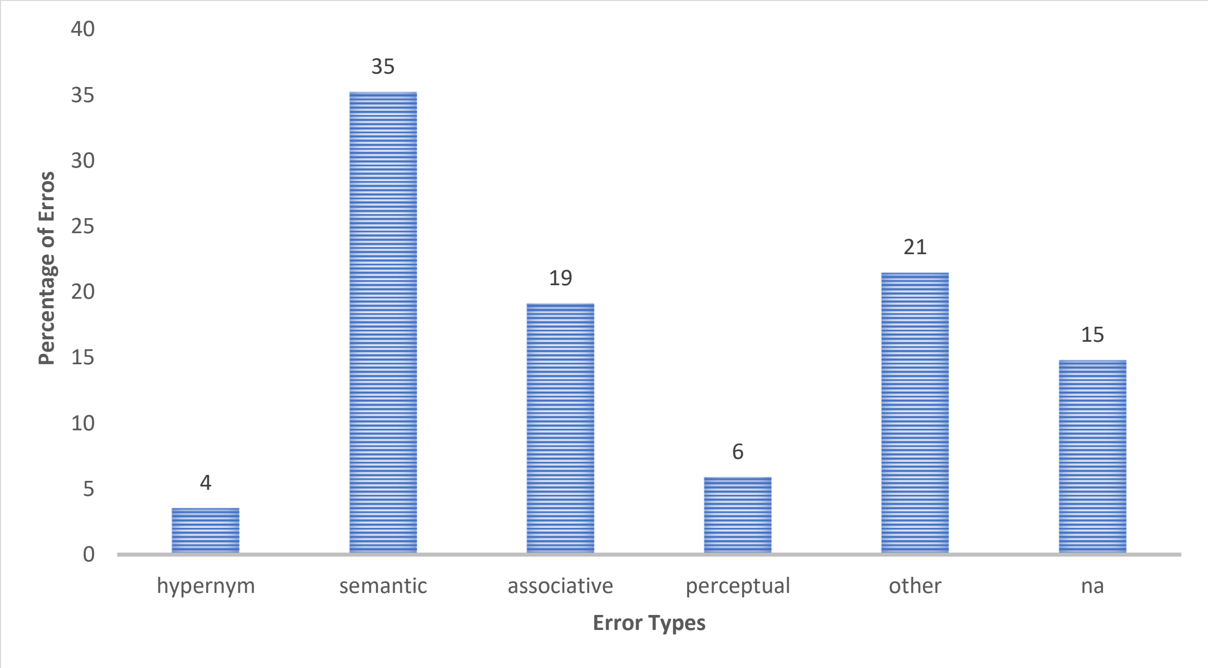

Figure 2.3 Distribution of Lexical Errors in Each Error Type

Interaction between Language Status and Error Type. Table 2.10 elaborates on the interaction between language status and error type via post-hoc (emmeans pairwise comparisons) analysis. This shows that EM children committed significantly more other errors (which could not be categorized) and na (children replying with I don't know) errors. On the other hand, Maj children produced significantly more semantic confusion errors. The number of hypernyms, associative, and perceptual errors were not significantly different in both language groups. The visual interpretation of this interaction is also presented in Figure 2.4.

Table 2.10 *Differences in Error Frequencies in Each Error Type by Language Status*

Error Types	*β*	*SE*	*z*	*p*
Hypernym	-.386	.398	-.972	.331
Semantic confusion	.340	.133	2.56	.010*
Associative confusion	.030	.173	.175	.861
Perceptual confusion	.188	.307	.614	.539
Other	-.442	.168	-2.63	.008**
na	-1.58	.25	-6.14	<.0001***

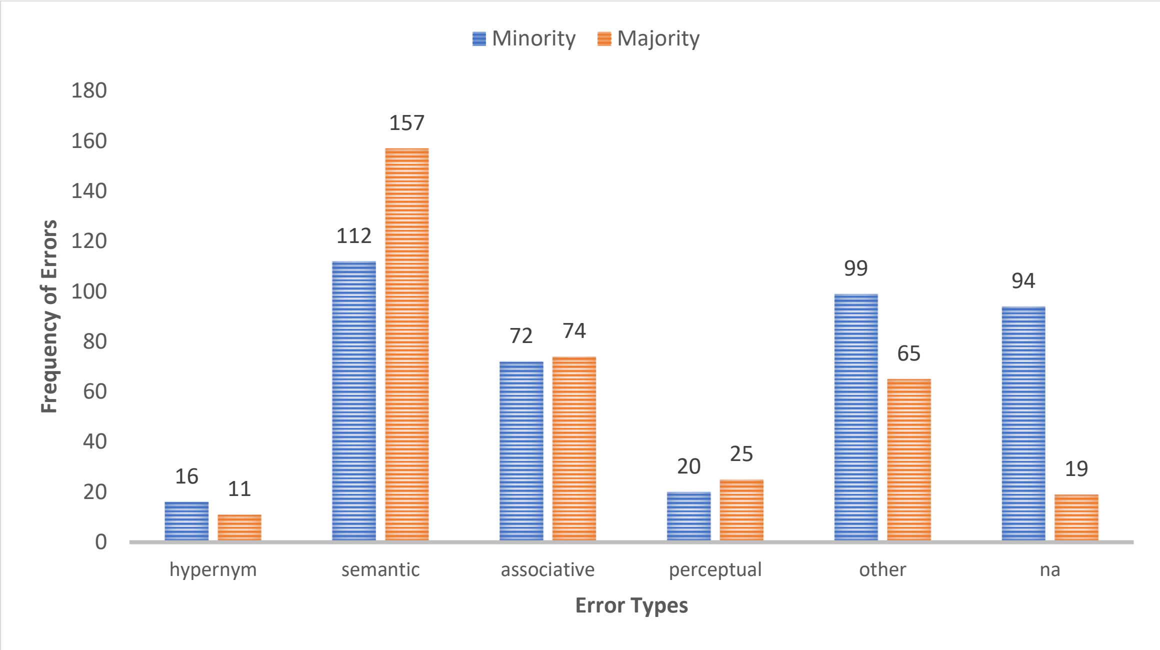

Figure 2.4 Distribution of Lexical Errors in Each Error Type by Language Status

Interaction between Age and Error Type. Table 2.11 presents this interaction's post-hoc (emmeans pairwise comparisons) results. This table and Figure 2.5 show that younger children produced a significantly greater number of perceptual and other lexical errors. For the rest of the error types, the difference between both age groups was non-significant.

Table 2.11 *Differences in Error Frequencies in Each Error Type by Age*

Error Types	β	SE	z	p
Hypernym	.309	.394	.785	.432
Semantic confusion	-.021	.132	-.163	.870
Associative confusion	-.020	.173	-.116	.907
Perceptual confusion	.650	.318	2.04	.040*
Other	.456	.167	2.72	.006**
na	-.205	.204	-1.00	.313

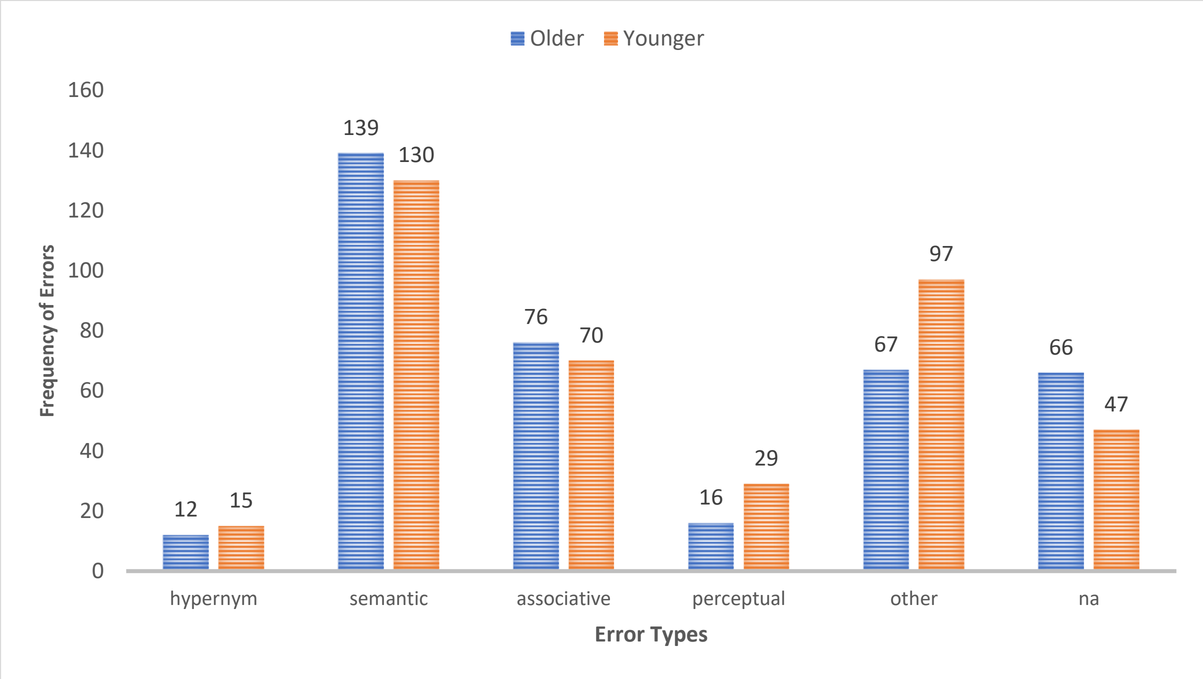

Figure 2.5 Distribution of Lexical Errors in Each Error Type by Age

2.4.2 Morphosyntactic Abilities

Morphosyntactic abilities were assessed using the SRep task. Three different types of scoring schemes were employed. The findings based on each scoring scheme are presented individually. The maximal models included language status (sum-coded; EM = "1"; Maj = "-1"), age (mean-centered), amount of Urdu input (mean-centered), maternal education (mean-centered), sentence length (mean-centered), and sentence structures (sum contrast coding), interactions between language status and age, and language status and sentence structures as fixed effects, and participants and items as random effects (Barr et al. 2013).

Binary Scoring (0-1 score). This was a relatively strict coding scheme, and the score was awarded only if the item was repeated verbatim. Table 2.12 shows that language status had a significant effect, and as predicted, the EM children performed significantly worse than their control peers. Next, there was also a significant effect of age, and the older children outperformed the younger ones. The effect of sentence item length was also significant, and the performance on relatively shorter sentences appeared to be significantly better. The effect of sentence structure was also significant. Figure 2.6 allows to visualize this finding, and it can be seen that the accuracy proportion differs across sentence structures. The findings show that overall, children found repeating items with the subordinate structure to be the most difficult, followed by the wh-object Q and ORC structure items. The effects of the amount of Urdu input and maternal education were non-significant. The interaction between language status and age was not significant. Lastly, there was a significant interaction between language status and sentence structure.

Table 2.12 *GLMM Analysis Summary for Fixed Effects Predicting SRep Scores (Binary Scoring; Max. Score 35)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.57	.181	8.65	<.0001***
Language Status (EM vs. Maj)	-.594	.186	-3.18	.001**
Age	.020	.008	2.46	.014*
Amount of Urdu Input	.013	.007	1.97	.055
Maternal Education	-.016	.034	-.484	.628
Item Length	-.406	.147	-2.75	.005**
Structure1-1auxiliary/modal	.100	.350	.287	.774
Structure2-auxiliary/modal+negation	.789	.440	1.81	.069
Structure3-coordination	.483	.549	.880	.379
Structure4-2auxiliary/1auxiliary+modal	.371	.405	.917	.358
Structure 5-auxiliary+modal+negation	.898	.454	1.97	.048*
Structure6-subordination	.784	.499	1.57	.116
Structure7-complement/adjunct	-.267	.399	-.670	.502
Structure8-Wh Object Question	.294	.543	.541	.588
Structure9-ORC	-1.99	.410	-.485	<.0001***
Language Status: Age	-.004	.008	-.492	.622

Fixed Effects	*B*	*SE*	*z*	*p*
Language Status: Structure1	-.064	.151	-.428	.668
Language Status: Structure2	.385	.240	1.60	.109
Language Status: Structure3	-.485	.293	-1.65	.097
Language Status: Structure4	.094	.204	.464	.642
Language Status: Structure5	-.049	.269	-.184	.854
Language Status: Structure6	-.573	.220	-2.59	.009**
Language Status: Structure7	.172	.187	.919	.358
Language Status: Structure8	-.123	.177	-.692	.488
Language Status: Structure9	.174	.176	.992	.321

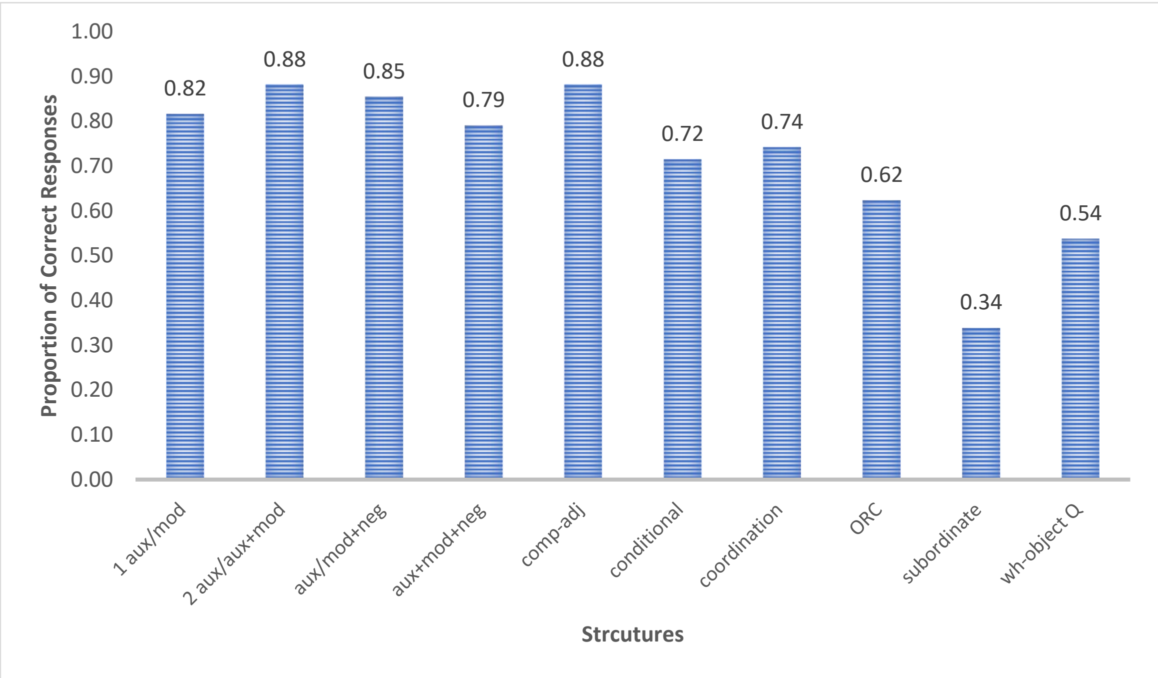

Figure 2.6 Accuracy Proportion in Each Sentence Structure

Interaction between Language Status and Sentence Structures. Table 2.13 presents the results from the post-hoc analysis (emmeans pairwise comparisons) for this interaction. It registers that the EM children performed significantly worse in different sentence structures, including SOV with one auxiliary/model (test items where nominal and verbal morphology was targeted), SOV with one auxiliary/model + negation, conditionals, and ORC. The interpretation of this interaction is further assisted by Figure 2.7.

Table 2.13 *Differences in Performance on Each Sentence Structure by Language Status*

Structures	*β*	*SE*	*z*	*p*
1auxiliary/modal	1.31	.458	2.87	.004**
2auxiliary/1auxiliary+modal	.419	.622	.673	.501
auxiliary/modal+negation	2.16	.726	2.97	.002**
auxiliary+modal+negation	1.00	.553	1.80	.070
complement/adjunct	1.28	.67	1.90	.057
conditional	2.33	.582	4.01	.0001***
coordination	.844	.521	1.62	.105
ORC	1.43	.501	2.86	.004**
subordination	.840	.491	1.71	.086
Wh Object Question	.252	.473	.533	.594

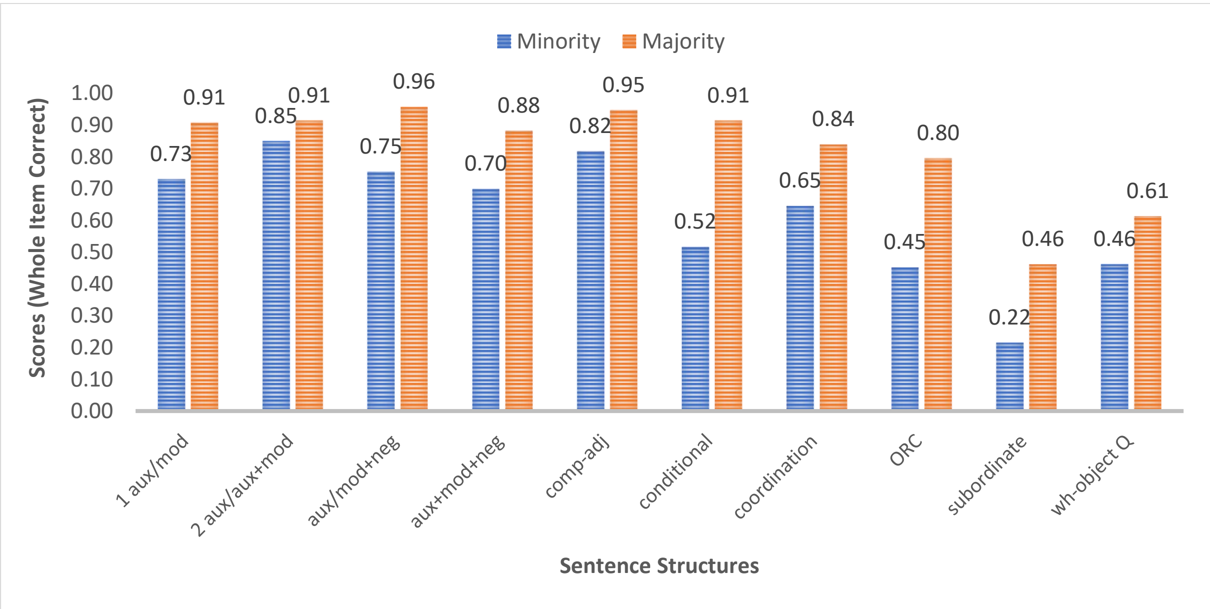

Figure 2.7 Differences in Performance on Each Sentence Structure by Language Status

Error Scoring (0-3 score). This was a comparatively lenient coding scheme, and the score was awarded based on the number of errors (see Morphosyntactic Abilities under Section

2.3.3). Table 2.14 showed that when this lenient scoring was employed, the effect of language status was non-significant, but there was a positive significant effect of the amount of Urdu input. It shows that children who received a higher amount of exposure scored well. There was also a significant effect of age. The older children performed significantly better than the younger ones. There was also a significant effect of sentence item length, and the performance in shorter sentences was better. There was a non-significant effect of maternal education. The effect of sentence structure was also significant. Figure 2.8 visualizes this main effect, showing that accuracy proportion varies across sentence structures. It shows that, like binary scoring, children generally found repeating items with subordinate structure the most difficult followed by wh-object Q and ORC structure items. None of the interactions were significant.

Table 2.14 *GLMM Analysis Summary for Fixed Effects Predicting SRep Scores (Error Scoring; Max. Score 105)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	.943	.014	65.1	<.0001***
Language Status (EM vs. Maj)	-.030	.019	-1.54	.122
Age	.001	.000	2.19	.028*
Amount of Urdu Input	.008	.000	2.43	.014*
Maternal Education	-.002	.003	-.062	.530
Item Length	-.042	.015	-2.70	.006**
Structure1-1auxiliary/modal	.002	.037	.069	.944
Structure2-auxiliary/modal+negation	.070	.043	1.64	.099
Structure3-coordination	-.005	.052	-.096	.923
Structure4-2auxiliary/1auxiliary+modal	.057	.042	1.33	.180
Structure 5-auxiliary+modal+negation	.072	.042	1.68	.092
Structure6-subordination	.082	.053	1.55	.121

Fixed Effects	*B*	*SE*	*z*	*p*
Structure7-complement/adjunct	.035	.042	.818	.413
Structure8-Wh Object Question	.014	.061	.236	.813
Structure9-ORC	-.154	.049	-3.11	.001**
Language Status: Age	.000	.000	.327	.743
Language Status: Structure1	.012	.028	.437	.662
Language Status: Structure2	.049	.041	.198	.843
Language Status: Structure3	.008	.041	.198	.843
Language Status: Structure4	.012	.042	.288	.773
Language Status: Structure5	.028	.041	.683	.494
Language Status: Structure6	-.067	.043	-1.55	.119
Language Status: Structure7	.014	.042	.352	.724
Language Status: Structure8	-.071	.045	-1.57	.114
Language Status: Structure9	-.001	.047	-.030	.976

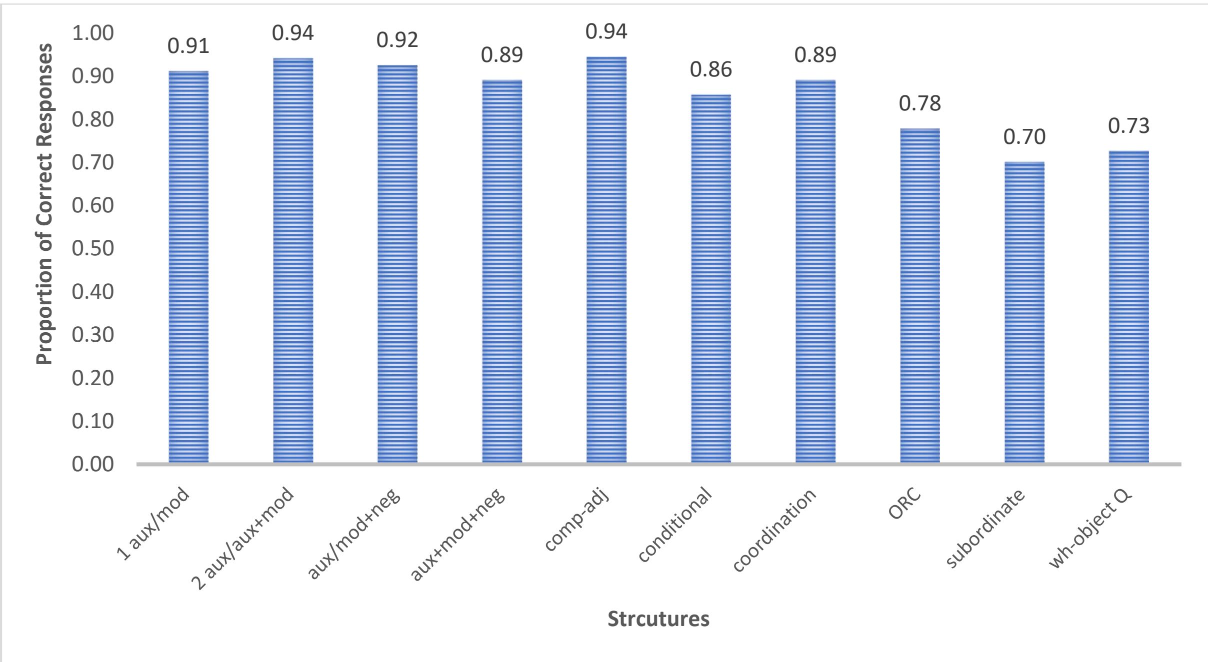

Figure 2.8 Accuracy Proportion in Each Sentence Structure (0-3 Scoring)

Structure Scoring (0-1 score). This scoring scheme awarded the score if the target structure was produced correctly. Table 2.15 shows that there was a significant effect of age, and the older children outperformed the younger ones. The impact of language status and amount of Urdu input was non-significant, showing that the EM children who are acquiring their L1 under reduced input conditions perform at par with their Maj peers when assessed solely for the maintenance of the grammatical structures and when their sentence repetitions were not penalized for errors such as lexical substitutions that do not affect target structure of the sentence.

Furthermore, the effects of maternal education, length, and sentence structure were also not significant. The overall effect of sentence structure was also double-checked via an omnibus test, and it turned out that, unlike the maximal model output, there was a significant effect of sentence structure ($\chi 2$ (9) = 56.39, p < .0001) on the maintenance of target structure. This effect could have been modulated in the presence of other predictors in the maximal model. The substantial effect of sentence structure can also be seen in Figure 2.9, where children showed varied performance across structures and scored significantly worse during the repetition of items targeting subordinate and wh-object question structures. There were no significant interactions.

Table 2.15 *GLMM Analysis Summary for Fixed Effects Predicting SRep Scores (Structure Scoring; Max. Score 35)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	3.46	78.4	.044	.964
Language Status (EM vs. Maj)	-1.19	78.4	-.015	.987
Age	.028	.009	3.04	.002**
Amount of Urdu Input	.012	.007	1.52	.127
Maternal Education	.007	.038	.192	.848
Item Length	-.100	.118	-.844	.398
Structure1-1auxiliary/modal	.045	78.4	.001	.999

Fixed Effects	*B*	*SE*	*z*	*p*
Structure2-auxiliary/modal+negation	.967	78.4	.012	.990
Structure3-coordination	8.14	706	.012	.990
Structure4-2auxiliary/1auxiliary+modal	-.540	78.4	-.007	.994
Structure 5-auxiliary+modal+negation	.722	78.45	.009	.992
Structure6-subordination	-.233	78.44	-.003	.997
Structure7-complement/adjunct	-1.63	78.44	-.021	.987
Structure8-Wh Object Question	-1.23	78.44	-.016	.987
Structure9-ORC	-3.49	78.44	-.045	.964
Language Status: Age	-.007	.009	-.727	.467
Language Status: Structure1	.349	78.44	.004	.996
Language Status: Structure2	1.28	78.45	.016	.986
Language Status: Structure3	-6.77	706	-.010	.992
Language Status: Structure4	.727	78.44	.009	.992
Language Status: Structure5	.318	78.44	.004	.996
Language Status: Structure6	.593	78.45	.008	.993
Language Status: Structure7	.848	78.44	.011	.991
Language Status: Structure8	.689	78.44	.009	.992
Language Status: Structure9	.883	78.44	.011	.991

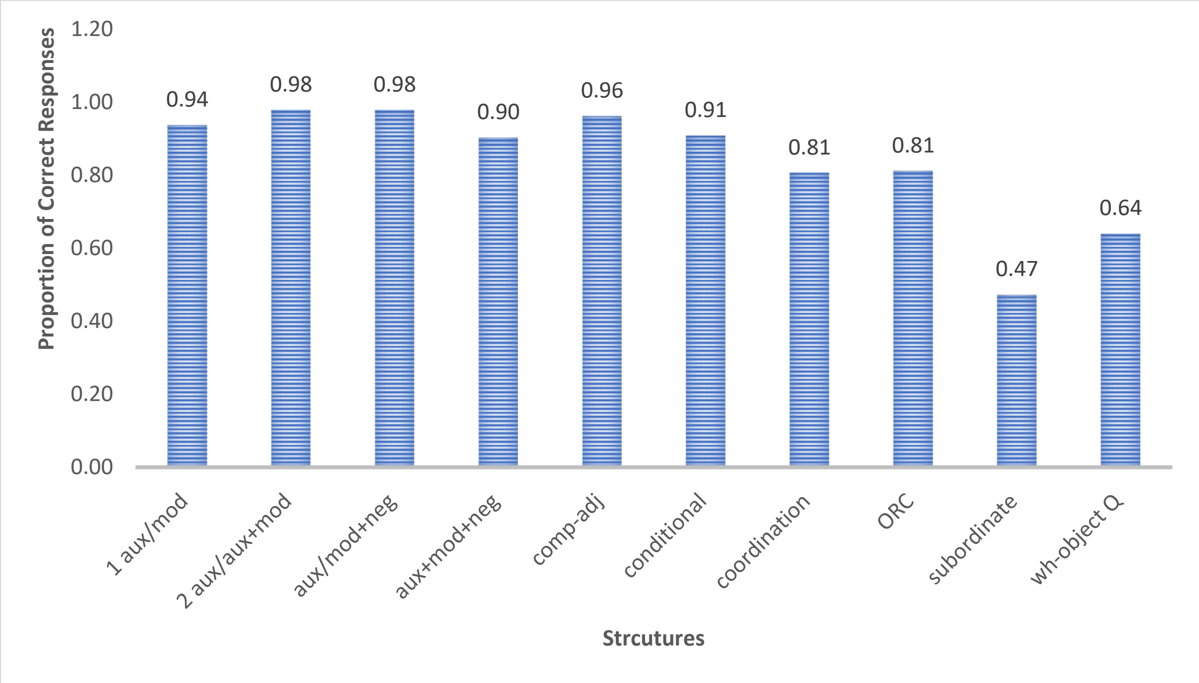

Figure 2.9 Accuracy Proportion in Each Sentence Structure (Structure Correct)

Error Analyses. This section provides insights into the EM children's linguistic vulnerabilities by presenting morphosyntactic error patterns observed in both language groups. Table 2.16 shows that the EM children made more errors in all categories. However, the error pattern is largely similar in both language status groups. The highest number of errors were committed while repeating items from wh-object questions. It indicates that this structure was challenging for both groups. For this structure, both groups committed a higher number of word order errors followed by postposition errors. The EM group committed the next highest number of errors during the repetition of items with one aux/model structure. However, these were simple items, and nominal and verbal morphology were tested through these items. In the rest of the items, number, gender, and tense were controlled within each structure. This implies that morphology is challenging for EM children even when tested in the simplest structures. The least number of errors were committed while repeating biclausal coordination sentences in both groups.

Table 2.16 *Error Distribution in Each Structure by Language Status*

Language Groups	Minority					Majority				
Error Types	Total	WO	GM	NM	PP	Total	WO	GM	NM	PP
Sentence Structures	*(N)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*(N)*	*N (%)*	*N (%)*	*N (%)*	*N (%)*
1auxiliary/modal	44	10 (23)	4 (9)	12 (27)	18 (41)	5	3 (60)	0	1 (20)	1 (20)
2auxiliary/1auxiliary+modal	4	3 (75)	0	0	1 (25)	2	1 (50)	1 (50)	0	0
auxiliary/modal+negation	23	7 (30)	1 (4)	0	15 (65)	6	1 (17)	0	0	5 (83)
auxiliary+modal+negation	17	15 (88)	2 (12)	0	0	6	6 (100)	0	0	0
complement/adjunct	11	8 (73)	1 (9)	0	2 (18)	5	3 (60)	0	0	2 (40)
conditional	15	9 (60)	0	4 (27)	2 (13)	1	0	0	0	1 (100)
coordination	2	2 (100)	0	0	0	1	0	0	0	1 (100)
ORC	31	17 (55)	0	0	14 (45)	9	3 (33)	0	0	6 (67)
subordination	20	15 (75)	0	1	4 (20)	16	12 (75)	0	0	4 (25)
wh-Object Question	53	32 (60)	0	0	21 (40)	31	18 (58)	0	0	13 (42)

Note. WO = Word Order; GM = Gender Marking; NM = Number Marking; PP = Postposition

Table 2.17 shows the effect of different factors on the production of morphosyntactic errors. The maximal model included language status (sum-coded), age (mean-centered), error type (sum contrast coding), and their interactions as fixed effects, and participants and items as random effects (Barr et al., 2013). This table highlights that language status had a significant effect, and the EM children committed a significantly greater number of morphosyntactic errors than their Maj counterparts. The effect of age was non-significant. A significant effect of error type can be seen in Figure 2.10, which displays the varied distribution of morphosyntactic errors. It shows that children committed a higher number of word order errors followed by postposition errors. The overall percentage of gender and number marking errors was lower than other error types.

The interactions between language status and age and age and error type were non-significant. On the other hand, the interaction between language status and error type was also non-significant. Figure 2.11 also shows that the EM children committed a higher number of errors in each category.

Table 2.17 *GLMM Analysis Summary for Fixed Effects Predicting the Production of Morphosyntactic Error*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	-5.08	.284	-17.8	<.0001***
Language Status (EM vs. Maj)	.872	.218	3.99	<.0001***
Age	-.012	.104	1.16	.243
Error_type1(Word Order)	1.60	.199	8.02	<.0001***
Error_type2(Gender)	-1.61	.415	-3.88	.0001***
Error_type3(Number)	-1.33	.419	-3.18	.001**
Language Status: Age	.000	.008	.116	.907
Language Status: Error_type1(Word Order)	-.363	.196	-1.84	.065
Language Status: Error_type2(Gender)	.180	.411	.440	.659

Fixed Effects	*B*	*SE*	*z*	*p*
Language Status: Error_type3(Number)	.596	.412	1.44	.148
Age: Error_type1(Word Order)	-.001	.008	-.125	.900
Age: Error_type2(Gender)	-.000	.017	-.002	.998
Age: Error_type3(Number)	-.009	.013	-.719	.472

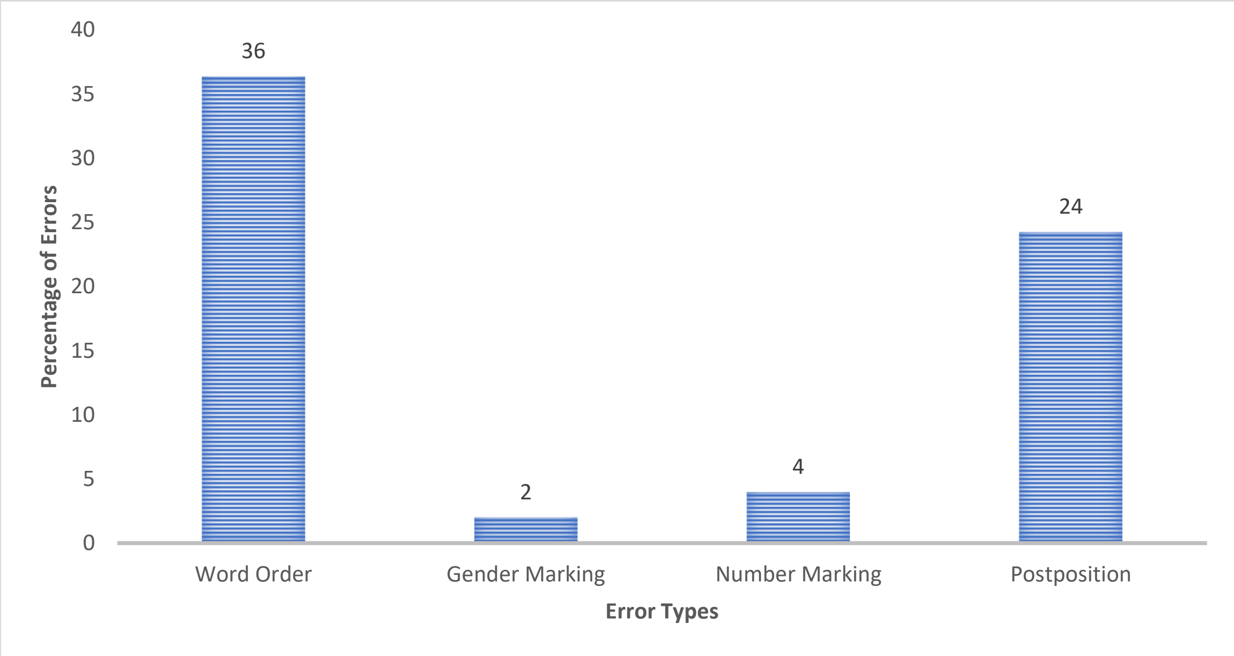

Figure 2.10 Distribution of Morphosyntactic Errors in Each Error Type

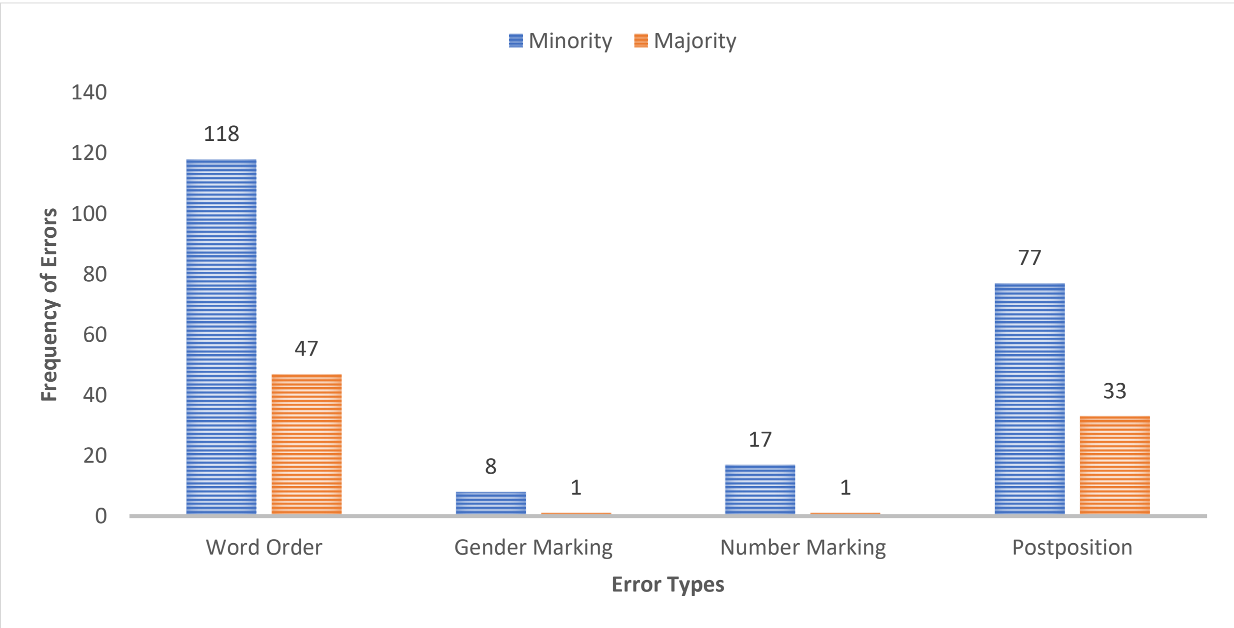

Figure 2.11 Frequencies of Morphosyntactic Errors in Each Error Type by Language Status

2.4.3 Narrative Abihlities

Narrative abilities were assessed at both comprehension and production levels using the MAIN. The maximal model for narrative comprehension included language status (sum-coded; EM = "1"; Maj = "-1"), age (mean-centered), their interaction, amount of Urdu input (mean-centered), and maternal education (mean-centered) as fixed effects, and participants as random effects (Barr et al., 2013).

Four stories were told by each child using two elicitation modes (retelling vs. telling) during the production part. So, in the narrative production models, elicitation mode was also included as a fixed effect in addition to the above-mentioned predictor variables. The random effects for narrative production models were also modified by adding elicitation mode as a random slope. These maximal models also included three two-way interactions (language status: age; language status: elicitation mode; age: elicitation mode) and one three-way (language status: age: elicitation mode) interaction.

Narrative Comprehension. Table 2.18 presents results for narrative comprehension. It shows a statistically non-significant effect of language status, and both language groups, despite minority vs. majority status, obtained comparable scores. The effects of age, amount of Urdu input, maternal education, and interaction between language status and age were also non-significant.

Table 2.18 *GLMM Analysis Summary for Fixed Effects Predicting Narrative Comprehension Scores (Max. Score 20)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	2.58	.836	3.08	.002**
Language Status (EM vs. Maj)	-.117	.394	-.298	.765
Age	.020	.012	1.69	.091

Fixed Effects	*B*	*SE*	*z*	*p*
Amount of Urdu Input	-.004	.008	-.546	.585
Maternal Education	-.001	.036	-.051	.959
Language Status: Age	-.010	.017	-.597	.550

Narrative Macrostructure. The production of narrative macrostructure was measured using story structure (SS) and story complexity (SC). Three different scoring schemes were used to examine SC (see Section 2.2.3 for details). Tables 2.19–2.22 show that both EM and Maj groups obtained comparable scores at the macrostructural level (SS and SC). The effect of age was significant for SS scores only, and the older children outperformed the younger ones.

Moreover, the effects of the amount of Urdu input and maternal education were non-significant for both SS and SC. The impact of the elicitation mode was also non-significant, suggesting that the performance in retelling and telling conditions was similar. There were no significant two-way and three-way interactions.

Table 2.19 *GLMM Analysis Summary for Fixed Effects Predicting SS Scores (Max. Score 34)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	2.81	.031	68.53	<.0001***
Language Status (EM vs. Maj)	.040	.039	1.01	.308
Age	.005	.002	2.53	.011*
Amount of Urdu Input	.000	.003	.000	.999
Maternal Education	.008	.006	1.35	.176
Mode (Retell vs. Tell)	-.013	.042	-.319	.750
Language Status: Age	-.000	.002	-.382	.702
Language Status: Mode	-.011	.042	-.268	.788
Age: Mode	.000	.002	.049	.960
Language Status: Age: Mode	.000	.002	.036	.971

Table 2.20 *GLMM Analysis Summary for Fixed Effects Predicting SC Scores (Sheng et al., 2020; Max. Score 6)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	-.035	.112	-.315	.001**
Language Status (EM vs. Maj)	.181	.138	1.31	.189
Age	-.005	.007	-.716	.474
Mode (Retell vs. Tell)	-.277	.164	-1.68	.091
Amount of Urdu Input	.000	.004	.130	.896
Maternal Education	.012	.022	.576	.564
Language Status: Age	-.007	.007	-.971	.331
Language Status: Mode	.079	.164	.483	.628
Age: Mode	.007	.010	.742	.458
Language Status: Age: Mode	.004	.010	.434	.664

Table 2.21 *GLMM Analysis Summary for Fixed Effects Predicting SC Scores (Maviş et al., 2016; Max. Score 18)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.09	.060	18.1	<.0001***
Language Status (EM vs. Maj)	.033	.076	.432	.666
Age	.003	.003	.993	.321
Amount of Urdu Input	-.002	.002	-1.11	.264
Maternal Education	.004	.012	.393	.694
Mode (Retell vs. Tell)	-.066	.074	-.901	.368
Language Status: Age	-.003	.003	-.876	.381
Language Status: Mode	-.039	.074	-.536	.592
Age: Mode	.001	.004	.381	.703
Language Status: Age: Mode	.002	.004	.438	.662

Table 2.22 *GLMM Analysis Summary for Fixed Effects Predicting SC Scores (Gagarina et al., 2019b; Max. Score 36)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.94	.042	45.5	<.0001***
Language Status (EM vs. Maj)	.026	.054	.497	.619
Age	.001	.002	.522	.602
Amount of Urdu Input	-.001	.001	-.895	.371
Maternal Education	.003	.008	.363	.717
Mode (Retell vs. Tell)	-.051	.048	-1.06	.287
Language Status: Age	-.001	.002	-.698	.485
Language Status: Mode	-.015	.048	-.327	.743
Age: Mode	.002	.003	.655	.513
Language Status: Age: Mode	.001	.003	.634	.526

Narrative microstructure. Different measures were used to assess the narrative microstructure abilities.

Productivity. The total number of C-units (TNC) and the total number of words (TNW) without mazes, code-switching, and unintelligible words were used to measure narrative productivity. Tables 2.23 and 2.24 present results for TNC and TNW. Language status had a significant effect on both TNC and TNW. The EM children obtained higher scores, implying that they produced comparatively lengthier narratives than their Maj peers. There was a significant effect of age for TNW only, and the older children produced more TNW. The effect of the elicitation mode was non-significant for both measures of productivity, illustrating that these children produced almost comparable TNC and TNW in both elicitation conditions. Maternal education had a significant positive effect on both measures of productivity, showing that narrative length improved with an increase in the mother's level of education. The effect of Urdu input and all interactions were non-significant.

Table 2.23 *GLMM Analysis Summary for Fixed Effects Predicting TNC*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	2.68	.028	93.1	<.0001***
Language Status (EM vs. Maj)	.084	.037	2.23	.025*
Age	.002	.001	1.36	.170
Amount of Urdu Input	-.000	.001	-.178	.858
Maternal Education	.017	.006	2,78	.005**
Mode (Retell vs. Tell)	.035	.033	1.06	.287
Language Status: Age	-.000	.001	-.102	.918
Language Status: Mode	.017	.033	.526	.599
Age: Mode	.000	.002	.437	.662
Language Status: Age: Mode	.000	.002	.244	.806

Table 2.24 *GLMM Analysis Summary for Fixed Effects Predicting TNW*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	4.68	.033	138	<.0001***
Language Status (EM vs. Maj)	.100	.046	2.15	.031*
Age	.005	.002	2.61	.008**
Amount of Urdu Input	.002	.001	1.34	.179
Maternal Education	.019	.008	2.37	.017*
Mode (Retell vs. Tell)	.030	.031	.964	.335
Language Status: Age	-.003	.002	-1.48	.137
Language Status: Mode	.034	.031	1.10	.268
Age: Mode	-.000	.002	-.096	.923
Language Status: Age: Mode	.001	.002	.787	.431

Lexical Diversity. The number of different words (NDW) was used to measure lexical diversity. Table 2.25 shows a non-significant effect of language status, demonstrating that both EM and Maj children produced a comparable number of different words. Age had a significant effect, and the older children produced significantly more NDW. There were non-significant effects of elicitation mode, amount of Urdu input, and maternal education. There was a significant interaction between language status and elicitation mode, suggesting that the effect of elicitation mode for NDW was not uniform across the two language groups. Figure 2.12 illustrates that the EM group produced more NDW during the telling tasks, and the Maj group produced more NDW during the retelling tasks.

Table 2.25 *GLMM Analysis Summary for Fixed Effects Predicting NDW*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	3.83	.023	162	<.0001***
Language Status (EM vs. Maj)	.021	.032	.648	.517
Age	.004	.001	2.68	.007**
Amount of Urdu Input	.001	.001	1.04	.297
Maternal Education	.009	.005	1.67	.094
Mode (Retell vs. Tell)	-.005	.019	-.264	.792
Language Status: Age	-.002	.001	-1.95	.050
Language Status: Mode	.043	.019	2.22	.026*
Age: Mode	.001	.001	.810	.418
Language Status: Age: Mode	.000	.001	.362	.717

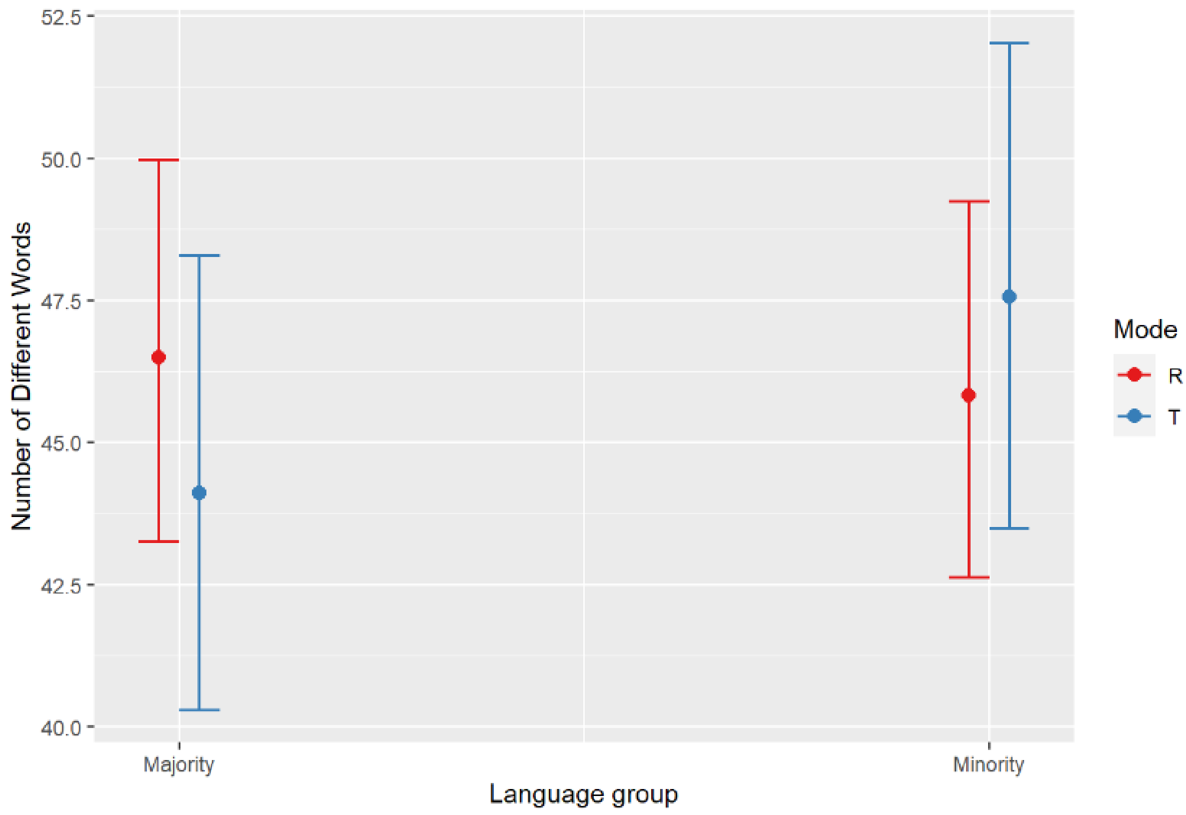

Figure 2.12 Children's NDW in Each Language Group by Elicitation Mode

Syntactic Complexity. The mean length of C-units (MLCU), the proportion of complex clauses (Cprop), and the proportion of subordinate conjunction clauses (Sprop) were used to measure syntactic complexity (Tables 2.26–2.28). There was a non-significant effect of language status for all measures of syntactic complexity. The effect of age was significant for MLCU only, and the older children outperformed the younger ones. There was a significant effect of elicitation mode only for Sprop, and it appeared that clauses with a relatively higher number of subordinate conjunctions were produced during retelling tasks. There was a significant positive effect of the amount of Urdu input for all three syntactic complexity measures, suggesting that children who received higher input in the target language produced more syntactically complex narratives. The effect of maternal education was non-significant for all three measures of syntactic complexity.

Moreover, the interaction between language status and age was significant only for MLCU, indicating that the effect of age was not uniform across the two language groups. The interaction plot (Figure 2.13), together with post-hoc emmeans pairwise comparisons, revealed that the age-based difference for MLCU was significant in the Maj group only (p = .004) and not in the EM group (p > .05).

Table 2.26 *LMM Analysis Summary for Fixed Effects Predicting MLCU*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	5.79	.706	8.20	<.0001***
Language Status (EM vs. Maj)	.160	.212	.754	.454
Age	.021	.009	2.18	.033*
Amount of Urdu Input	.022	.008	2.65	.010*
Maternal Education	.027	.037	.730	.468
Mode (Retell vs. Tell)	.010	.135	.080	.936
Language Status: Age	-.021	.009	-2.17	.034*
Language Status: Mode	.105	.135	.782	.437
Age: Mode	-.001	.008	-.212	.832
Language Status: Age: Mode	.003	.008	.462	.646

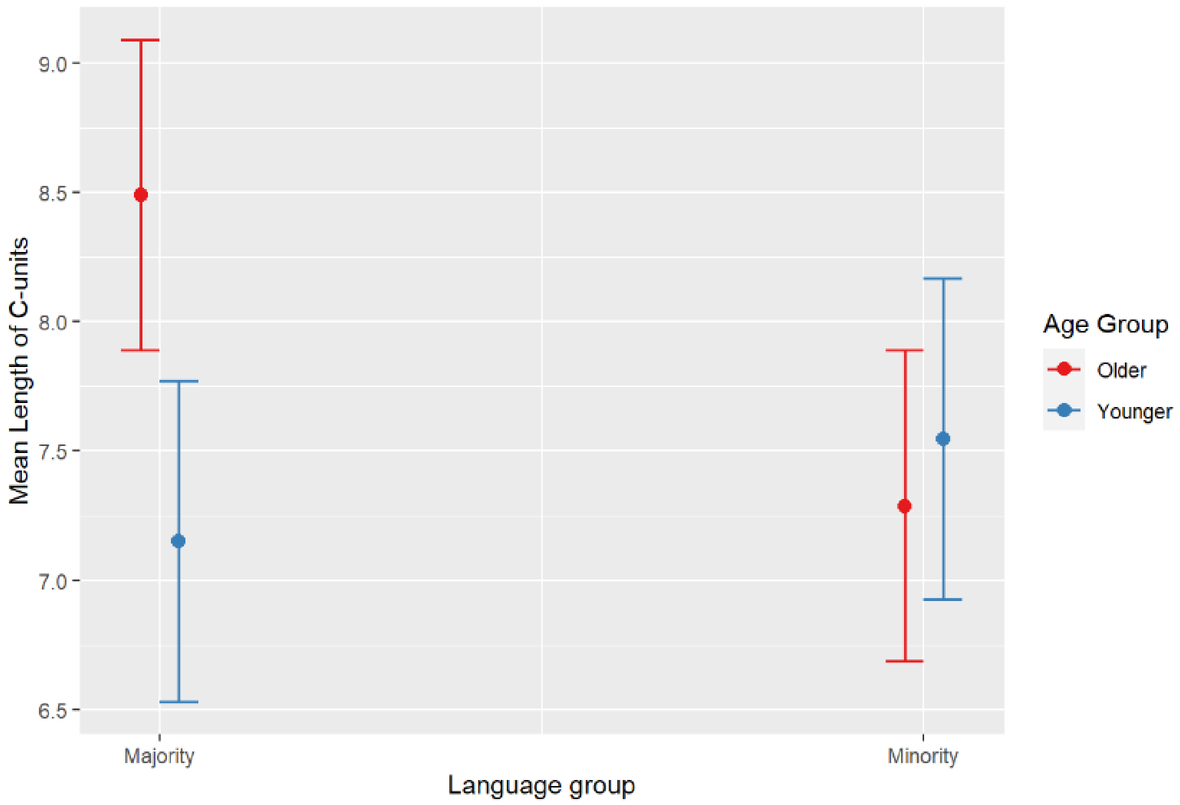

Figure 2.13 Children's MLCU in Each Language Group by Age

Table 2.27 *LMM Analysis Summary for Fixed Effects Predicting Cprop*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	.131	.067	1.93	.057
Language Status (EM vs. Maj)	.008	.020	.428	.670
Age	.001	.000	1.87	.065
Amount of Urdu Input	.001	.000	2.47	.016*
Maternal Education	.001	.003	.291	.771
Mode (Retell vs. Tell)	-.011	.012	-.959	.341
Language Status: Age	-.001	.000	-1.51	.135
Language Status: Mode	.001	.012	.108	.914
Age: Mode	-.000	.000	-.533	.596
Language Status: Age: Mode	.000	.000	1.09	.279

Table 2.28 *LMM Analysis Summary for Fixed Effects Predicting Sprop*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	-.044	.051	-.870	.388
Language Status (EM vs. Maj)	.021	.016	1.34	.185
Age	.000	.000	1.20	.233
Amount of Urdu Input	.001	.000	2.93	.004**
Maternal Education	.004	.002	1.61	.111
Mode (Retell vs. Tell)	-.023	.009	-2.42	.018*
Language Status: Age	-.000	.000	-1.08	.282
Language Status: Mode	.015	.009	1.57	.120
Age: Mode	-.000	.000	-.188	.851
Language Status: Age: Mode	.000	.000	.851	.398

Grammaticality. Two measures, namely the proportion of grammatical C-units (Gprop) and verb accuracy (VA), were used as measures of grammaticality. Tables 2.29 and 2.30 show that language status significantly affected only Gprop, demonstrating that EM children produced a significantly lower number of grammatical utterances than their Maj peers. Moreover, there was also a significant effect of age for Gprop only, and the older children outperformed the younger ones. There was also a significant positive effect of the amount of Urdu input for VA only, and the children who received more input in Urdu used verbs more accurately in their narratives. Maternal education appeared to significantly affect the production of grammatical C-units, not VA. The direction of this significant effect was negative, indicating that the children with a relatively lower level of maternal education produced more grammatical utterances. Elicitation mode had a non-significant effect on both measures.

There was a significant interaction between language status and age for VA, revealing that the effect of age was not uniform across the two language groups. The interaction plot (Figure 2.14), together with post-hoc analysis, shows that age-related differences in the Maj group were not significant ($p > .05$), and both younger and older children performed close to the ceiling. On the other hand, this age-based difference was significant for the EM group, and the older children outperformed the younger ones ($p = .028$).

Table 2.29 *LMM Analysis Summary for Fixed Effects Predicting Gprop*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	.774	.064	12.0	<.0001***
Language Status (EM vs. Maj)	-.053	.020	-2.62	.010*
Age	.002	.000	2.64	.010*
Amount of Urdu Input	.000	.000	1.16	.249
Maternal Education	-.010	.003	-3.04	.003**
Mode (Retell vs. Tell)	.002	.012	.195	.846
Language Status: Age	.001	.000	2.00	.050
Language Status: Mode	.006	.012	.533	.595
Age: Mode	-.000	.000	-.204	.838
Language Status: Age: Mode	.000	.000	.575	.567

Table 2.30 *LMM Analysis Summary for Fixed Effects Predicting Verb Accuracy*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	.944	.019	48.7	<.0001***
Language Status (EM vs. Maj)	-.002	.005	-.476	.635
Age	.000	.000	.940	.350
Amount of Urdu Input	.000	.000	2.11	.038*
Maternal Education	.001	.001	-1.70	.093
Mode (Retell vs. Tell)	-.004	.003	-1.13	.259
Language Status: Age	.000	.000	2.14	.035*
Language Status: Mode	-.001	.003	-.340	.733
Age: Mode	.000	.000	1.92	.055
Language Status: Age: Mode	.000	.000	.296	.767

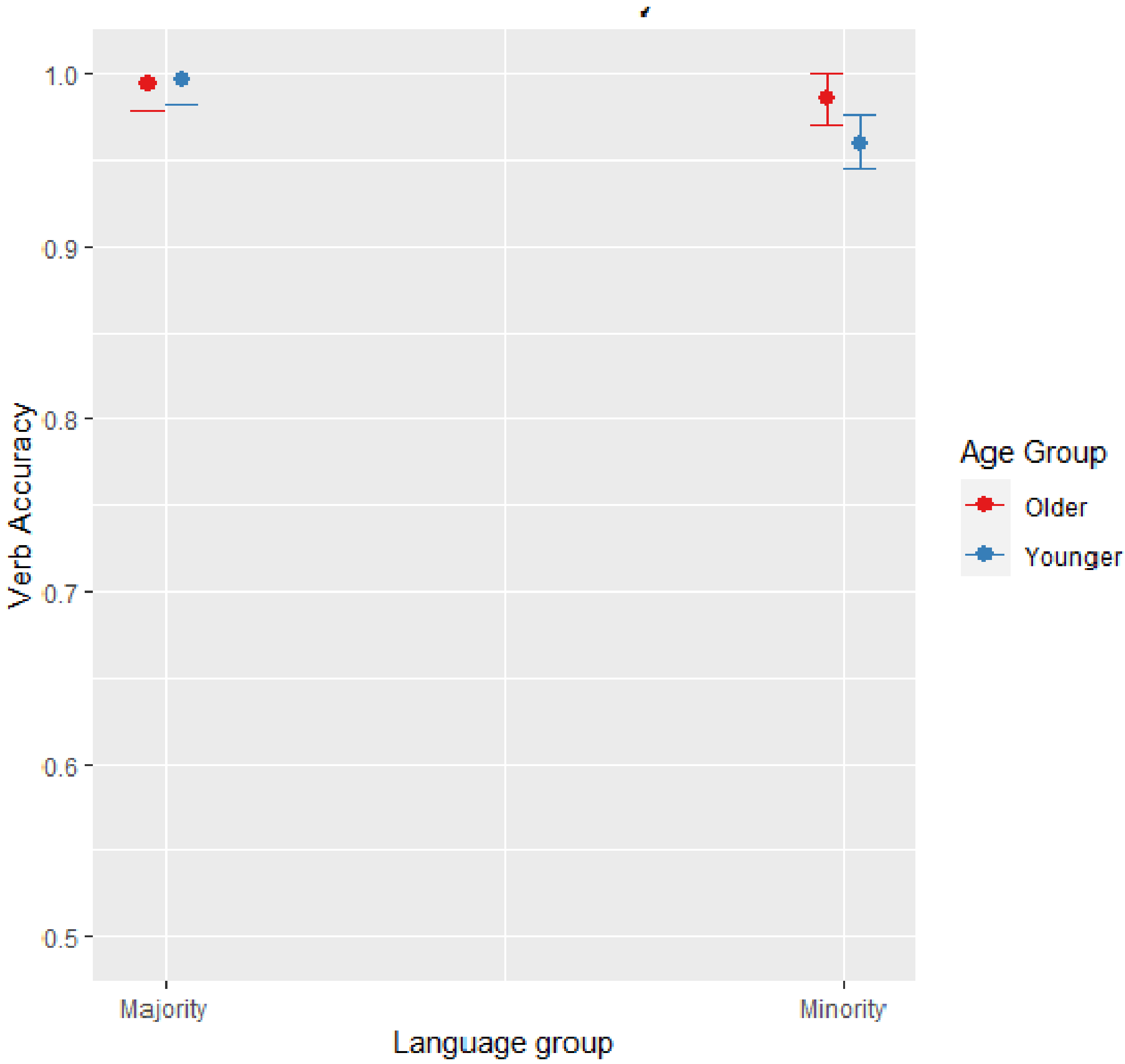

Figure 2.14 Children's VA in Each Language Group by Age

To sum up, the narrative microstructure is suggested to be more language-specific than the macrostructure, and it was anticipated that the Maj group with relatively abundant language input would outperform the EM group with restricted language input. The support for this claim was found in the results of the grammaticality measure. Significant age-based differences were observed for TNW, NDW, MLCU, and Gprop; the older children outperformed the younger ones. Significant differences based on elicitation mode appeared for Sprop, where the performance in retelling was better than telling. Significant positive and negative effects of maternal education impacting the narrative microstructure scores across different measures (TNW, TNC, and Gprop) were also observed. A significant positive effect of the amount of

Urdu input was found for syntactic complexity measures. The interaction between language status and age was significant for MLCU and VA, suggesting that, unlike the EM group, the Maj group showed more age-related progress.

Error Analyses. The results have shown that certain microstructure domains, such as grammaticality, are more sensitive to exposure. The EM children performed significantly worse than their Maj peers on this narrative measure. This section presents the findings related to language status and age-based differences in the error frequencies. This section highlights the error patterns observed in the ungrammatical productions. The initial model included language status (sum-coded; EM = "1"; Maj = "-1"), age (mean-centered), error types (sum contrast coding), and their interactions as fixed effects, and participants as random effects with elicitation mode as the random slope (Barr et al., 2013). Instead of raw frequencies, the relative frequencies of each error type were utilized in the model (see Section 2.3.5 for details). The figures, however, include raw frequencies to visualize the error distribution better.

Table 2.31 shows that language status had a significant effect, and the EM children committed significantly more grammatical errors than their Maj peers. The effect of age was also significant, and the younger children committed a significantly higher number of total errors than older children. There was a significant effect of error types. The interaction between language status and error type was also significant, suggesting that the effect of language status was not uniform across different error types. The interaction between age and error type was also significant, implying that the effect of age was not uniform across error types. The interaction between language status and age was non-significant.

Table 2.31 *GLMM Analysis Summary for Fixed Effects Predicting the Production of Grammatical Error*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	.025	.003	8.50	<.0001***
Language Status (EM vs. Maj)	.010	.003	3.40	.001**
Age	-.000	.000	-2.53	.013*
Error_type1(Number)	-.012	.002	-4.51	<.0001***
Error_type2(Gender)	.001	.002	.664	.506
Error_type3(Incomplete)	-.010	.002	-3.77	.000***
Error_type4(Word Order)	.005	.002	1.93	.053
Error_type5(Case/Postposition)	.015	.002	5.51	<.0001***
Language Status: Age	-.000	.000	-1.67	.099
Language Status: Error_type1(Number)	-.006	.002	-2.23	.025*
Language Status: Error_type2(Gender)	.007	.002	2.66	.007**
Language Status: Error_type3(Incomplete)	-.008	.002	-3.11	.001**
Language Status: Error_type4(Word Order)	.008	.002	-3.11	.001**
Language Status: Error_type5 (Case/Postposition)	.001	.003	.614	.539
Age: Error_type1(Number)	.000	.000	1.92	.054
Age: Error_type2(Gender)	-.000	.000	-1.10	.268
Age: Error_type3(Incomplete)	.000	.000	1.74	.081
Age: Error_type4(Word Order)	-.000	.000	-4.36	<.0001***
Age: Error_type5(Case/Postposition)	.000	.000	.845	.398

Figure 2.15 presents the overall distribution of errors in each category. It shows that children produced a higher number of case/postposition errors followed by word order and gender marking errors.

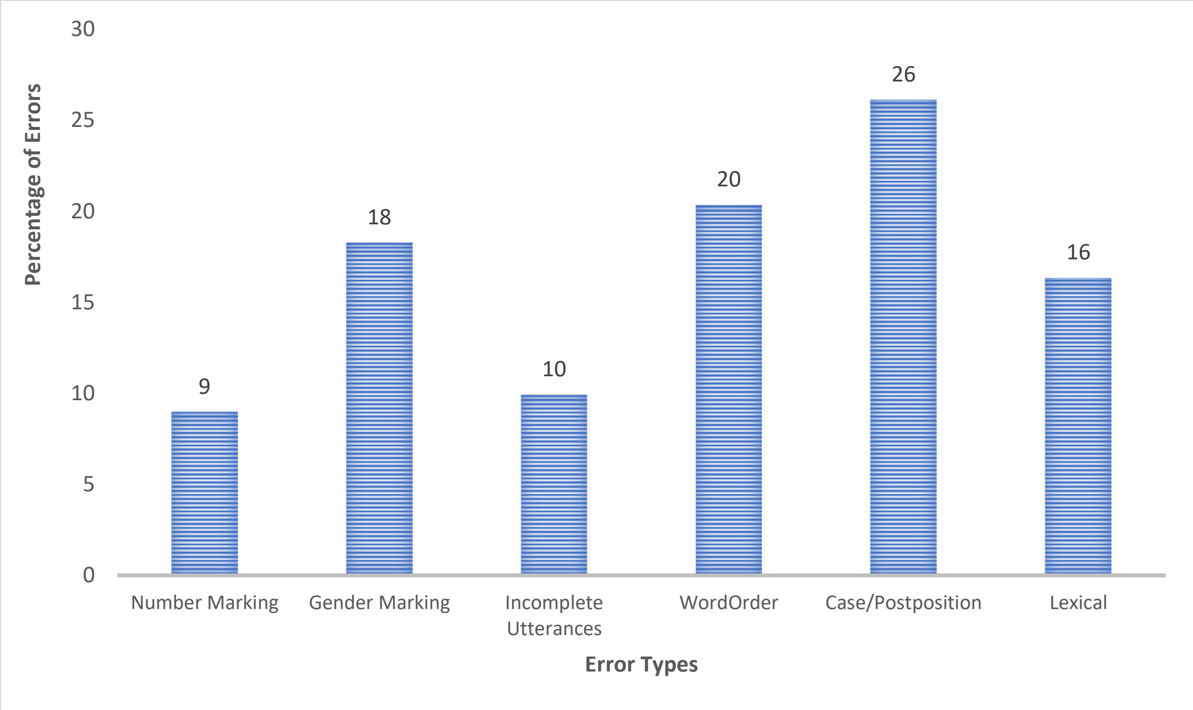

Figure 2.15 Percentage of Different Error Types

Interaction Between Error Types and Language Status. Table 2.32 presents the post-hoc analyses for this interaction, which can also be visualized in Figure 2.16. The findings show that EM children produced a significantly higher number of gender marking, word order, and case/postposition errors than Maj children. These results highlight the areas of morphosyntactic/ grammatical vulnerabilities in these EM children acquiring Urdu under reduced input conditions.

Table 2.32 *Differences in Error Frequencies in Each Error Type by Language Group*

Error Types	*β*	*SE*	*z*	*p*
Number Marking	-.008	.008	-1.011	.313
Gender Marking	-.034	.008	-4.26	<.0001***
Incomplete Utterances	-.003	.008	-0.42	.670
Word Order	-.037	.008	-4.62	<.0001***
Case/Postposition	-.023	.008	-2.90	.004**
Lexical	-.014	.008	-1.74	.083

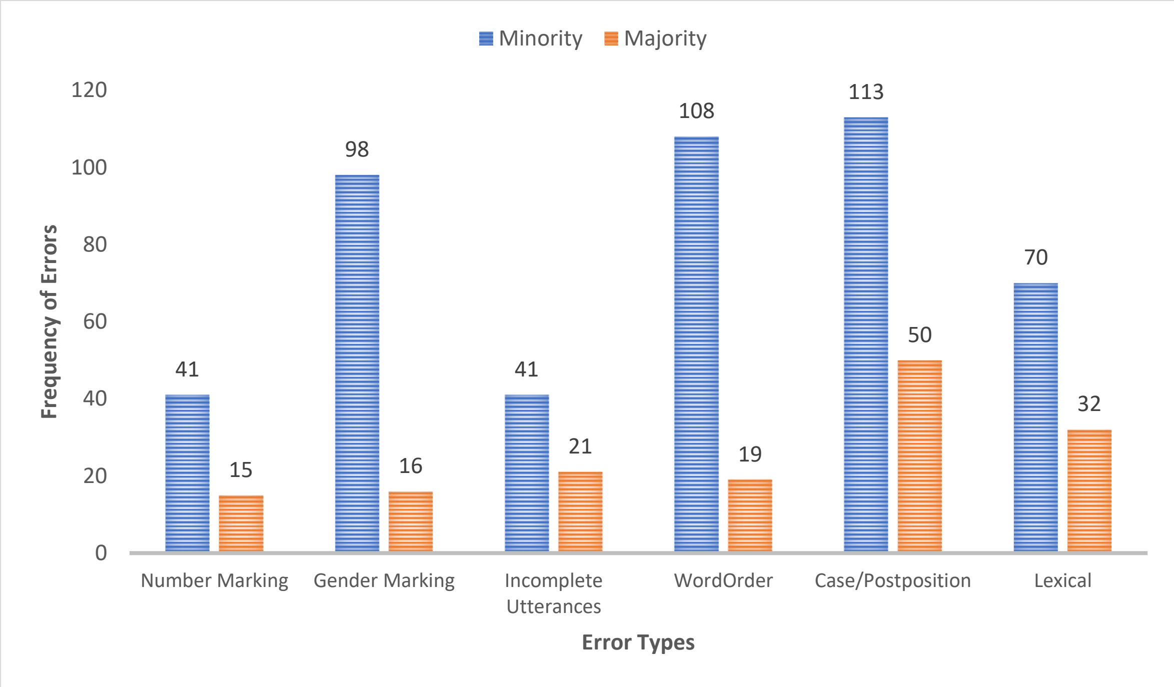

Figure 2.16 Frequency of Errors in Each Error Type by Language Group

Interaction Between Error Types and Age. Table 2.33 presents the post-hoc analyses for this interaction, which can also be visualized in Figure 2.17. The findings suggest that younger children produced a significantly higher number of gender marking and word order errors than older children, signifying that these two types of errors might be developmental in nature.

Table 2.33 *Differences in Error Frequencies in Each Error Type by Age*

Error Types	*β*	*SE*	*z*	*p*
Number Marking	-.005	.008	-.642	.521
Gender Marking	-.018	.008	-2.28	.023*
Incomplete Utterances	-.003	.008	-.383	.702
Word Order	-.041	.008	-4.99	<.0001***
Case/Postposition	-.006	.008	-.841	.401
Lexical	-.010	.008	-1.30	.193

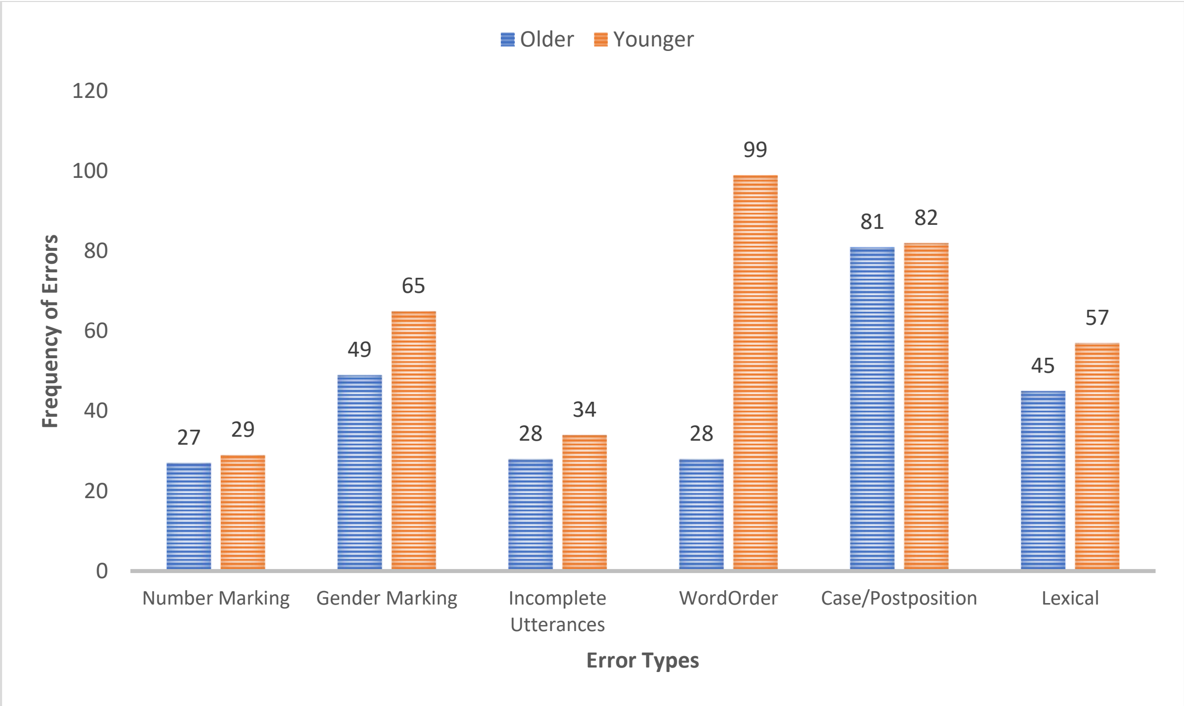

Figure 2.17 Frequency of Errors in Each Error Type by Age

Internal State Terms (ISTs). Both IST tokens (total number of ISTs used) and IST types (use of different ISTs; see Table A in Supplementary Materials) in a narrative were calculated. Table 2.34 shows a significant effect of language status, and the EM children produced a significantly greater number of IST tokens than their Maj peers. The effects of age, amount of Urdu input, and elicitation mode were non-significant. Table 2.35 demonstrates that there were non-significant effects of language status, age, amount of Urdu input, and elicitation mode for IST types. However, a positive significant effect of maternal education was found for IST tokens only and not for IST types. This means the children whose mothers had comparatively higher education levels used more IST tokens. There were no significant interactions.

Table 2.34 *GLMM Analysis Summary for Fixed Effects Predicting IST Tokens*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.86	.049	37.64	<.0001***
Language Status (EM vs. Maj)	.148	.065	2.27	.022*
Age	.005	.003	1.89	.058

Fixed Effects	*B*	*SE*	*z*	*p*
Amount of Urdu Input	.003	.002	1.35	.176
Maternal Education	.042	.011	3.77	.000***
Mode (Retell vs. Tell)	-.067	.049	-1.34	.179
Language Status: Age	-.003	.003	-1.02	.306
Language Status: Mode	.050	.049	1.01	.311
Age: Mode	.001	.003	.570	.568
Language Status: Age: Mode	-.000	.003	-.109	.913

Table 2.35 *GLMM Analysis Summary for Fixed Effects Predicting IST Types*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.14	.050	22.6	<.0001***
Language Status (EM vs. Maj)	.039	.062	.637	.524
Age	.002	.003	.836	.403
Amount of Urdu Input	-.001	.002	-.631	.528
Maternal Education	.001	.009	.167	.867
Mode (Retell vs. Tell)	-.062	.072	-.836	.388
Language Status: Age	-.004	.003	-1.45	.146
Language Status: Mode	-.012	.072	-.171	.864
Age: Mode	-.002	.004	-.490	.624
Language Status: Age: Mode	.001	.004	.327	.744

2.4.4 Correlations between Amount of Urdu Input and Different Language Measures

Table 2.36 shows the correlations between the amount of Urdu input and language performance across different language measures. Although in the mixed model tables presented above, the amount of Urdu input was included as a separate continuous variable, the effect of this predictor variable might have been modulated in the presence of other variables in the mixed models. Therefore, in the table below the nature of the relationship between the amount of Urdu input and various language measures was explored exclusively. The findings showed that there were significant positive correlations between the amount of Urdu input and children's scores for measures assessing their lexical, morphosyntactic, and expressive narrative microstructure including syntactic complexity (MLCU, Sprop, Cprop), and grammaticality (Gprop and verb accuracy) abilities. This shows that the children who received higher Urdu input also obtained higher scores for these language domains. This also signifies that these language areas might be more sensitive to the amount of language input than the others. In contrast, the association between one expressive narrative microstructure measure that is total number of C-units (length and productivity) was significant but negative. This indicates that the children who received less Urdu input produced a higher number of C-units.

Moreover, the associations between the amount of Urdu input and narrative comprehension, expressive narrative macrostructure including story structure and story complexity, and ISTs were non-significant. These findings show that these language areas might not be sensitive to the amount of language input.

Table 2.36 *Correlations between Amount of Urdu Input and Different Language Measures*

Language Measures	Amount of Urdu Input
Lexicon (CLT Scores)	.216***
Morphosyntax (SRep Stringent Scores)	.212***
Morphosyntax (SRep Lenient Scores)	.222***
Narrative Comprehension	-.056
Narrative Macrostructure	
Story Structure	-.049
Story Complexity (Maviş et al., 2016)	-.111
Story Complexity (Gagarina et al., 2019b)	-.111
Story Complexity (Sheng et al., 2020)	-.116
Narrative Microstructure	
Total Number of Words (TNW)	-.005
Total Number of C-units (TNC)	-.183**
Number of Different Words (NDW)	.094
Mean Length of C-units (MLCU)	.279***
Proportion of Complex C-units (Cprop)	.320***

Language Measures	Amount of Urdu Input
Proportion of Subordinate C-units (Sprop)	.211***
Proportion of Grammatical C-units (Gprop)	.279***
Verb Accuracy (VA)	.226***
Internal State Terms (ISTs)	-.011

Note. *p < .05; **p < .01; ***p < .001.

2.5 Discussion

The present study examined the impact of reduced input on the linguistic competence of multilingual EM children acquiring Urdu as their first/heritage language in a minority context. The effect of language status (minority vs. majority) and input quantity was studied across multiple language domains, including lexicon (comprehension and production), morphosyntax (production), and narrative (comprehension and production), in the same group of children. Different assessment tools from the LITMUS battery were employed to measure linguistic competence.

The aim was to identify EM multilingual children's areas of linguistic strengths and vulnerabilities relative to control peers acquiring L1 Urdu in a majority context. To do so, a control group of age, grade, and gender-matched Maj peers (acquiring L1 Urdu under relatively abundant conditions) from Pakistan was also included.

The overall results of this study show that the EM children lagged behind their Maj peers in different language domains, even in their strongest language (Urdu was indicated as the strongest language by the parents/caregivers). The EM children performed significantly worse in some domains, including vocabulary, morphosyntax, and narrative microstructure (grammaticality). However, the EM group's performance was comparable to their Maj counterparts across various narrative assessment measures, including narrative comprehension, expressive narrative macrostructure, expressive narrative microstructure (narrative length, lexical diversity, syntactic complexity), and internal state terms (ISTs). These findings are mainly in line with the previous literature concerning multilingual development.

The forthcoming sections first discuss the main findings, including the impact of various factors (e.g., language status, age, and input quantity) across different language domains. Then, some other findings related to domain-specific factors like modality, word category (for lexicon), sentence structure (for morphosyntax), elicitation mode (for narrative), and error analyses are presented.

2.5.1 Main Findings

Language Status Effects. This section mainly discusses findings related to the impact of language status on the performance of EM and Maj groups across multiple linguistic domains. The lexical competence was assessed at receptive and expressive levels using Crosslinguistic Lexical Tasks (Haman et al., 2015; Hamdani et al., 2020a). Results have shown that the effect of language status was significant for the lexical abilities, and the EM children performed significantly worse than their Maj counterparts. At the outset of this study, it was predicted that the EM children may perform significantly worse than their control peers on both comprehension and production lexical tasks. This prediction draws support from the non-significant interaction between language status and modality, highlighting that the EM children performed significantly worse than their Maj peers at both productive and receptive lexical levels. Some previous studies have also shown that EM multilingual children exhibit a smaller productive and receptive lexicon than their age-matched monolingual peers acquiring language in a majority context (e.g., Altman et al., 2017; Lindgren & Bohnacker, 2020; O'Toole et al., 2017; Pearson et al., 1993), even when assessed in their L1 (e.g., Haman et al., 2017; Miękisz et al., 2017; Pearson et al., 1997). Present findings also add to this line of research.

This study also assessed the morphosyntactic abilities using the Sentence Repetition task (Chiat, 2015; Hamdani et al., 2020b). The results from the most stringent condition (score was awarded if the test item was repeated verbatim) showed a significant effect of language status. As predicted, the EM children performed significantly worse than their Maj peers. This implies that when assessed in more stringent conditions, the EM children perform significantly worse than their Maj peers on grammar-related measures. These results resonate with the findings from Haman et al. (2017) and Thordardottir and Brandeker (2013), who also compared L1 morphosyntactic competence using the SRep task in children acquiring L1 in minority vs. majority contexts and concluded that the EM groups obtained significantly poor scores.

Interestingly, the effect of language status became non-significant when children were scored using a lenient scoring scheme (0–3 scores based on errors; see scoring under Section 2.3.3 for details). This illustrates that when shown some flexibility in scoring, the difference in the morphosyntactic scores of both language groups reduced to the level of non-significance. On the other hand, this also shows that a stringent scoring scheme could be a more sensitive rubric for detecting differences in children's performance from minority vs. majority linguistic backgrounds.

Furthermore, children from both language groups obtained comparable scores when assessed for maintaining the target structure (score for correct structure), where the repetitions were not penalized for lexical errors. This implies that the EM children might have acquired syntactic structures like their Maj peers. Still, the verbatim repetitions (whole item correct) were challenging based on other factors, including length and lexical items. Hence, it shows that when it comes to acquiring only syntactic structures, both EM and Maj groups obtained almost parallel scores. However, when dimensions like morphology and lexicality were considered, the accuracy rate of EM children dropped significantly. This finding also supports the claim that EM children, like their Maj peers, follow a similar developmental trajectory for the acquisition of grammar and that the rate of this development might be delayed but not deviant (Bernardini & Schlyter, 2004; Meir, 2018).

Moving on to the findings from the narrative abilities that were assessed using the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2019a; Hamdani et al., 2020c). Results have shown that children from both EM and Maj groups obtained comparable scores in the narrative comprehension task. This finding aligns with the prediction and some previous literature stating similar results. Previous studies (e.g., Rodina, 2017) that also employed the MAIN framework concluded that acquiring L1 as a minority language did not significantly undermine children's performance during the narrative comprehension tasks and that children from both minority vs. majority contexts obtained comparable narrative comprehension scores. This finding illustrates that skills like narrative comprehension, which involve interrelated cognitive competence (Bohnacker & Gagarina, 2020) and are not exclusively dependent on language-specific abilities, thus, do not put multilingual children, even from EM groups, at a disadvantage.

Narrative production was assessed at the macrostructural level via story structure (SS) and story complexity (SC). EM and Maj children obtained comparable SS and SC scores. This is in line with the prediction and lends credibility to the notion that macrostructure abilities are language-independent, and children acquiring languages under reduced vs. abundant conditions exhibit similar performances (Haman et al., 2017; Kunnari et al., 2016; Rodina, 2017). This finding could be related to the fact that generating a coherent narrative demands robust cognitive abilities essential for constructing a logical storyline that might transcend their language-specific skills (Gagarina et al., 2016). Therefore, discursive abilities at this global level appeared unaffected or less affected by the language input conditions.

Moreover, studies also suggest that there could be a carry-over of certain macrostructure components across the two languages of multilingual children, even if linguistic skills in one language are weaker (Gagarina, 2016). So, even though multilingual children might lack some language-specific skills, their overall narrative macrostructure abilities would develop interdependently in all languages (Rodina, 2017) and might even enhance each other's growth rate (Otwinowska et al., 2018).

Different measures targeting productivity, lexical diversity, syntactic complexity, and grammaticality were included to assess narratives at a microstructural level. The results revealed a significant effect of language status for productivity and grammaticality only. The EM children produced lengthier narratives than their Maj peers. The same findings were also reported by Balčiūnienė and Dabašinskienė (2019) and Blažienė (2015), where Lithuanian-English multilinguals (acquiring L1 in a minority context) produced a higher total number of words (TNW) and the total number of C-units (TNC) in their L1 than age-matched monolinguals (acquiring L1 in a majority context).

On the other hand, the EM children used significantly fewer grammatical utterances in their stories (Gprop: proportion of grammatical utterances) than their Maj peers. This result coincides with some previous findings (Gámez & González, 2019; Rodina, 2017) stating that children acquiring L1 as a minority language (EM multilinguals) exhibited significantly poor performance on the narrative measures of grammaticality than their peers acquiring it in a majority context (monolinguals). These findings signify that EM children might lag behind their Maj peers when the tasks involve skills at the more local level (syntax, morphology), such as producing grammatical utterances.

The above findings from narrative productivity and grammaticality suggest that producing lengthy narratives does not always warrant high narrative quality in terms of coherence and grammaticality. The results also support this speculation, as the EM children who outperformed Maj peers in narrative length measures produced significantly fewer grammatical utterances.

Further, syntactic complexity was assessed via the mean length of C-units (MLCU), the proportion of complex C-units (Cprop), and the proportion of subordinate C-units (Sprop). No significant differences were found between the performances of the two language groups across these three syntactic complexity measures. This indicates that the EM children could produce narratives of almost parallel syntactic complexity like their Maj peers. Some previous studies (e.g., Cahill et al., 2020; Rodina, 2017; Witkowska et al., 2022) have also reported no group differences in the syntactic complexity of narratives produced by children acquiring the target language in minority vs. majority contexts.

Furthermore, the effect of language status was also non-significant for the number of different words (NDW). Children from both groups exhibited almost similar lexical diversity. Previous studies examining the L1 lexical diversity in Spanish-English (Gámez & González, 2019) and Lithuanian-English (Blažienė, 2015) speaking children also reported that there were no significant differences in the NDW produced by children acquiring L1 in minority vs. majority language backgrounds. At the same time, these findings are not in line with the results of some other studies (e.g., Balčiūnienė & Dabašinskienė, 2019; Rodina, 2017) that concluded that the EM multilingual children, in comparison to the Maj monolingual counterparts performed significantly worse on the measures of lexical diversity in their L1 narrative productions. These two studies also used stories from the MAIN, but unlike the present study, where the NDW from both telling and retelling modes were taken into account, these studies tested performance only during telling. This highlights that data from both elicitation modes should be considered to obtain a comprehensive profile of narrative abilities in EM children, as the results from only one mode might inflate disparities between the scores of TD EM and Maj children for specific measures.

Next, the use of ISTs in the narrative productions was also assessed. In line with the prediction, language status had no effects, suggesting that children from both language groups used almost similar types and tokens of ISTs in their narratives. Kunnari et al. (2016) also reported the same findings that studied narrative abilities in the Finnish-Swedish multilinguals and their age-matched Finnish-speaking monolingual peers. They stated that the use of ISTs was not statistically different in both groups.

At the beginning of this study, it was predicted that as narrative comprehension and macrostructure production are relatively language-independent, the EM children might not be disadvantaged when assessed for these narrative skills. The results of this study supported the prediction, and the EM and Maj groups obtained comparable scores on both narrative measures. Furthermore, it was also predicted that as microstructure abilities are more language-dependent and demand language-specific skills, the EM children might perform significantly worse at this level. This prediction was only supported by the results of Gprop, where the EM children produced significantly fewer grammatical utterances.

Now, back to the first general prediction made at the inception of this study: reduced input associated with language status might affect certain linguistic abilities more than others. This prediction was supported by the results from different domains, including lexicon, morphosyntax (stringent scoring scheme), and narrative microstructure (grammaticality), as the EM children performed significantly worse than their Maj peers on these measures. Adding more to this point is the evidence from the error analyses (discussed below) indicating that the EM children committed more morphosyntactic (during narrative productions and SRep) and lexical errors. These findings highlight grammar and vocabulary as the areas of linguistic vulnerabilities in these L1 Urdu EM multilingual children acquiring L1 Urdu under constrained input conditions.

Age Effects. The children in this study were divided into two groups based on age. The children from younger (grades P1 and P2) and older (grades P3 and P4) groups had mean ages of 6;9 and 9;2 [years; months], respectively. Overall, the results showed that age had varying effects across language domains.

For the lexical abilities, the main effect of age was non-significant. However, a significant interaction between age and modality (comprehension vs. production) revealed that the gap between older and younger children varies within modalities. The post-hoc analysis further showed that the age-related differences were highly significant only at the comprehension level, and the older children outperformed the younger children. The older children obtained relatively higher scores in the production task as well, but the difference between older and younger children's scores was statistically non-significant. This finding entails that L1 lexical comprehension, compared to production, might be a more sensitive measure of capturing developmental progress in younger multilingual children.

Moreover, the results revealed significant age effects on morphosyntax across all three scoring levels, and the older children obtained significantly higher scores than the younger ones. This finding aligns with the study results of Komeili et al. (2020). They also assessed the morphosyntactic abilities in 6;3 to 11;6-year-old Farsi-English multilingual children using SRep. Their results showed a positive correlation between age and the total SRep scores, indicating an increase in morphosyntactic scores with age. Lastly, the interaction between age and language status was found to be non-significant, indicating that the effect of age was uniform across the two language groups, and the older children from both groups obtained higher scores than the younger ones.

The findings for narrative comprehension showed a non-significant effect of age, demonstrating that older and younger children in this study performed at par on this measure. This finding is in line with the previous work of Bohnacker and Gagarina (2020), who examined the narrative comprehension abilities in children aged between 2;10 to 9;9. Their results suggested that the age effects on narrative comprehension are steeper in the early years and start to taper off after 6;0. After the age of 6;6 or 7;0, the improvement rate at the group level was rather small. Bohnacker and Gagarina (2020) included findings based on children from diverse language backgrounds. The results from the present study also add to this line of research by revealing that the Urdu-speaking EM and Maj children also showed no significant age-related growth between ages 6;9 and 9;2. Both the main effect of age and its interaction with language status were non-significant.

For narrative macrostructure, the age effects were found to be significant only for SS (narrative macrostructure), and the older children obtained significantly higher scores than the younger children. The result showing non-significant interaction between language status and age implies that children from both EM and Maj groups are still undergoing developmental progress for their story-organizing abilities, even between the ages of 6;9 to 9;2. This finding concurs with the conclusions of Bamberg and Damrad-Frye (1991), Berman and Slobin (1994), Bishop and Donlan (2005), Pearson (2002), Trabasso and Rodkin (1994), and van den Broek (1997) stating that narrative abilities at the macrostructural level continue to develop even during the school-age years.

Furthermore, significant age effects for various narrative microstructure measures, including TNW, NDW, MLCU, and Gprop, were also found. In other words, the older children outperformed the younger ones in the areas of productivity, lexical diversity, syntactic complexity, and grammaticality, respectively. These results support the claim that microstructural development is more protracted and continues even after age 10 (Blankenstijn & Scheper's, 2003).

In addition to significant age effects, the interaction between language status and age was also significant for two narrative microstructure measures, including MLCU (syntactic complexity) and verb accuracy (VA; grammaticality). For MLCU, the age-related progress was evident in the Maj group only and not in the EM group. For VA, the interaction revealed that younger children in the Maj group had already reached close to the ceiling like older children. Hence, there were no significant age-based differences in the accurate use of verbs in the Maj group. On the contrary, the older children from the EM group attained significantly higher scores than younger ones.

The results further showed no age-related development for both tokens and types of ISTs. These results are in line with Gagarina's findings (2016). She examined the age-related growth among three age groups, including Russian-German children from preschool (mean age = 3;7), first grade (mean age = 7;0), and third grade (mean age = 9;2). She reported that substantial age-related progress for using ISTs was only evident between ages 3;7 to 7;0 years, and not later. In the present study, children were 6;0 and above, so following what has been reported by Gagarina (2016), the age-based differences at the group level for both IST tokens and types might have started to recede and even become non-significant after 6;0.

In the beginning, it was predicted that reduced input associated with minority language acquisition context might lead to more restricted age-related progress in the first/heritage language. This prediction was supported by the results of only two narrative microstructure measures, MLCU (syntactic complexity) and VA (grammaticality). In general, non-significant interaction between language status and age across various measures signifies that the EM children, like their Maj peers, might follow a similar developmental trajectory. This pattern could be delayed for some linguistic domains depending on various factors such as quantity and quality of language input, but it is not deviated (Antonova Ünlü & Li, 2016, 2018; Bernardini & Schlyter, 2004).

Urdu Input Effects. In addition to a categorical variable, language status (EM vs. Maj), the effect of Urdu input (mean of both school and non-school days) was also included as a separate continuous variable. The intention was to examine the effect of language input on language performance across multiple linguistic domains. As the impact of Urdu input might have been modulated when other factors were considered in the mixed effects models, the relationship between the amount of Urdu input and various language measures was also examined separately.

Even though the effect of the amount of Urdu input was not significant for lexical abilities in the GLMM model, the correlation between these two variables was highly significant and positive. The children with higher Urdu language exposure also obtained higher lexical scores. This indicates that the quantity of language input might have a significant association with children's lexical skills in multilingual children. This finding is also in line with previous studies reporting that the frequency of input in a particular language led to better lexical scores (e.g., Dijkstra et al., 2016; Leseman, 2000).

Next, the input quantity was found to impact the morphosyntactic abilities in children. The effect was close to significant for the stringent scoring and significant for the lenient scoring. The correlation analyses also showed that the relationship between the Urdu input and morphosyntactic scores at both stringent and lenient levels was positively significant. This shows that the children who received relatively higher language input demonstrated better morphosyntactic abilities. This is in line with the findings of previous studies stating that the amount of language input in a particular language is positively correlated with morphosyntactic competence in that language (e.g., Chondrogianni & Marinis, 2011; Haman et al., 2017; Hoff et al., 2018).

Moreover, the effect of the amount of language input was found to be significant for some narrative microstructure measures, namely syntactic complexity (MLCU, Cprop, Sprop) and grammaticality (verb accuracy) in the Mixed-effects models. All these expressive narrative measures and the proportion of grammatical C-units (grammaticality) were also found to be positively associated with the amount of Urdu input. This indicates that children with a relatively high amount of language input produced syntactically more complex and grammatical utterances in their stories. These results also lend credibility to the claim that the degree of input is positively associated with better language outcomes in children in terms of syntactic complexity and grammaticality (Dixon et al., 2020; Hoff et al., 2012).

Furthermore, results showed that the effect of the amount of Urdu input was non-significant for both narrative comprehension, expressive narrative macrostructure, and ISTs. The correlation analyses also confirmed these findings as the associations between these narrative measures and Urdu input were found to be non-significant. These results resonate with previous literature stating that these two narrative measures are less sensitive to language exposure and that the EM children can attain comparable scores even in their L1 that is acquired under constrained input conditions (e.g., Haman et al., 2017; Kunnari et al., 2016; Rodina, 2017). The effect of the amount of language input was also non-significant for two expressive narrative microstructure measures, including lexical diversity (NDW) and productivity (TNW). The correlations between the amount of Urdu input and these two expressive narrative microstructure measures were also non-significant.

These findings highlight that the amount of Urdu input might impact children's language outcomes differently across various linguistic domains. The effect was significant for morphosyntax and some narrative microstructure measures (grammaticality and syntactic complexity). At the same time, it was non-significant for vocabulary, narrative comprehension, and macrostructure. This finding of diverse input effects across domains draws its support from previous studies (e.g., V. C. M. Gathercole, 2002; V. C. M. Gathercole & E. M. Thomas, 2009; Montrul & Potowski, 2007) that have also concluded that the effect of language input could differ for different linguistic phenomena.

2.5.2 Domain-specific Findings

Effects of Modality on Lexical Abilities. At the outset of this study, it was predicted that the performance in the comprehension tasks might be better than production. The results have supported this prediction, and the performance during comprehension, in general, was better than production. Previous studies (e.g., Altman et al., 2017; Bialystok et al., 2010; Gross et al., 2014; Uccelli & Páez, 2007) also reported a similar pattern, and children demonstrated significantly better performance during lexical comprehension tasks. Furthermore, the interaction between language status and modality was found to be non-significant, implying that irrespective of different language acquisition backgrounds (minority vs. majority context), acquiring vocabulary at the receptive level is relatively more straightforward than at the expressive level.

Effects of Word Category on Lexical Abilities. In the beginning, it was predicted that irrespective of language status, the performance in nouns might be better than in verbs. The findings have also supported this prediction, as the interaction between language status and word category was non-significant. The children from both EM and Maj groups obtained significantly higher scores during noun tasks than verbs. This is also in line with the results of some previous studies (e.g., Altman et al., 2016; Lindgren & Bohnacker, 2020) reporting that children performed better in nouns than verbs regardless of their language status. This better performance in nouns can be attributed to the fact that as nouns, in comparison to verbs, are more meaningful, unambiguous, concrete, and less context-dependent (Altman et al., 2017), they might be relatively easy to acquire. Moreover, when lexical knowledge is assessed via pictures, this better performance in nouns can also be ascribed to the fact that it is more convenient to present nouns using static pictures than verbs (Lindgren & Bohnacker, 2020).

Effect of Sentence Structure on Morphosyntactic Abilities. There was a significant effect of sentence structures on the morphosyntactic abilities. Children demonstrated varied performance across sentence structures. In general, the ORC (object relative clauses), wh-object questions, and subordinate sentence structures were revealed to be highly challenging for children. On the other hand, sentence structures like biclausal sentences with complement/adjunct, 2 auxiliaries/1 axillary+1 modal were comparatively easier for children to repeat.

Furthermore, there was also a significant interaction between language status and sentence structures (stringent scoring scheme). This showed that the effect of sentence structures was not uniform across the two language groups. Further analyses showed that the EM children obtained significantly poor scores on four out of 10 structures (SOV with 1 axillary/1 modal, axillary/modal + negation, conditionals, and ORC). The scores between the two language groups were comparable for the other sentence structures. These results also highlight the most challenging sentence structures for the TD EM children acquiring L1 Urdu as a minority language.

Results showed that the EM children obtained significantly low scores while repeating the simple SOV with 1 axillary/1 modal sentence. Although these were syntactically simple sentences, modifications were made at the morphological level to assess Urdu nominal and verbal morphology acquisition. The EM children were found to perform significantly worse than their Maj peers, indicating that nominal and verbal morphology is one of the areas of linguistic vulnerabilities, and children with reduced language input conditions find it difficult to acquire. Several studies have also suggested that acquiring and applying morphological rules is challenging even for TD EM children (Bolonyai, 2007; V. C. M. Gathercole & E. M. Thomas, 2009; Polinsky, 2008).

Elicitation Mode Effects on Narrative Production. It was predicted that based on the support from the model story, the performance in the retelling condition might be better than telling. This expected performance gap based on the two elicitation modes was found to be significant for only one narrative microstructure measure, Sprop (proportion of subordinate C-units). Children produced more subordinate clauses during story retelling. For the rest of the measures, the effect of the elicitation mode was non-significant, suggesting that performance in both retelling and telling conditions was comparable.

This is not in line with some previous studies reporting significantly better performance during retelling conditions (e.g., Kunnari et al., 2016; Otwinowska et al., 2018). The discrepancy in findings could be attributed to the variations in the methodology of the present study and previous studies. Following the MAIN framework, Kunnari et al. (2016) and Otwinowska et al. (2018) used Cat/Dog for retelling and Baby Birds/Baby Goats for telling. While in the present study, each child produced both stories two times (depending on the counterbalance order, if the children were assigned Cat and Baby Birds, then they were asked to produce both stories in both elicitation conditions, so in total, four stories from each child). The findings from the current study show that even though the Cat/Dog and Baby Birds/Baby Goats are structurally comparable, it looks like the differences in the scores between retelling and telling conditions mentioned in these two studies emerged mainly due to the differences in the stories (Cat/Dog vs. Baby Birds/Baby Goats) and the children might have found the Baby Birds/Baby Goats story content to be more challenging than the Cat/Dog. Hence, the differences in performances during retelling vs. telling conditions could not be attributed solely to different elicitation conditions.

2.5.3 Findings from the Error Analyses

Error analyses might serve as a potent tool in examining the linguistic knowledge of individuals. The error analyses showed that for all three linguistic domains, the EM children generally committed a significantly higher number of errors than Maj peers. The findings below highlight the areas of linguistic vulnerabilities in EM children acquiring language under reduced input conditions.

During the lexical tasks, children generally made a higher number of semantic and associative confusion errors during their verb productions and semantic and na (when the children responded they do not know the name) errors during their noun productions. Moreover, the results showed a significant effect of language status, and the EM children produced a higher number of lexical errors than their Maj counterparts. There was also a significant interaction between language status and lexical error types, revealing that the effect of language status was not comparable across error types. Further analyses confirmed that children from the Maj group committed a higher number of semantic confusion errors. This means they used more words from the same semantic category, for example, guava for pear.

On the contrary, the EM children produced a higher number of other (where it was not possible to label the error type under designated categories) and na (where the children responded that they do not know the answer) errors. These findings imply that errors committed by the Maj group were semantically closer to the expected word. At the same time, the EM children made errors that were semantically distant from the target item. Thus, Maj group errors signify intact competence with performance difficulties, while EM group errors unveil the gap between lexical and semantic competence.

Even though the age effects were non-significant, there was a significant interaction between lexical error types and age, indicating that the effect of age was not parallel across error types. Further analyses revealed that the younger group, compared to the older group, committed a significantly higher number of perceptual confusion and other errors. It shows that younger children produced lexical errors either due to their misperceptions of pictures or their unfamiliarity with the words in the target language.

The effect of language status was also significant for morphosyntactic errors, and the EM children made significantly more errors than their age-matched Maj peers. Additionally, there was a significant interaction between language status and error types, suggesting that the effect of language status was non-uniform across the error types. The children from the EM group committed significantly higher morphosyntactic errors. However, the magnitude of the statistical difference varied across error types. The EM children produced a significantly higher number of word order errors, followed by postposition, number marking, and gender marking errors. Lastly, the effect of age was non-significant, implying that children from both age groups committed an almost similar number of errors and that the morphosyntax might be equally challenging for older and younger children.

The results from the narrative analyses also highlight that the EM children produced a significantly higher number of grammatical errors. There was also a significant interaction between language status and error types, suggesting that the effect of language status was not uniform across various error types. The EM children, in particular, produced a significantly higher number of gender marking, word order, and case/postposition errors during their story productions.

Furthermore, the effect of age was also significant, and the younger children committed more grammatical errors during their narrative productions. There was also a significant interaction between age and error types, implying that the effect of age was not similar across different types of errors and that younger children produced considerably more gender marking and word order errors during their narrative productions. This finding also suggests that grammatical errors, including gender marking and word order, are not only influenced by language acquisition backgrounds but could also be developmental in nature. Therefore, young L1 Urdu-speaking children, regardless of their language status, find the application of gender marking and maintenance of correct word order to be challenging.

2.5.4 Other Findings

Effects of Maternal Education. Maternal education is suggested to be a broad linguistic environment factor as it indexes higher quality and quantity of language input/exposure to children as well as cultural capital and general family environment (Prevoo et al., 2014). The literature so far reports inconsistent findings related to the effect of maternal education on different linguistic domains, especially when it comes to multilingual children's L1. Findings from this study also revealed diverse maternal education effects of varying polarities across domains.

Findings revealed a negative significant effect on these children's lexical and grammaticality (Gprop) scores. This indicates that the children with lower levels of maternal education had higher lexical knowledge and produced more grammatical utterances during their narrative productions. The possible reason for this negative association could be that when mothers have a lower level of education, they might have few languages (or even just L1) at their disposal. As a result, they may not have many options to choose between different languages during their interactions with children in general. In return, this might inflate the quantity and quality of L1 input, resulting in better lexical and grammatical abilities in children.

For the narrative measures, there was a significant positive effect on narrative productivity and IST types, suggesting that children with higher maternal education produced lengthier narratives and used diverse types of ISTs. This kind of positive effect on narrative abilities has also attested in some previous studies (e.g., Dollaghan et al., 1999).

2.5.5 Strengths and Limitations

Before concluding, some strengths and limitations of this study are discussed here. This study has highlighted the unique profiles of multilingual children acquiring their languages under reduced input conditions. Based on these findings, it is proposed that professionals should refrain from assessing these children against norms devised for monolingual or even multilingual Maj children. This could certainly lead to over or under-identification of language disorders in these children. It is also recommended that when assessing these EM children, suggestive evidence concerning the presence or absence of language impairment should be obtained from multiple linguistic domains and not just any one language area, as it is evident from the results of this study and some previous ones that even for TD EM children, some linguistic areas are of relative strength, and others are rather weak. So, conducting comprehensive assessments by examining multiple linguistic domains could help professionals make informed decisions. Moreover, the results provide beneficial insights for practitioners and parents working with the EM children related to focused interventions by elucidating their linguistic weaknesses. Lastly, based on their feasibility to be adapted into any language and free open access (e.g., MAIN), these new Urdu assessment tools might be practical for many researchers.

As the current study had a small sample size, future studies with larger sample sizes would assist in further validating the current set of findings. Furthermore, the credibility of face-to-face versus online testing using these LITMUS tools for conducting assessments and collecting reference data should also need to be investigated in future studies.

2.6 Conclusion

This study is a first attempt at analyzing the effect of language status and other participant-related factors including age, amount of Urdu input, and maternal education on the lexical (comprehension and production), morphosyntactic (production), and narrative (comprehension and production) abilities of multilingual children acquiring L1 Urdu under minority (EM group) vs. majority (Maj group) acquisition contexts. All the child participants were TD with normal nonverbal abilities and Urdu as their strongest language. Present results support some previous studies suggesting that narrative comprehension, macrostructure production, and ISTs in multilingual children are resistant to language status (minority vs. majority) and the amount of language input. Thus, when it comes to these narrative measures, the EM multilingual children acquiring their L1 under reduced input conditions are not disadvantaged. On the other hand, in line with some minority heritage language literature, the EM children were found to perform significantly worse on certain language measures, including receptive and expressive vocabulary, morphosyntax (in stringent conditions), and narrative microstructure (grammaticality), implying that these language domains could be more sensitive to language input conditions.

Furthermore, the age effects were found to be uniform across the EM and Maj groups for all linguistic domains except for two narrative microstructure measures, including MLCU (syntactic complexity) and verb accuracy (grammaticality). For these two measures, the effect of age was not uniform across the two language groups, and the EM children exhibited restricted age-related progress. Overall, these findings elucidate that in their L1, the EM children, like their Maj peers, might follow a similar development course, but this pattern could be somewhat delayed. Next, the associations between the amount of Urdu input and different language measures, including lexicon, morphosyntax, and expressive narrative measures (syntactic complexity and grammaticality) were found to be highly significant. These language domains appeared to be more sensitive to the language input. In contrast, narrative comprehension, expressive narrative macrostructure, some expressive narrative macrostructure such as NDW and TNW, and ISTs turned out to be less affected by the amount of Urdu input, as the correlations between the Urdu input and these measures were non-significant.

This study also investigated the effects of some domain-specific factors. For the lexical abilities, the performance in comprehension was better than production, and verbs appeared to be more challenging than nouns. There was also a significant effect of sentence structure in the SRep task, and children found sentences with subordination, ORC, and wh-object questions to be the most challenging. The impact of elicitation mode (telling vs. retelling) on narrative productions was insignificant for all narrative measures except Sprop, and the children produced more subordinate clauses during the retelling condition. Lastly, the error analyses showed that the children from the EM group committed a significantly higher number of morphosyntactic and lexical errors.

To conclude, for the children acquiring L1 Urdu under reduced input conditions, narrative comprehension, production of narrative macrostructure, productivity, syntactic complexity, and lexical diversity (microstructure) emerged as the areas of strength. In contrast, vocabulary and morphosyntax (in both narrative production and sentence repetition) appeared as the areas of linguistic vulnerabilities. There is a high probability that these group differences might become even more significant and convert into incomplete L1 acquisition or even language loss (Montrul, 2008) as these EM children grow older and their exposure to the L2 (majority/societal language) dominant community increases (Kohnert, 2004). Given this situation, it could be assumed that for certain language abilities, these EM children would either exhibit developmental delay or might not be able to reach the age-appropriate proficiency levels like their monolingual or multilingual Maj peers (Montrul, 2022).

Chapter 3 - Identifying Developmental Language Disorder (DLD) in Multilingual Children Using the CATALISE Diagnostic Guidelines, LITMUS Assessment Tools, and Remote Online Testing

3.1 Introduction

Globally, 7–11% of 5-year-olds are estimated to be affected by difficulties in using their first language, which can affect their everyday communication, but those difficulties are not linked to a clear biomedical etiology (Bishop et al., 2017; Norbury et al., 2016; Tomblin et al., 1997). This condition is identified as Developmental Language Disorder (DLD), which negatively impacts an individual's academic progress, increases the risk of mental health problems in adolescence, and limits career choices in adulthood if left untreated.

A long-standing challenge in identifying DLD has been differentiating the effects of language experience from impairment in multilingual children. The present study demonstrates, via a case study, that it is feasible to identify DLD in a multilingual child using the CATALISE diagnostic criteria, Language Impairment Testing in Multilingual Settings (LITMUS) assessment tools, and remote online testing. Before presenting the case study, as background first, the CATALISE diagnostic criteria, a battery of assessment tools named LITMUS battery, and their cross-linguistic evidence of promise in identifying Language Disorder (LD) in children, and some current literature on remote online testing are highlighted in the following sections.

3.1.1 Diagnostic Guidelines by The CATALISE Consortium (Bishop et al., 2017)

The CATALISE consortium provides a binary decision tree that can be followed to help decide whether a multilingual child presenting with speech, language, and communication needs might potentially be identified as having DLD. For a multilingual child who shows difficulty using language in a way that affects everyday functioning, one needs to consider whether the child is unfamiliar with the majority language, especially when it is also the target language in which the child is being assessed. If the child is unfamiliar with the majority language but competent in another language, this may reflect language needs due to insufficient exposure to the majority language. This would not be considered language difficulties that constitute a disorder. However, suppose a child who is unfamiliar with the majority language is also not competent in another (or any) language. In that case, features of poor prognosis, such as difficulties in multiple domains, especially receptive ones, may suggest language difficulties, as the language problems are likely to persist.

For children identified with language difficulties, if it is not associated with a biomedical condition, it is termed DLD under the framework of Bishop et al. (2017). Otherwise, the condition is termed LD associated with that biomedical condition. Both LD and DLD can co-occur with disorders in other domains, such as Attention Deficit Hyperactivity Disorder (ADHD) and dyslexia, or manifest differently among children in terms of the language areas that are more or less impaired. It is important to note that having low nonverbal ability (but not intellectual disability) and/or the presence of biological or environmental risk factors do not exclude a child from being identified as having DLD.

Based on the CATALISE diagnostic criteria (Bishop et al., 2017), multilingual children presenting with the following characteristics could be identified as having DLD: multilingual children (suspected of) having DLD would have demonstrated evidence in terms of negative functional impact, lack of competence even in the best language, presence of poor prognostic features, and absence of associated biomedical conditions.

3.1.2 LITMUS Battery

Recall one major consideration in the CATALISE criteria in determining whether a multilingual child has LD/DLD or not is whether “there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al. 2017: Statement 4). Addressing this consideration requires the support of assessment tools to collect objective and informative evidence from multiple language domains to guide diagnostic decisions. A battery of assessment tools named LITMUS battery developed under The European Cooperation in Science and Technology (COST) Action IS0804 (2009–2013) is a good candidate for this purpose. This LITMUS battery offers a set of cross-linguistically and cross-culturally applicable tools that allow us to test multiple linguistic domains. Numerous cross-linguistic studies have consistently shown that these assessment tools can effectively differentiate between children with and without DLD (e.g., Altman et al., 2016; Armon-Lotem & Meir, 2016; Boerma et al., 2016; Gagarina et al., 2019b; Kapalková & Slančová, 2017; Saliby et al., 2017; Tsimpli et al., 2016). These assessment tools have been adapted into numerous languages. Their cross-linguistic and cross-cultural appeal allows researchers and clinicians to assess children acquiring diverse languages growing up in different parts of the world and assess multilingual children in their multiple languages using assessment tools with parallel designs. Table 1.1 (Chapter 1) displays the number of language versions available for each tool.

3.1.3 LITMUS Tools: Cross-Linguistic Evidence of Its Promise in Identifying Language Disorder in Children

The LITMUS battery includes different tools to assess multiple linguistic domains. For example, the CLT (Haman et al., 2015) assesses lexical knowledge, the CL-NWR (Chiat, 2015) targets nonword repetition abilities, the SRep (Marinis & Armon-Lotem, 2015) examines morphosyntactic skills, and the MAIN (Gagarina et al., 2019a) is used to assess narrative abilities.

CLT for Assessing Lexical Competence. Regarding lexical competence, cross-linguistic studies using the CLT have also reported TD/DLD significant group differentiation involving multilingual children acquiring Lebanese and English or French aged 5;7 to 7;10 [years;months] (Saliby et al., 2017), multilinguals acquiring Gaelic and English aged 6;0 to 8;0 (Chondrogianni et al., 2021), and monolinguals acquiring Slovak aged 4;3 to 5;5 (Kapalková & Slančová, 2017) in expressive and/or receptive lexical abilities. The results showed that the children with DLD performed significantly worse than their age-matched TD peers at both receptive and expressive lexical levels (Chondrogianni et al., 2021; Kapalková & Slančová, 2017; Saliby et al., 2017). Moreover, at the individual level, Eikerling et al. (2023) evaluated the diagnostic accuracy of the verb comprehension subtest in the Italian version of CLT, reporting a specificity of 78% and a sensitivity of 80% in multilingual Italian-speaking children with and without DLD.

SRep for Assessing Morphosyntactic Abilities. Different studies employed the SRep to examine the morphosyntactic abilities in children with and without DLD from different language backgrounds including French-speaking children aged 5;0 to 8;0 (de Almeida et al., 2017), French-speaking children aged 5;2 to 8;9 (Fleckstein et al., 2018), French or German-speaking children aged 5;6 to 8;11 (Tuller et al., 2018), Russian and German or Turkish children aged 4;5 (Gagarina et al., 2019b), Russian and Hebrew children aged 5;5 to 6;8 (Armon-Lotem and Meir, 2016). Findings showed that the SRep could effectively distinguish between children with and without DLD (Armon-Lotem and Meir, 2016; de Almeida et al., 2017; Fleckstein et al., 2018; Gagarina et al., 2019a; Tuller et al., 2018). Some of these studies also measured the diagnostic accuracy of the SRep task in distinguishing multilingual children with and without DLD. de Almeida and colleagues (2017) reported the diagnostic specificity and sensitivity of the French SRep to be around 72% and 76%, respectively. Armon-Lotem and Meir (2016) reported the diagnostic accuracy of the L1 Russian SRep tasks to be around 81%. They also reported the sensitivity and specificity of the L2 Hebrew SRep tasks to be 100% and 89%, respectively.

NWR for Assessing Nonword Repetition Abilities. Some of the studies cited above (e.g., Armon-Lotem and Meir, 2016; de Almeida et al., 2017; Tuller et al., 2018) and others (e.g., Grimm, 2022: 8;0 –10;0-year-old monolingual and L2 learners of German) examined the NWR abilities using the NWR task. These studies reported that the NWR test could successfully differentiate between children with and without DLD. Armon-Lotem & Meir (2016) also reported the diagnostic sensitivity and specificity of the L1 Russian NWR task in detecting multilingual children with and without DLD to be around 70% and 76%, and that of the L2 Hebrew NWR task to be approximately 81% and 79%, respectively. Moreover, de Almeida et al. (2017) reported the diagnostic accuracy of the French NWR task to be around 80% in identifying children with DLD.

MAIN for Assessing Narrative Abilities. The MAIN has been used in multiple studies to examine the narrative competence of children from diverse linguistic backgrounds, including multilingual children acquiring Russian and German or Turkish aged 4;5 (Gagarina et al., 2019b), Russian and Hebrew aged 5;6 to 6;6 (Fichman et al., 2017), English and Hebrew aged 5;3 to 6;5 (Altman et al., 2016), Greek aged 9;0–9;1 (Tsimpli et al., 2016), and Greek and Albanian aged 6;0 to 8;0 (Peristeri et al., 2020). All these studies reported that children with DLD scored significantly lower in narrative comprehension or production than their age-matched TD peers (Altman et al., 2016; Fichman & Altman, 2019; Fichman et al., 2017; Gagarina et al., 2019b; Peristeri et al., 2020; Tsimpli et al., 2016). Moreover, Boerma et al. (2016) examined the diagnostic accuracy of the Dutch version of MAIN. In the monolingual group, the MAIN was able to identify a total of 85% of the children with language disorder (sensitivity) and 79% of the children with typical development (specificity). The reported sensitivity and specificity in the multilingual group were 79% and 88%, respectively.

3.1.4 Remote Online Testing

While the feasibility of providing virtual assessments for families residing in remote and rural areas has been established for some time (Coleman et al., 2015; Sutherland, 2016, 2017; Wright, 2018), the outbreak of COVID-19 in late 2019 transfigured this understudied and alternative mode of delivery into a pressing need. Sudden lockdown orders disrupted in-person activities, including face-to-face testing, implying that remote online testing was no longer seen as an alternative mode of delivery but the only available mode for conducting assessments during this extraordinary period (Nelson & Plante, 2022).

Lately, there have been some efforts to evaluate the efficacy of telepractice for conducting diagnostic assessments to identify children with and without language disorders, and the results are promising. For example, Pratt et al. (2022) compared the narrative comprehension and morphosyntactic skills in 10 adult-child dyads in in-person and virtual modes. Children were aged between 4;1 to 8;5. Narrative comprehension and morphosyntactic skills were assessed using the MAIN (Gagarina et al., 2019a) and the Bilingual English–Spanish Assessment (BESA; Peña et al., 2018), respectively. The findings indicated a significant correlation between the virtual and in-person narrative comprehension and morphosyntactic scores. In a different study, Nelson and Plante (2022) compared the Test of Integrated Language and Literacy Skills (TILLS) results when administered in person and online. A total of 51 participants with and without language or literacy disorders aged between 6;0 and 18;0 years were tested using different subtests of the TILLS. The results yielded 96% agreement between the two testing modes in identifying children with and without language or literacy disorders. In another study, Eikerling et al. (2023) validated the remote screening of DLD in multilingual Spanish-Italian-speaking children using a new web-based application called MuLiMi. Their study recruited 36 4;0 to 6;0-year-old Spanish-speaking preschoolers; 16 of them were already diagnosed with DLD. Language abilities in L1 and L2 were assessed using dynamic tasks (novel word learning) and static tasks (verb comprehension, grammaticality judgment, and nonword repetition). They reported significant correlations between the screening task scores and outcome measures from parents, SLT, and teacher questionnaires, and also screening task scores and standardized test scores. This provided supporting evidence that the Spanish-Italian MuLiMi application has the diagnostic potential to identify children at risk for DLD.

3.2 Present Study

This study aims to offer a tutorial with a case study to illustrate the feasibility of identifying DLD in multilingual children using the CATALISE criteria in combination with LITMUS tools via remote online testing.

Like many other places in the world, Hong Kong has a growing population of multilingual ethnic minority (EM) children, with Pakistanis as a significant minority population. The size of the Pakistani population in Hong Kong has increased from 18,094 in 2016 to 24,385 in 2021, constituting 24 percent of the total South Asian people residing in Hong Kong (Census and Statistics Department, Government of the Hong Kong SAR, 2021). Many Pakistani children in Hong Kong acquire Urdu as their home language and first language (L1) and Cantonese as a second language (L2) in their school and community. SLTs have increasingly encountered multilingual Pakistani children in recent years. As they are acquiring both Urdu and Cantonese under reduced input, they may lag behind their monolingual peers in each language. As a result, they might face difficulties in accessing education in schools where they need L2 (Cantonese), as well as less input to support the development and maintenance of L1 (Urdu) as their heritage language (e.g., Bosch & Sebastián-Gallés, 2003, Paradis, 2010). However, there are obstacles in providing adequate and appropriate language support, as SLTs and educators in Hong Kong know little about the speech and language development of these children, including L1 Urdu and its influence on the development of the L2 and community language Cantonese (e.g., Cummins, 2000; Gu & Patkin, 2013). In addition, speech and language assessment tools are not available for Urdu, and tools for Cantonese only have monolingual norms and may not apply to this multilingual population. Therefore, language disorders cannot be reliably identified in these children- a common global issue faced by researchers and clinicians when assessing multilingual EM children (Armon-Lotem, 2018; Bedore & Peña, 2008). In order to improve the differentiation of Urdu-Cantonese children in Hong Kong with and without DLD, the LITMUS tools have been adapted to Urdu and Cantonese by our research team.

This study demonstrates how online adaptations of the LITMUS tools can provide informative data when we identify potential DLD in a multilingual child in light of the CATALISE diagnostic criteria, using Urdu-Cantonese multilingual EM children in Hong Kong as an illustrative example. Notably, the online versions of these tests have the potential to support the identification of multilingual DLD in any target language in terms of global impact. This study mainly discusses the conceptual and methodological considerations in the context of a case study, presented as a proof of concept that could pave the way to more wide-ranging research that would have positive societal implications in other parts of the world.

3.3 Methods

3.3.1 Participants

The participant in this case study identified as potentially having DLD (suspected DLD, S-DLD child) was a female aged 6;8 at the time of testing. She was referred to the in-house speech therapy clinic of The Hong Kong Polytechnic University by her schoolteacher, who expressed concerns about her language abilities. She was born in Hong Kong, and her mother reported typical pregnancy and delivery. She had no significant biomedical history, and her mother reported typical motor, emotional, and social development. Although Urdu and Punjabi were reported as family/home languages, Urdu was reported as being used more frequently and the participant's strongest language. The participant also spoke English and Cantonese. Cantonese, the societal and school language, was reported as the weakest language (as per mother's ranking). She started schooling at the age of 3;6, where the medium of instruction was Cantonese. Her mother expressed concerns related to speech intelligibility, memory when related to academic tasks, and learning ability. The school also expressed concerns about the participant's language abilities, which affected everyday social interactions and educational progress. However, her Raven's Progressive Matrices standard score of 91 suggested no intellectual disability.

The other participants considered TD in this study included seven Urdu-Cantonese multilingual EM age- and grade-matched peers (three females and four males; TD-1 to TD-7) who were shortlisted and recruited from Study 1. They were aged between 6;1 and 7;4 (mean age = 6;6, SD = 4 months). Four out of seven children were +/- 3 months, two were 7 months younger and scored better than the S-DLD child, and one was older by 8 months but was grade-matched. Although this sample size is relatively small, this is common in language acquisition studies on multilingual ethnic minority children (e.g., Farndale et al., 2016; Hu et al., 2014; Serrano-Hidalgo, 2018).

All these participants, like the S-DLD child, were enrolled in public primary schools with Cantonese as the medium of instruction and were studying at the same grade level (P1). All the participants achieved the Raven's Progressive Matrices standard scores of above 75, indicating no intellectual disability (mean = 92.43, SD = 9.13). All participants spoke Urdu as their strongest and home language as ranked by the caretakers (mothers), consistent with information on the amount of language exposure in the parental questionnaires (Table 3.1 provides information related to each participant's language exposure). All participant families, including the S-DLD child, have comparable socio-economic status (SES; low-middle). Education and occupation of the parents were used as a proxy of SES (e.g., Antonijevic-Elliott et al., 2020; Lindgren & Bohnacker, 2020).

One participant (TD-1) had suspected speech sound disorder based on her substitution errors in speech that were not age-appropriate, despite no parental concerns about her language development and her performance in the other receptive and expressive language tasks comparable to the other age- and grade-matched peers. Some of her errors were inconsistent and were marked as incorrect, but the error of substituting retroflex sounds /ʈ, ɖ/ with dental sounds /t̪, d̪/, respectively, was relatively consistent and was not penalized following the standard scoring principle for giving allowance to constant substitution errors.

Table 3.1 *Information on L1 Urdu Exposure and Relative Proficiency of Languages in the Participants as Reported in Parental Questionnaires*

Participant	Weekly L1/Urdu Language Exposure (in hours)	No. of Languages Each Child Uses in Daily Routine (Ranked based on proficiency by mothers; 1 indicates the best language)
S-DLD	69	1. Urdu 2. Punjabi 3. English 4. Cantonese
TD-1	80	1. Urdu 2. Cantonese 3. English
TD-2	83	1. Urdu 2. English 3. Cantonese
TD-3	64	1. Urdu, Pashto 2. Cantonese 3. English
TD-4	25	1. Urdu, Mirpuri 2. English 3. Cantonese
TD-5	90	1. Urdu 2. English

Participant	Weekly L1/Urdu Language Exposure (in hours)	No. of Languages Each Child Uses in Daily Routine (Ranked based on proficiency by mothers; 1 indicates the best language)
		3. Pashto 4. Cantonese
TD-6	72	1. Urdu 2. Cantonese 3. English
TD-7	37	1. Urdu 2. Punjabi 3. English 4. Cantonese

3.3.2 Instruments

The Urdu versions of the following tests were used: the Crosslinguistic lexical tasks (CLT, Hamdani et al., 2020a), the Crosslinguistic Nonword Repetition Test (CL-NWR, Hamdani et al., 2020d), and the Sentence Repetition task (SRep, Hamdani et al., 2020b), and the Multilingual Assessment Instrument for Narratives (MAIN, Hamdani et al., 2020c).

Crosslinguistic lexical tasks (CLT). The CLT assesses lexical abilities and includes production and comprehension modes, with 32 nouns and 32 verbs assessed for each mode (including two trial items per mode). Participants are asked to name the illustration on the screen orally for the production mode. For the comprehension mode, participants are asked to select the illustration matching the audio stimuli from four options by saying the number of the selected illustration, with the position of the target illustrations varied across trials. The target words, balanced for the semantic type and level of difficulty, are chosen from a pool of illustrations standardized for all language versions. The semantic types for noun stimuli include animate natural objects, inanimate natural objects, and artifacts, while the verb stimuli include actions performed by humans, animals, and states/unintentional actions. The difficulty level is determined by the complexity index, which depends on grammatical and phonological features of the lexical item, and the age of acquisition of the target word, which is determined by ratings from 25 native speakers of Urdu. Regarding the order of presentation, tasks of the same mode are presented together, with the order of mode and order of noun vs. verb counterbalanced across participants. The order of presentation for trials within a sub-task is fixed.

Crosslinguistic Nonword Repetition Test (CL-NWR). The CL-NWR, as the name indicates, is used to assess nonword repetition abilities. Difficulty with repeating nonwords has been put forward as a clinical marker of DLD cross-linguistically (Schwob et al., 2021). The CL-NWR task is presented as a game in which the participant is asked to fix a necklace by repeating the stimuli (Polišenská & Kapalková, 2014). Stimuli are 2–5 syllables long, with four items per length. They are selected from a pool of 4–6 candidates per item that are matched in length (allowing selection of different alternatives for different language versions in case one or more of the candidates is a real word in the language). An additional criterion was that the component phonemes should be acquired by Urdu-speaking children between 3;0 to 3;6 years. (Ambreen, 2023; Bari & Ajmal, 2016; Zahra, 2016). The stimuli are presented in one randomized order. Two training stimuli are also administered before the task begins.

Sentence Repetition Task (SRep). The SRep is used to measure morphosyntactic abilities. The task is presented as a treasure-hunting game, where the participant is requested to repeat sentences to help a bear find a treasure. There are three blocks of stimuli corresponding to three levels of syntactic difficulty. The length of the stimuli in each block is shown in Table B (Supplementary materials). A total of 10 syntactic structures are tested, with three in the 'language-specific' condition and seven in the 'language independent' condition, which is considered challenging for children with DLD in the target language and across languages, respectively. The target structures were chosen considering the structures tested in the original English SRep task (Marinis & Armon-Lotem, 2015), crosslinguistic literature on DLD regarding vulnerabilities, the typological characteristics of Urdu, and clinical experiences with Urdu-speaking children having language disorders. Each structure (except one) is evaluated using three sentences, with nominal and verbal inflectional morphology signaling number, gender, and tense controlled within items across structures. One structure (SOV with 1 auxiliary/modal verb) has eight sentences. The intention was to assess nominal and verbal morphology (number and gender agreement and case marking), which are vulnerable in children with language disorders, more comprehensively via this simple structure. The task is presented in three randomized orders. Four training stimuli were also administered before the task began.

Multilingual Assessment Instrument for Narratives (MAIN). The MAIN is used to assess narrative abilities and consists of narrative telling, retelling, and comprehension tasks. The instructions for telling and retelling are delivered through videos that show the hands folding and unfolding the story pictures, with recorded audio narration. The narrator is shown as a cartoon character of a girl, who emphasizes that only the participant can see the pictures. For the retelling mode, the whole recorded story script is played to the participants before they are prompted to retell the story. For each story, up to 10 questions for the comprehension mode are presented in live voice because the wording of the questions depends on the participants' responses to previous questions. Specific relevant pictures are highlighted in turn by special effects of a red-colored frame and "pulse" animation in PowerPoint for each comprehension question.

In this study, participants first listen to a model story script with picture support and have to retell the story with picture support. Then, they answer ten standard comprehension questions about the story. This procedure is repeated for another story to collect narrative data from two stories. After completing some other language tasks as a "gap", they would be asked to tell the two stories again with the same picture support (but without prior listening to a story script). That is, each child produces a total of four narrative samples, two during retelling and two during telling.

Raven's Progressive Matrices. A non-verbal IQ test (Raven's Progressive Matrices, Raven, & Court, 1998) was administered to participants to screen out intellectual disability. In the Raven's Progressive Matrices, the participants answer 60 multiple-choice questions, where they have to identify the missing piece that completes a pattern from six to eight options. A standard score of 70 or above is considered within the normal range.

Parents of Bilingual Children Questionnaire (LITMUS-PaBiQ; Tuller, 2015). The Urdu version of the LITMUS-PaBiQ was also administered before the assessment session(s) during the recruitment phase. The Ph.D. student, an experienced Urdu-speaking SLT, contacted the caregivers (mothers) of the participants and filled in the questionnaire over a phone call.

The LITMUS-PaBiQ (Tuller, 2015), available for free, collects data about the participant's language environment, language exposure, receptive and expressive competence in each language, frequency of usage of different languages in daily routines, early language, and other developmental history, and family history related to speech, language, and hearing issues. This parental questionnaire also gathers information about the parent/caregiver's education level, occupation, language competence in different languages, and frequency of usage in each language with the child.

The LITMUS-PaBiQ can help obtain important information about a child's developmental language history, language exposure, language use, and the presence or absence of biomedical conditions. This information could help clinicians determine whether a child's (poor) linguistic competence is associated with certain risk factors related to LD/DLD, factors related to multilingualism, or any significant biomedical condition. This also ensures that a parent/caregiver can share their observations and concerns about a child's speech and language difficulties with the researcher/clinician in a structured way. The LITMUS-PaBiQ has been reported to be informative in differentiating children with and without LD or DLD or in predicting LD or DLD in monolingual and bilingual children in several cross-linguistic studies (Boerma & Blom, 2017; Hreich & Messara, 2013; Tuller et al., 2018).

3.3.3 Online Testing Procedure

An Urdu-speaking SLT (the Ph.D. student) conducted all the testing. The LITMUS tests were computerized versions adapted for remote testing, which were shown to the participants via the 'Share Screen' function of Zoom. The testing procedures of the original tests designed for face-to-face testing were followed and only numbers were added to each picture in the CLT tasks to facilitate online testing (i.e., instead of pointing at pictures, participants were asked to say the numbers added to the pictures). Standard headphones, equipped with a microphone, were mailed to the participants before the testing session. This assisted in the standardization of audio quality. The participants were offered to keep these headphones as a token of appreciation for their participation. The experimenter also used the same microphone for each testing session. In addition, participants were required to use a computer, laptop, or tablet in a quiet environment with stable Wi-Fi and with only an accompanying parent or guardian (testing in public was not recommended). They were requested to switch on their cameras during testing, which took place at the participants' homes. Most testing sessions were completed in one round (around 60–70 minutes), but if required to suit children's attention, it was divided into two rounds.

The parents/caregivers were requested not to instruct or interrupt the participants during the testing session. They were further advised to use the highest volume level of the headset, although the volume could also be controlled by the computer. Therefore, the exact volume level was not standardized across participants. The volume level was controlled with the following arrangement: at the outset, the participants were asked to indicate whether the speech sounds they were hearing through the headset were loud and clear enough on their side. Moreover, some simple practice trials were conducted before moving on to the test items in each of the non-word repetition, sentence repetition, and cross-linguistic lexical tasks to obtain some objective evidence that they could hear the auditory stimuli clearly. Each session began with some warm-up questions. The sequence of testing mainly remained the same for all the participants. Assessment session(s) were sequenced as follows:

1. NWR
2. SRep
3. retelling and comprehension parts of MAIN
4. two sub-tasks of CLT
5. tell the two MAIN stories again without support from a model story
6. two sub-tasks of CLT
7. the test for nonverbal IQ

This order of task presentation was chosen based on the following considerations: Tasks 1 (NWR) and 2 (SRep) were in a game-like format and serve as a good base for building rapport with children. In addition, these two repetition tasks also provided a good warmup for making these children talk before a more expressive and linguistically demanding task like story retelling. The children in this study were required first to retell (with support from the model story) and then tell (without any support from the model story) the same stories. Therefore, a break was required between Task 3 (story retelling) and Task 5 (telling). For this purpose, and also following the CLT authors' prescribed method of administration (i.e., Two tasks for the same part of speech (comprehension and production) should not go as consecutive sessions. A break after each session or after at most two sessions was required), Task 4 (i.e., between story retelling and telling) and Task 6 (after story telling) were decided to be either CLT production or comprehension sub-tasks depending on the counterbalancing order. Lastly, Raven's Colored Progressive Matrices (Task 7) was chosen to be at the end, thinking that if the session gets too long for the children, it could be relegated to the new test session.

3.3.4 Scoring

The following scoring conventions were employed for each test:

- For the CLT, scores are calculated excluding the trial items. There are four sub-tasks (noun comprehension, verb comprehension, noun production, and verb production), each having 30 items. Participants are given one score for each correct item.
- For the CL-NWR, there are 16 test items in total, excluding the training items. The maximum score is 16 for whole items and 56 for segments correct.
- In the SRep, there are 35 test items, excluding the trial items. Participants get one score for producing the test sentence identically. The maximum score for this task is 35.
- For the comprehension task of the MAIN, participants answer 10 questions per story at most, with 1 point per question. As participants tell two different stories in Urdu, the maximum score is 20. For the production task, scores for storytelling and retelling are calculated separately. For each mode, participants are given scores on Story Structure (SS, max = 17 per story). Participants are given one score for each of the 16 scoring items, apart from the first one, where participants can get a maximum of 2 points. Scoring items for SS include the setting for the story (Time and Place reference) and Internal State Terms (IST) as initiating event[2], Goal (G), Attempt (A), Outcome (O), and IST as a reaction for each of the three episodes.

[2] "Initiating Event" has been commonly considered as a story grammar element in the story grammar framework of analyzing narrative macrostructural competence in the literature (e.g., Stein & Glenn, 1979). "Initiating Event" refers to an event that triggers/initiates the intentionality of a story character (in the form of a "Goal" (another story grammar element)) which motivates the character to carry out a goal-directed action (manifested as an "Attempt" (another story grammar element)). An "Initiating Event" can be expressed with or without the use of an Internal State Term (IST), but according to the MAIN scoring criteria, it adopts a more stringent scoring criterion for the story grammar element "Initiating Event" that an utterance expressing an Initiating Event with the use of an IST (e.g., Baby goat was **scared**, or the mother Goat **saw** that the baby goat was **scared**, or the mother Goat was **worried** about the baby goat in the water) would only be scored one point, and it therefore uses the phrase/term "IST as initiating event" in its scoring manual.

Story Complexity (SC) counts the number of sequences or episodes in terms of AO, GA/GO, GAO, and single G. Since there is currently no gold standard in terms of the scoring approach one should use that best captures the DLD/TD differences in story complexity. This study has chosen three scoring schemes that have been used in cross-linguistic studies for which score weighting assignment aligns with the level of story structure complexity based on the binary decision tree by Westby (2005). It represents three levels of granularity in terms of structural complexity in macrostructure. Among them, the scoring scheme by Gagarina et al. (2019) has the finest level of differentiation (6 levels), the scoring scheme by Sheng et al. (2020) has the coarsest level of differentiation (2 levels), and the scoring scheme by Maviş et al. (2016) is in the middle (4 levels). For further information, see Table C in Supplementary Materials. Furthermore, the scoring schemes by Gagarina et al. (2019) and Sheng et al. (2020) have both successfully captured significant DLD/TD group differences in story complexity.

To the best of my knowledge, no studies have examined and compared how different scoring schemes that differ in the level of granularity of differentiating between the different levels of story structural complexity might affect (the degree of) differentiation between TD children and those with DLD. It is, however, interesting and essential to address this methodological issue as it might affect diagnostic accuracy. Since different scoring approaches may vary in their ability to capture the DLD/TD gap, this tutorial chose to use all three scoring schemes for a more informative comparison. See Table D in Supplementary Materials for an example of parts of a narrative sample to illustrate the scoring points.

Furthermore, the use of ISTs was also assessed. ISTs refer to terms that describe the internal states of a character, mainly referring to their mental states and feelings, such as emotions, thoughts, intentions, and reactions (Gagarina et al., 2015). Table A (Supplementary Materials) presents some examples of ISTs in English and Urdu according to the subtypes of ISTs listed in the MAIN manual. This study scored the use of ISTs in both token and type measures. Token measures count the number of individual words in a narrative sample (i.e., also counting the repetitions of individual words). In contrast, type measures count the number of unique word forms (i.e., repetitions of an individual word will only be counted once).

3.3.5 Interrater Reliability

For the MAIN, CLT, and SRep, all the data were transcribed by a part-time research assistant (an Urdu-speaking SLT) and double-checked by the Ph.D. student (also an Urdu-speaking SLT). All the scorings for these three measures were completed by one rater and then double-checked by the second rater (both Urdu-speaking SLTs) to ensure that all the utterances and items were scored correctly, not left unscored or scored twice. The initial percentage of agreement between the two raters for different tasks was as follows. Regarding the CLT, all the responses in the four sub-tasks (N-C, V-C, N-P, V-P) were scored independently by the two raters (total 1024 items). There was disagreement over three items (V-C task). The percentage of agreement between the two raters was 99.7% (1021 over 1024 items). For the SRep, all the utterances (35 items) from the 8 participants were scored independently by the two raters (total 280 items). The percentage of agreement between the two raters was 99.6% (279 over 280 items). For the MAIN, the narrative samples were scored for narrative comprehension (2 stories) and narrative macrostructure production (4 stories per child). The percentage of agreement between the two raters for the narrative comprehension questions was 100% (160 over 160 items). The agreement across different macrostructure measures was as follows: SS = 99.6% (542 over 544 items), SC-M = 99.6% (287 over 288 items), SC-Sh = 100%, (96 over 96 items), SC-Ga = 99.8% (575 over 576 items). There were very few minor disagreements due to occasional omissions and typos, which were quickly resolved to 100 % after double-checking.

For the CL-NWR, two raters independently transcribed all the data. Both raters were Urdu-speaking SLTs and had relevant training in phonology. Out of a total of 128 items (16 target items per participant, 8 participants in total), there was disagreement on nine items where both raters perceived and transcribed the phoneme differently and the disagreement was not resolved by discussion. In this case, a third rater who was also an Urdu-speaking SLT and was working in the field of Urdu phonology was invited to transcribe the discrepant items independently. The transcription for the discrepant items was finalized based on the dominant number of votes (choosing the transcription with 2 out of 3 votes). Overall, the percentage of agreement between the two raters was 93% (119 items over 128 items).

3.4 Results

This section presents results for different language domains assessed using all four measures.

3.4.1 Lexical abilities

Table 3.2 shows the scores for CLT. The S-DLD child scored lower scores in different sub-tasks compared to her peers. In comprehension, she scored a total of 28, with 16 in nouns and 12 in verbs. For production, she scored a total of 16, with 9 in nouns and 7 in verbs. Overall, her age-and-grade-matched peers achieved higher scores for nouns and verbs in the comprehension mode, with an average of 25.14 (SD = 2.73; Range = 21–27) out of 30 in nouns and 25.29 (SD = 2.06; Range = 23–28) out of 30 in verbs. In the production mode, scores in nouns (mean = 17.57, SD = 6.16; Range = 10–26) out of 30 and verbs (mean = 14.71, SD = 2.43; Range = 11–18) out of 30 were also higher.

Table 3.2 *Scores Representing Lexical Abilities Assessed Through CLT Tasks*

Participant	Comprehension			Production		
	Total (Max. Score 60)	Noun (Max. Score 30)	Verb (Max. Score 30)	Total (Max. Score 60)	Noun (Max. Score 30)	Verb (Max. Score 30)
S-DLD	28	16	12	16	9	7
TD-1	46	21	25	28	12	16
TD-2	52	25	27	25	10	15
TD-3	57	29	28	37	22	15
TD-4	47	24	23	24	12	12
TD-5	46	23	23	38	20	18
TD-6	54	27	27	42	26	16
TD-7	51	27	24	32	21	11
Mean-TD	50.43	25.14	25.29	32.29	17.57	14.71
SD-TD	4.28	2.73	2.06	6.95	6.16	2.43

3.4.2 Nonword repetition abilities

Table 3.3 presents the CL-NWR scores. Overall, the S-DLD child scored the lowest compared to all other participants (including TD-1), with 3 when measured by item and 17 when measured by segment. On average, her age- and grade-matched peers scored 7.00 (SD = 1.53; Range = 5–9) out of 16 when measured by item and 42.00 (SD = 3.46; Range = 37–47) out of 56 when measured by segment.

Table 3.3 *Scores Highlighting Nonword Repetition Abilities Assessed Via CL-NWR Task*

Participant	Whole item Correct (Max. Score 16)	Segment Correct (Max. Score 56)
S-DLD	3	17
TD-1	5	43
TD-2	9	47
TD-3	9	42
TD-4	6	43

Participant	Whole item Correct (Max. Score 16)	Segment Correct (Max. Score 56)
TD-5	7	44
TD-6	6	38
TD-7	7	37
Mean-TD	7.00	42.00
SD-TD	1.53	3.46

3.4.3 Morphosyntactic abilities

Table 3.4 shows the scores from the SRep task. The S-DLD child scored the lowest at 1 out of 35, compared to the overall score of 21.43 (SD = 5.77; Range = 14–32) obtained by her age- and grade-matched peers.

Table 3.4 *Scores Featuring Morphosyntactic Abilities Assessed Using SRep Task*

Participant	SRep Score (Max. Score 35)
S-DLD	1
TD-1	22
TD-2	14
TD-3	21
TD-4	22
TD-5	32
TD-6	23
TD-7	16
Mean-TD	21.43
SD-TD	5.77

3.4.4 Narrative Comprehension

Table 3.5 presents the narrative comprehension scores. The S-DLD child could correctly answer only 4 out of 20 questions. On the other hand, her age and grade-matched peers scored on average 16.71 (SD = 2.93) with a range of 13–20.

Table 3.5 *Scores of Narrative Comprehension Tasks Assessed Via MAIN*

Participant	Comprehension (Max. Score 20)
S-DLD	4
TD-1	14
TD-2	20
TD-3	15
TD-4	20
TD-5	13
TD-6	16
TD-7	19
Mean-TD	16.71
SD-TD	2.93

3.4.5 Narrative Production

Table 3.6 shows the results of narrative production. Overall, scores in the retelling mode (which was also conducted first) were slightly higher than in the telling mode. For story structure, the S-DLD child scored 16 in the retelling mode and 4 in the telling mode. Overall, her age and grade- matched peers scored on average 15.71 (SD = 2.56; Range = 11–18) out of 34 in the retelling mode, compared to 16.00 out of 34 in the telling mode (SD = 3.27; Range = 11–21). Therefore, while the S-DLD child scored similarly relative to her peers in the retelling, she had a noticeable drop in performance when switched to telling without the support of a prior script, unlike her peers who could maintain a relatively stable performance across story retelling and telling modes (c.f. Sheng et al., 2020).

Scores for story complexity depended on the scoring scheme used. The S-DLD child scored 8 in retelling and 2 in telling according to Maviş et al.'s scheme, 2 in retelling and 1 in telling according to Sheng et al.'s scheme, and 16 in retelling and 4 in telling according to Gagarina et al.'s scheme. Overall, her age-and-grade matched peers scored on average 6.43 (SD = 2.94; Range = 2–12) out of 18 in retelling compared to 6.29 (SD = 2.75; Range = 3–10) out of 18 in telling according to Maviş et al.'s scheme, on average 2.00 (SD = 1.00; Range = 1–4) out of 6 in retelling compared to 1.71 (SD = 1.38; Range = 0–3) out of 6 in telling according to Sheng et al.'s scheme, or on average 14.57 (SD = 5.09; Range = 8–25) out of 36 in retelling compared to 14.29 (SD = 4.42; Range = 10–17) out of 36 in telling according to Gagarina et al.'s scheme. This shows that the S-DLD child scored significantly lower than other participants in telling mode, especially according to Maviş et al. and Gagarina et al.'s schemes. She also produced no complete GAO episodes, similar to the TD children, who overall produced very few complete GAO episodes, from 0–3 for retelling and 0–2 for telling.

Even though the S-DLD child produced a relatively high number of ISTs (tokens) compared to her peers, 24 tokens in retelling and 28 tokens in telling, the type measures were relatively low (4 in retelling, 5 in telling). The S-DLD child actually produced longer narratives (in terms of length) than her TD peers. She mainly repeated the same sentences after some intervals, leading to lengthy narratives with high IST tokens but limited IST types. It seems like she overused the same types of ISTs during her narrative productions. For the age-and-grade matched peers, the IST token measures were on average 13 (SD = 4.43; Range = 9–21) in the retelling, and on average 11.86 (SD = 4.18; Range = 7–19) in telling; while the type measures were comparable to the S-DLD child, on average 5.86 (SD = 1.46; Range = 4–9) in the retelling, and on average 5 (SD = 1.51; Range = 3–8) in telling.

Table 3.6 *Scores of Story Retelling and Telling Tasks Assessed Via MAIN*

Participant	Mode	SS	SC-M	SC-Sh	SC-Ga	GAO	IST Tokens (type)
	Retelling						
S-DLD		16	8	2	16	0	24 (4)
TD-1		18	12	4	25	3	21 (9)
TD-2		17	7	2	14	0	11 (4)
TD-3		18	6	2	14	0	17 (6)
TD-4		17	6	1	13	0	9 (5)
TD-5		11	2	1	8	0	9 (6)
TD-6		15	6	2	14	1	12 (6)
TD-7		14	6	2	14	1	12 (5)
Mean for TD participants		15.71	6.43	2.00	14.57	0.71	13.00 (5.86)
SD for TD participants		2.56	2.94	1.00	5.09	1.11	4.43 (1.46)
	Telling						
S-DLD		4	2	1	4	0	28 (5)
TD-1		18	8	3	17	0	19 (4)
TD-2		16	4	0	10	0	11 (4)
TD-3		21	10	3	20	2	15 (5)
TD-4		17	4	0	10	0	13 (8)

Participant	Mode	SS	SC-M	SC-Sh	SC-Ga	GAO	IST Tokens (type)
TD-5		11	3	1	10	0	7 (6)
TD-6		16	6	2	14	0	10 (5)
TD-7		13	9	3	19	2	8 (3)
Mean for TD participants		16.00	6.29	1.71	14.29	0.57	11.86 (5)
SD for TD participants		3.27	2.75	1.38	4.42	0.98	4.18 (1.51)

Note. EM = Elicitation Mode; SS = Story Structure; SC-M = Story Complexity by Maviş et al. (2016); SC-Sh = Story Complexity by Sheng et al. (2020); SC-GA = Story Complexity by Gagarina et al. (2019); GAO = Goal Attempt Outcome; IST = Internal State Terms.

3.5 Discussion

This study demonstrated the feasibility of identifying DLD in multilingual children using the CATALISE diagnostic criteria as the conceptual framework, with the support of the LITMUS tools to collect objective and informative evidence from multiple language areas to guide the diagnostic decisions.

The binary decision tree proposed by Bishop et al. (2017) was used to guide diagnostic decision-making to aid the identification of DLD in the S-DLD child. The S-DLD child's mother and her school reported that the S-DLD child faced difficulty in using language in a way that affected her everyday functioning and educational progress. Since the input she received in the majority language (Cantonese), is likely reduced compared to other predominantly monolingual L1 Cantonese-speaking children in Hong Kong, following the binary decision tree, the next step to consider is whether the S-DLD child is competent in another language. In this case, Urdu, reported as S-DLD's best language, was assessed. Urdu versions of the CLT (Hamdani et al., 2020b), the CL-NWR (Hamdani et al., 2020c), the SRep (Hamdani et al., 2020d), and the MAIN (Hamdani et al., 2020a) were used to collect objective and informative evidence from multiple language areas.

Results showed that the S-DLD child obtained particularly low scores in most of the tests and sub-tests, including all the CLT sub-tasks, SRep, CL-NWR, and comprehension, story structure (telling mode), and story complexity (telling mode) in the MAIN according to both scoring schemes when compared to her age and grade-matched peers, suggesting the S-DLD child is in fact not competent for her age in Urdu. As Urdu was this participant's L1 and best language, the results suggest language difficulties affecting multiple domains of language. Previous studies using LITMUS tasks have also found that compared to TD children, children with DLD show weaker narrative production and comprehension (e.g., Boerma et al., 2016; Kraljević et al., 2020; Tsimpli et al., 2016), lower nonword and sentence repetition accuracy (e.g., Boerma et al., 2015; Dos Santos & Ferré, 2018; Fleckstein et al., 2018; Gagarina et al., 2019b), and lower picture naming and selection accuracy (e.g., Kapalková & Slančová, 2017; Saliby et al., 2017). In addition, the parental questionnaire also suggested that the difficulties had been observed for some time and persisted even when the child reached school age. This together with the evidence of lack of competence even in receptive language, suggests a poor prognosis, according to Bishop et al. (2017). Since no associated biomedical conditions were reported, the S-DLD child can be identified as having DLD.

There are two further observations from the data which could be clinically valuable. First, recall in the narrative assessment, each child was first tested in story retelling, which provides a model story for reference, and then had a break doing some other language tasks and was assessed in telling the same stories without a model story. The S-DLD child had a noticeable drop in story structure and story complexity scores from retelling to telling narratives. This contrasts with her TD peers, who showed more consistent performance across both stories retell and tell tasks. This, in turn, led to a marked difference in performance on story structure and story complexity between the S-DLD child and TDs in the story-telling task but not in the model-supported retelling task. This pattern of findings was also reported by Sheng et al. (2020): while their at-risk-for-DLD group of children exhibited comparable performance on story structure and syntactic complexity as their TD peers in story retelling when supported with a prior adult model, their story structure and syntactic complexity scores decreased significantly once the adult model was removed in the story telling task, unlike their TD peers who could maintain a high level of performance across both tasks. Sheng et al. (2020) associated this phenomenon with a primary assumption of dynamic assessment, in which TD children having intact language learning abilities are expected to show evidence of a more substantial learning potential in response to some support, such as training or modeling, than children with DLD having a weaker language learning capacity. In the current context, the benefit of learning the sophisticated language and richer story structure modeled in the first retell task resulted in more sustainable performance across retell and tell tasks in the TD children. Still, this benefit was fragile and diminished quickly for the child with or at risk for DLD once the model was removed in the telling task requiring the child to generate a story from pictures. Although the current paradigm is not equivalent to a dynamic assessment, the differences in performance from the retelling task to the telling task between TD and (at risk) DLD suggest that it could be clinically informative to evaluate a child's sustainability of modifiability (improvement) upon modeling/scaffolding in assessing multilinguals. To the extent that some DLDs may benefit from modeling, we hypothesized that these benefits would be more fragile/transient from story retelling to telling relative to the typically developing peers.

The second observation concerns the different schemes for scoring story complexity in macrostructure. Different scoring approaches may vary in their ability to capture the DLD/TD group differences. In general, a scoring scheme with finer levels of granularity (i.e., more levels of score weightings) to differentiate between the different levels of structural complexity should be better able to capture the possible DLD/TD differences if the two groups differ in their story complexity. Recall the three scoring schemes used in the current study had different levels of differentiation: 1) Six levels (Gagarina et al., 2019b); 2) Four levels (Maviş et al., 2016); and 3) Two levels (Sheng et al., 2020). The findings showed that the S-DLD child scored noticeably lower than TD participants in telling mode, especially according to Maviş et al. and Gagarina et al.'s scoring schemes. While it is hard to make a firm recommendation based on this case study featuring only one child with DLD and seven TD children, we have also taken note from another ongoing research of our team examining a group of Cantonese-speaking children with DLD (N=25) and TD (N=25) using the Cantonese MAIN and compared these three scoring schemes/systems in story complexity.

The results also suggested that compared to Sheng et al.'s scoring scheme (2020), the scoring schemes of Maviş et al. (2016) and Gagarina et al. (2019) were better able to differentiate between the two groups of children. This result aligns with the idea that since the latter two scoring schemes offered relatively finer levels of granularity (i.e., more levels of score weightings) to differentiate between the different levels of structural complexity, they were better able to distinguish the two groups of children who differed in their story complexity.

Based on these findings, we could offer some preliminary advice that the scoring schemes of Maviş et al. (2016) and Gagarina et al. (2019) appeared to be better than the scoring scheme of Sheng et al. (2020) in differentiating DLD/TD. However, it is hard to conclude from these findings whether the scoring scheme of Gagarina et al. (2019) is better than that of Maviş et al. (2016). Further investigation of this observation including larger sample sizes and examining different age ranges and diagnostic/classification accuracy such as sensitivity and specificity is warranted. Moreover, practicality also needs to be considered if recommendations are made for clinicians. While a scoring scheme with the finest level of granularity is likely most differentiating, it could also be slower/harder to score. Having more informative data in the future comparing the scoring schemes of Maviş et al. (2016) and Gagarina et al. (2019) would allow us to consider both diagnostic/classification accuracy and practicality to make more concrete recommendations on which scoring system is ideal or good enough for use by clinicians in speech and language therapy clinics.

3.5.1 Strengths and Limitations

Before closing, some novelties and limitations of the current study and suggestions for further research are pointed out here. To the best of my knowledge, this study is the first to have adapted the LITMUS tools into Urdu, and there is no published research documenting the diagnostic potential/accuracy of Urdu LITMUS tools. This study is likely the first to address the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children in this major but understudied world language. This study also presents the first published data on the Urdu materials as far as the LITMUS Cross-linguistic Lexical Task (CLT), Cross-linguistic Non-Word Repetition (CL-NWR) and Sentence Repetition (SRep) data are concerned. As for the Urdu MAIN, Chan et al., 2023 published some TD data featuring another group of older elementary school children on Urdu narratives using the Urdu MAIN, examining the relationship between narrative macrostructure and microstructure. Therefore, if referring to the Urdu MAIN data from (suspected) DLD children and new data addressing the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children, this study also presents the first published data using the Urdu MAIN in this regard. Despite these novelties, the diagnostic accuracy of these new assessment tools in identifying Urdu-speaking multilingual children with and without LD or DLD, still needs to be established as we step up to future research with a larger sample size. The diagnostic accuracy of in-person versus remote testing using these LITMUS should also be explored. Regarding diagnostic accuracy measures for the Urdu-speaking multilingual children suspected of LD or DLD, further studies are needed to examine and compare the diagnostic accuracy of each of our newly adapted Urdu LITMUS tools.

This section acknowledges the following general principles from the current knowledge base in improving the diagnosis of LD or DLD in multilingual children: (i) using a combination of measures for collective evidence to improve the diagnosis of LD or DLD in multilingual children; (ii) using measures that would not disadvantage multilingual children with reduced experience to the target language (e.g., measures that are less affected by language-specific vocabulary and morphosyntactic knowledge such as nonword repetition and dynamic assessment); and (iii) using parental questionnaires (in combination with other measures) tapping information on developmental history and language background and experience. Future studies, similar to Li'el et al. (2019), could also be conducted to address this issue.

One merit of the present study is that it compares the performance of the S-DLD child with her age- and grade- matched TD peers from similar language backgrounds. Using TD peers with similar language backgrounds as a reference group is necessary to guide the developmental expectations. This is consistent with the principle that multilingual children should be benchmarked against multilingual norms or reference data but not data generated from monolingual norms or samples (Armon-Lotem, 2018) because comparing multilinguals to a monolingual sample would likely disadvantage multilingual children with reduced target language experience, leading to the risk of over-identification of DLD.

Moreover, recall that one major consideration in the CATALISE criteria in determining whether a multilingual child has LD or DLD or not is whether "there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al. 2017: Statement 4). We can therefore infer that if a multilingual child has a clear language dominance profile, then evidence showing that the child does not have age-appropriate (or developmentally appropriate (i.e., appropriate with respect to the developmental expectations for this age and this language background/exposure condition) skills even in the child's best language would constitute adequate evidence that the child would very likely not have age-appropriate skills in any language because she is likely even less proficient in her other weaker/non-dominant languages. This case study features a multilingual child with a clear language dominance profile, with her first language, Urdu, being the best/strongest language and Cantonese and other languages being the second/weaker languages. In this regard, therefore, evidence from Urdu would be informative and adequate in addressing this major consideration in the CATALISE criteria, and evidence from Cantonese was not necessary/critical in guiding the diagnostic decisions for this child in this context.

Having said that, we are aware that multilingual children are heterogeneous, and therefore for children with a different language dominance profile (for example, children with a more balanced dominance profile or children whose second language, Cantonese is their stronger/best language instead), evidence from Cantonese or languages other than Urdu would become informative in addressing whether "there is evidence that the child does not have age-appropriate skills in any language" (Bishop et al., 2017: Statement 4). Further work could demonstrate how this approach can be applied to multilingual children with different language dominance profiles. When one extends the investigation to multilingual children with diverse language dominance profiles, there are two more reminders to pay attention to. First, some skills may look "inadequate" when tested only in one language of a multilingual child, even in a child without DLD, e.g., in the case of distributed vocabulary. For these skills, it could be informative to ascertain whether there is a lack of demonstrated and expected competency even when all languages are considered for a child suspected of having DLD. Second, while the L1 of the multilingual children in this study was also the best language of these children, L1 does not always coincide with the best language for all multilingual children. Therefore, one needs to be cautious in interpreting the demonstrated competence of a multilingual child's L1 in the case of language attrition or subtractive multilingualism if dominance is undergoing a shift.

3.6 Conclusion

To conclude, the current study adds further evidence to the literature that it is possible to identify DLD in multilingual children using remote online testing. The LITMUS tools are designed to be appropriate for identifying DLD in multilingual children from different cultures and can be administered online with video conferencing software. These tools can also be used to establish reference data in TD children to guide developmental expectations for specific populations where there are not yet appropriate tools and multilingual norms for assessing multilingual children. The Urdu adaptations of the LITMUS tools and user instructions are available for free use by the international clinical and research community. They can be used to support remote online testing in Urdu in different countries and facilitate language testing for Urdu-speaking children. Notably, the results demonstrate the promise of using the CATALISE diagnostic criteria with support from the LITMUS battery tasks adapted to online testing for identifying DLD and collecting reference data in multilingual children, not only in the multilingual context presented here but also in any target language(s).

Chapter 4 - Nonword Repetition Abilities in Urdu-speaking Children with and without Developmental Language Disorder

4.1 Introduction

Nonword repetition (NWR) tasks require individuals to repeat novel words that neither have been heard before nor are part of their lexical repertoire (Chiat, 2015). This task of repeating nonwords is suggested to be less dependent on one's prior knowledge of the target language and, therefore, is claimed to have an advantage over other language assessment tasks. Some previous studies have also reported that NWR in comparison to vocabulary and grammar assessments is relatively less affected by language experience and knowledge in both monolingual (e.g., Campbell et al., 1997; Roy & Chiat, 2013) and bilingual (Thordardottir & Brandeker, 2013) populations, and is a good predictor of an individual's language-related skills (S. E. Gathercole & A. D. Baddeley, 1989; Szewczyk et al., 2018).

Moreover, NWR tasks are also reported to be robust in the face of some other individual factors like nonverbal abilities (e.g., Bishop et al., 1996; Botting & Conti-Ramsden, 2001; Stothard et al., 1998), gender (e.g., Chiat & Roy, 2007; Washington & Craig, 2004), and maternal education (e.g., Balladares et al., 2016; Chiat & Polišenská, 2016; Farabolini et al., 2021; Huls, 2017; Meir & Armon-Lotem, 2017). These findings, in general, indicate that NWR tasks might be less loaded towards one's individual characteristics. Hence, this kind of memory task could be used as a more precise marker for disentangling limited language experience from genuine language impairment (Stothard et al., 1998).

The present study aims to examine the diagnostic potential of the new Urdu NWR test in discriminating the effects of restricted language experience and Developmental Language Disorder (DLD) in the L1 Urdu-speaking multilingual children from diverse linguistic backgrounds (minority vs. majority language contexts). It also investigates the impact of some participant and item-related factors on the overall performance of children's NWR accuracy. This chapter begins by introducing some previous studies that have analyzed the clinical utility of NWR and the influence of various factors on children's NWR performance. This is then followed by methods, results, and discussion sections, and ends by concluding the present research findings.

4.1.1 NWR and Influence of Language Status and Input/Exposure

Language status refers to the socio-political significance of the minority (home/heritage language) and majority (societal language) languages spoken by multilingual populations in the host country. The younger ethnic minority (EM) multilingual children are mainly exposed to their minority language at home, as it is not spoken in the wider community. The amount of input in the minority language starts to decline as these EM children grow older and their immersion in the outer world, where the majority language is more prevalent, increases (Montrul, 2016). Hence, it is comprehensible that the sociolinguistic status of language mainly correlates with the overall amount of language input (Paradis, 2010). Therefore, acquiring the same language in the minority (as multilinguals) versus majority (as monolinguals or multilinguals) contexts could influence various linguistic abilities to different extents. For example, vocabulary and morphosyntax are significantly impacted by language status (e.g., Haman et al., 2017). At the same time, narrative macrostructure (e.g., Kunnari et al., 2016; Rodina, 2017) and NWR precision (e.g., Lee et al., 2013) are less significantly influenced by language status.

Even though NWR accuracy is mainly conceded to be less influenced by language experience (Lee et al., 2013; Thordardottir & Reid, 2022), some studies pointed out that NWR abilities are not entirely independent of language specificities. Evidence shows that children tend to repeat nonwords that share phonological features of real words in the ambient language more accurately (Chiat, 2015). As this knowledge certainly depends on language input, children with varying levels of language exposure depending on the minority vs. majority status of the target language (that also impacts children's familiarity with lexical phonology) might vary in their NWR performance. The literature presents mixed findings.

In a study, Leet et al. (2013) assessed the NWR accuracy of preschool Korean-speaking children acquiring L1 Korean as a minority (living in the USA) vs. the majority (living in Korea) language. The results revealed no significant differences in the Korean NWR performance of both groups. Several other studies, for example, Core and colleagues (2017), de Almeida et al. (2017), and Farabolini et al. (2021), also reported non-significant associations between NWR abilities and language input conditions.

In contrast, some studies have reported significant language input effects on NWR performance. For example, K. M. Sharp & V. C. M. Gathercole (2013) assessed the NWR performance of Welsh-English multilinguals in their minority language (Welsh). Children had varying levels of exposure to minority languages in homes. The findings showed a significant positive association between the amount of minority language input/exposure and NWR accuracy, especially for sounds exclusive to Welsh. Similarly, Meziane and Macleod (2021), Parra et al. (2011), and Summers et al. (2010) have also reported positive associations between language input and NWR performance in the ambient language. These inconclusive findings indeed demand further exploration.

4.1.2 NWR and DLD

NWR has been identified as a potential clinical marker for DLD (previously known as Specific Language Impairment). Several studies (e.g., Bishop et al., 1996; Ortiz, 2021; Schwob et al., 2021) have reported significantly poor nonword accuracy in children with DLD compared to their typically developing (TD) peers. DLD encompasses expressive and receptive problems in multiple areas of language, including phonology (Elliot et al., 1989), word learning (Trauner et al., 2000), and morphosyntax (Bortolini et al., 1997) in the absence of any associated biomedical condition(s) such as hearing loss, intellectual disability, etc. (Bishop et al., 2017). This language condition, which adversely impacts one's communicative abilities (Kulkarni et al., 2022), is reported to affect two children in every class of 30 with an estimated global prevalence of around 7.58 % (Norbury et al., 2016; Tomblin et al., 1997).

The diagnosis of DLD is not straightforward, as children with this condition exhibit heterogeneous linguistic profiles with a wide range of problems (Bishop et al., 2017). The identification of DLD becomes even more arduous in the multilingual context. The paucity of suitable assessment tools and appropriate multilingual norms makes the discrimination between the children with persistent language problems (clinical population) and those with transient language difficulties due to inadequate language exposure (multilingual population) even more taxing for both clinicians and researchers (Armon-Lotem & de Jong, 2015). Additionally, monolingual children with DLD and TD multilingual children reportedly have strikingly comparable language difficulties in multiple linguistic domains like vocabulary and morphosyntax (Bedore & Peña, 2008; Paradis, 2010). In such scenarios, assessment tools like NWR that do not solely rely on existing linguistic knowledge or prior language experience could be considered an efficient tool that could tease apart the effects of language impairment and multilingualism (e.g., Antonijevic-Elliott et al., 2020; Eikerling et al., 2022; Farabolini et al., 2021; Scherger, 2022).

Successful nonword repetition incorporates speech perception followed by phonological segmenting (splitting the acoustic signal into speech units that could be saved in memory), phonological assembly (formulating motor plans that combine the related speech units), and articulation. Furthermore, it also demands a sturdy representation of fundamental speech units and ample memory to store and work on the novel phonological sequence briefly. Deficit in any of these intrinsic abilities could cause imprecise NWR accuracy (for review, see Coady & Evans, 2008). This makes the NWR task a robust screening tool for DLD, as individuals with

DLD are often reported to demonstrate deficits in these areas (S. E. Gathercole & A. D. Baddeley, 1990; Montgomery, 1995; Munson et al., 2005).

The diagnostic potential of NWR has been reported in several studies targeting different languages, for example, Dutch (de Bree et al., 2007), Italian (Dispaldro et al., 2013), Russian (Kavitskaya et al., 2011), French (Thordardottir & Brandeker, 2013), and others. The only exception has been a study by Stokes and colleagues (2006) testing Cantonese-speaking children, where no significant differences were found between the NWR accuracy of children with and without DLD. Recently, Fu et al. (Resubmitted) revisited the diagnostic efficiency of NWR tasks in Cantonese-speaking children and reported contradicting findings where children with DLD performed significantly worse than their TD peers on NWR tasks. The authors suggested that the lack of significant findings in Stokes et al.'s (2006) study might be because of the low lexicality levels (i.e., less Cantonese-like nonwords) of NWR stimuli. When the NWR stimuli with high lexicality levels (i.e., more Cantonese-like nonwords) were used, the difference between TD vs. DLD groups became significant.

4.1.3 NWR and Age Effects

Similar to some other factors, studies examining the role of age on NWR accuracy also present inconsistent findings. Developmental effects on NWR abilities have not been robustly examined and demonstrated in the literature (Reid, 2019). Previous studies targeting NWR have mainly recruited school-aged children (5;0–11;0 [years;months]), and their results exhibit inconsistent findings. Some studies have suggested age-related progress in NWR accuracy. For instance, three studies employing the NWR test by Dollaghan and Campbell (1998) in TD English-speaking children reported that the accuracy scores of 7;0 (Weismer et al., 2000), 8;0 (Windsor et al., 2010), and 9;0 (Thordardottir, 2008) year olds were 82.4%, 88.3%, and 93%, respectively. These results indicate systematic age-related progress in the NWR precision of children aged between 7;0 to 9;0.

On the contrary, other studies compared NWR performance across different age bands and concluded no significant developmental effects. For example, Montgomery (2004) reported no age-related differences in the NWR accuracy of 6;5 to 10;5-year-olds (almost similar age range to the present study). Other studies that included children aged between 7;0 to 12;0 (Taylor et al.,1989), 7;10–13;11 years (Kohnert et al., 2006), 8;0 to12;0 (Reid, 2019), and 9;10 to12;5 (Dollaghan et al., 1995) also confirmed non-significant age effects on the NWR competence.

4.1.4 NWR and Lexical Knowledge

Initially, S. E. Gathercole and A. D. Baddeley (1989; 1990) proposed that phonological short-term memory (pSTM) was a sterling measure of NWR. Therefore, individuals with better pSTM capacity to hold more information in their pSTM can maintain a temporary representation of nonwords long enough to reiterate them more precisely. Individuals with poorer pSTM, on the other hand, are unable to retain the nonwords in memory and consequently have limited NWR accuracy. This idea was criticized early on, and it was suggested that in addition to pSTM, the NWR task might also involve phonological skills, articulatory abilities, and lexical knowledge (Bowey, 1996; Snowling et al., 1991).

Lexical knowledge here stands for phoneme sequences that refer to a known morpheme or lexical item (Jones & Witherstone, 2011). S. E. Gathercole (1999; 2006) further used the term redintegration (Schweickert, 1993) to explain the influence of lexical knowledge on NWR accuracy. According to this idea, nonwords with worldlike phonological or morphemic patterns might boost the support lent by long-term lexical knowledge during NWR by employing redintegration. This lexical knowledge is thus used to restore incomplete or decaying traces of the NWR stimuli.

Several recent studies (e.g., Edwards et al., 2004; Kehoe & Havy, 2019; Meziane & MacLeod, 2021; Parra et al., 2011) have also acknowledged the significant influence of lexical knowledge on overall NWR accuracy by stating a positive correlation between vocabulary size and NWR performance in both L1 and L2. Their results indicated that children who obtained higher scores in lexical tasks also gained higher scores during NWR tasks.

4.1.5 NWR and Item Length

The item length has also been identified as another important factor that influences the NWR accuracy in individuals. Different studies (e.g., Marini et al., 2017; Gibson et al., 2015; Schwob et al., 2021; Stokes et al., 2006; Summers et al., 2010) have consistently reported that NWR accuracy is inversely proportional to items' length, implying that the level of performance drops with an increase in length of the nonword items.

The length effects are suggested to be the same for children with and without genuine language impairment (Windsor et al., 2010). Weismer and colleagues (2000) assessed the NWR accuracy in school-aged children in a study. Three groups were included: children with TD, children with DLD, and children with any non-specific language disorder. The results showed that interaction between length and participant groups was non-significant, highlighting the uniform effect of length across three groups of children. The NWR accuracy in all groups declined as the item length increased. This finding signifies that item length affects the NWR precision in children irrespective of their clinical status.

4.1.6 NWR and Other Item-related Factors

Lastly, the overall NWR accuracy in children is also reported to be influenced by some other item-related factors, including wordlikness (S. E. Gathercole, 1995), neighborhood density (Vitevitch & Luce, 2005), word frequency (Bowey, 1996), phonotactic probability (Coady & Aslin, 2004; Messer et al., 2010; Zamuner, 2009), CV attestedness (Beckman & Edwards, 2000; Munson, 2001; Stokes et al., 2006), and others. These factors would not be detailed here as the focus of the present study does not include investigating the effects of different item-related factors on the NWR accuracy of Urdu-speaking children.

4.2 Present Study

This study employed the cross-linguistic NWR test (CL-NWR) developed under the COST Action IS0804 by Chiat (2015). The CL-NWR was designed using a framework that could easily be applied across typologically diverse languages with apposite adaptations and cautious attention to the possible influence of children's L1 and their exposure to any target language if it is not their L1. The overarching objective that motivated the design of the CL-NWR test was the development of a sensitive enough assessment tool for multilingual children that, on the one hand, could maximize the gap between children with and without language impairment and, on the other hand, could minimize the gap between TD children with less vs. more language experience in any given language (Chiat, 2015).

The CL-NWR was adapted into Urdu following the standard procedures (see Methods for details). The aim was to scrutinize the sensitivity of the new Urdu CL-NWR in differentiating genuine language impairment vs. the effects of limited language experience.

4.2.1 Research Questions

This study aims to answer the following research questions:

1. Does the new Urdu CL-NWR test disadvantage TD children with reduced language experience?
    - Compare EM TD vs. Maj TD
2. Does the new Urdu CL-NWR test differentiate between children with DLD and children with TD, even for TD children with reduced language experience?
    - Compare DLD vs. Maj TD
    - Compare DLD vs. EM TD

4.3 Methods

4.3.1 Participants

Three groups of children were recruited:

i) L1 Urdu TD multilingual children acquiring Urdu as a minority heritage language in Hong Kong who have reduced experience of the language of assessment (EMTD group, N=31; mean age = 8;1, range = 6;1–10;11)

ii) L1 Urdu age- gender- and grade-matched TD peers acquiring Urdu as a majority language in Pakistan (MajTD group, N=31; mean age = 8;1, range = 6;1–10;10) -

iii) L1 Urdu majority language age-matched children with DLD (MajDLD group, N=14; mean age = 8;1, range = 6;1–10;11)

The EMTD and MajTD groups included the same children who participated in Study 1. For the MajDLD group, nine out of 14 were children who had previously received SLT services for the DLD. They were retested by the Ph.D. student (also an Urdu-speaking SLT) using the CATALISE diagnostic criteria: multilingual children (suspected of) having DLD would have demonstrated evidence in terms of negative functional impact, lack of competence even in the best language (range of domains - LITMUS tools), presence of negative poor prognostic features, and absence of associated biomedical conditions (Bishop et al., 2017). Their performance was then compared to the age-matched peers from the MajTD group. The other five children with DLD were recruited from the caseload of SLTs (diagnosed with DLD and receiving SLT services). The Ph.D. student also obtained additional evidence to substantiate their diagnosis by obtaining information related to the absence of biomedical conditions, presence of functional impairment and poor prognostic feature that is age above five years (via LITMUS-PaBiQ), and lack of competence even in the L1 before recruiting these children. The lack of competence, even in the best language, was confirmed through their significantly lower lexical scores in the CLT tasks than both TD EM and Maj groups (this aligned with the clinical judgment of SLTs). The total sample included 42 females and 34 males.

The LITMUS-PaBiQ (Tuller, 2015) was used to gain information related to the child participants' L1, developmental milestones, presence/absence of any significant associated medical conditions, and/or speech, language, hearing difficulties, family history of speech and/or language impairment, parental education and working status, the quantity of exposure to different languages on daily basis, usage of different languages in daily routine, child and parents' receptive and expressive competence in different languages. Urdu was indicated as the best/strongest language in all groups. Parental consent for the participation of all children was obtained before testing.

4.3.2 Study Design

In addition to language and clinical status, this study also examined the effects of the following participant and item-related factors:

- vocabulary
- age
- nonverbal abilities
- gender
- maternal education
- item length

4.3.3 Assessment Measures

NWR Stimuli. The Urdu version of the LTMUS CL-NWR (Chiat, 2015; Hamdani et al., 2020d) was used to assess NWR abilities. The LITMUS CL-NWR uses a limited range of phonemes that are considered typical across world languages, including /p, b, t, d, k, g, s, z, l, m, n, a, i, u/. These phonemes are combined into simple CV (consonant-vowel) syllable structures. It contains 16 nonwords of varying lengths (2, 3, 4, and 5 syllables) with four items.

The Urdu version was created following standard adaptation procedures (Chiat, 2015). The Quasi-universal test (with quasi-neutral prosody) was used from the three available tests. In this test, the target stimuli are produced using neutral prosody by avoiding language-specific prosodic patterns and equally stressing all the syllables. All the syllables were produced using even pitch and length except the final one. The final syllable has a slightly longer duration and is produced with a falling pitch that denotes the end of production.

All the target stimuli were selected from an available candidate list by keeping two main factors in mind: 1) the nonwords must contain the sounds that are expected to be acquired by TD Urdu-speaking children aged three years (Ambreen, 2023; Bari & Ajmal, 2016; Zahra, 2016); and 2) any nonword should not resemble any real word in Urdu.

Lexical Scores. The Urdu version of the LITMUS Cross-linguistic lexical tasks (CLT; Haman et al., 2015; Hamdani et al., 2020a) was used to obtain expressive lexical scores. CLT is a picture-based test that assesses receptive and expressive lexical knowledge. The details about this test have already been included in Chapter 2 (Methods section) of this dissertation. Only the scores from the production sub-tasks were employed in this study.

Nonverbal Ability. Nonverbal ability was assessed using Raven's Colored Progressive Matrices (Raven & Court 1998). The test was adapted into an online format using PowerPoint, and the instructions were given in the children's strongest language (Urdu). The children with the standard score of 70 or above were considered in the normal range.

4.3.4 Procedure

Testing sessions were conducted online using "Zoom" software due to the pandemic. All the children used the same headset with a microphone to ensure input and output quality. All the children participated using laptops or tablets. Cell phones were not allowed. The participants' video was turned on during the testing sessions so the examiner could monitor the situation. The examiner also used the same input device during all sessions to standardize the quality of the stimuli provided to each child. The parents/caregivers were advised to turn on the highest volume of the devices they were using to ensure that the speech sounds were loud and clear on their sides. Some practice trials were also conducted before testing, and the children were asked to repeat some easy syllable strings to obtain some objective evidence that they could hear the audio stimuli clearly before proceeding to the testing phase.

The parents/caregivers were advised to arrange a quiet environment before testing. It was ensured that the child was accompanied by only one parent/caregiver during the testing session(s). The parents/caregivers were further requested not to interfere or help during the assessment. The sessions were audio-recorded for future transcriptions and data analyses.

The Urdu CL-NWR tasks were embedded into a PowerPoint (Polišenská & Kapalková, 2014) presenting a broken necklace, and the children in the PowerPoint (names were used according to the culture) wanted to repair it for their mother. So, the participants were asked to repeat the magic words in the PowerPoint to help the children. The participants were told that a bead would appear upon every repetition. Two practice items were also included. Before presenting the target stimuli, it was confirmed via repetition of practice items that children understood the task correctly. Presenting the task in a game-like format helped retain the child participants' interest (see Figure J in Supplementary Materials). Moreover, embedding both instructions and stimuli also helped to maximize the input quality and standardize the testing procedure.

The expressive lexical scores for EMTD and MajTD were taken from the previous study (Study 1, Chapter 2 in this thesis). The same procedure (see Methods in Chapter 2) was employed to obtain lexical scores of MajDLD children. They were asked to name the CLT pictures (nouns and verb sub-tasks) in Urdu.

4.3.5 Scoring

The scoring instructions provided by Chiat (2015) were followed. Two different types of scoring schemes were employed. The score was assigned if the whole item was correct for the stringent whole item correct scoring scheme. An item was considered correct if all the segments in the target were produced in the exact order. The maximum score was 16. For the second by-segment scoring scheme, the score was awarded if the child correctly repeated any syllable(s) in the target item. The maximum score was 56. Substitutions, omissions, and additions were considered as errors. Allowances were provided for consistent phonological patterns (e.g., fronting of velar stops) representing the child's immature speech.

The expressive lexical data from the MajDLD group was scored using the same convention provided for the CLT by Haman and colleagues (2015) employed in Study 1. The items were marked using a binary 0-1 coding scheme. The information regarding categorizing correct and incorrect responses has been detailed in Chapter 2 (Methods section).

4.3.6 Inter-rater Reliability

Two raters independently transcribed 11% of the CL-NWR data. Both raters were Urdu-speaking SLTs and had relevant training in phonology. Out of 128 total NWR items (16 target items per participant, 8 participants in total), there was a disagreement on nine items where both raters perceived and transcribed the phoneme(s) differently, and the disagreement was not resolved by discussion. In this case, a third rater who was also an Urdu-speaking SLT and was working in the field of Urdu phonology was invited to transcribe the discrepant items independently. The transcription for the discrepant items was finalized based on the dominant number of votes (by choosing the transcription with 2 out of 3 votes). Overall, the percentage of agreement between the two raters was 93% (119 items over 128 items).

4.3.7 Statistical Analyses

The statistical analyses were completed by the Ph.D. student using R (version 4.2.1, R Core Development Team, 2022). Mixed-effects modeling was used as it allows a researcher to examine the variables of interest (fixed effects) and their possible interactions while also considering variability within and across participants and items simultaneously (random effects and random slopes). The lme4 package (Bates et al., 2015) was used to run the Generalized Linear Mixed Effect Models (GLMM). The dominance analysis package (Bustos, 2020) was used to obtain the average contribution/dominance of each significant predictor in the model. The post-hoc results were computed using the emmeans package (Lenth, 2022). The findings were considered significant with p values less than 0.05.

There were also some missing values. The IQ scores of five out of 14 children with DLD (the ones who were recruited from the caseload of SLTs) were missing. Instead of using the maximal model approach, the stepwise forward selection approach was used while selecting the predictors in the final model. R omits all the rows (data from the subjects) with missing values, as there were only 14 subjects with DLD. Therefore, the stepwise forward selection approach was used based on the reasoning that comparing the models using the Likelihood ratio test would allow to include data from all the subjects if IQ does not turn out to be the significant predictor (and that was the case in this study). Consequently, IQ was not included in the final model after testing the significance of fixed effects using the likelihood comparisons.

4.4 Results

This section presents the findings related to the NWR abilities of Urdu-speaking children.

The effects of different factors on both scoring conventions (whole item correct and segment correct) are discussed.

4.4.1 NWR Accuracy at the Whole Item Correct Level

The first phase of analyses explored the effects of various participant-related (language/clinical status, age, gender, nonverbal IQ, lexical scores, and maternal education) and item-related (item length) factors on NWR performance at the whole item level. The NWR scores were included as an outcome variable. Language group (MajDLD vs. MajTD vs. EMTD), lexical scores, age (mean-centered), gender, maternal education (mean-centered), nonverbal IQ, item length (2 vs. 3 vs. 4 vs. 5 syllables), and their interactions were added as fixed effects, and participants and items as random effects. Variables were added to the final model using a stepwise forward selection approach, and the models were compared using the likelihood ratio tests. None of the interactions were significant. The final model was as follows: NWRScores ~ LanguageGroups + LexicalScores + Length + (1|participant) + (1|Item).

Table 4.1 shows the fixed effects of the model. As the table indicates, the impact of language/clinical status diminished when other factors were considered. This indicates that the whole item correct scoring is likely a more stringent level of scoring, and even TD children may find it challenging to repeat the whole items correctly. Therefore, the group differences were not significant at this level.

On the other hand, the effect of lexical scores was significant, denoting that the children with higher lexical scores had higher NWR accuracy at the whole item level. There was also a significant negative effect of item length, showing that the NWR accuracy declined as the item length increased. Furthermore, the average contribution of each significant predictor was also calculated. The analysis showed that item length was stronger than lexical scores in predicting the NWR accuracy at this stringent whole item level (Figure 4.1).

Table 4.1 *GLMM Analysis Summary for Fixed Effects Predicting NWR Accuracy Scores at the Whole Item Correct Level (Max. Score 16)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	1.18	.878	1.34	.178
Language/Clinical Status (MajDLD vs. MajTD)	.621	.391	1.58	.112
Language/Clinical Status (MajDLD vs. EMTD)	.442	.297	1.48	.137
Lexical Score	.037	.017	2.10	.035*
Item Length	-.787	.195	-4.03	<.0001***

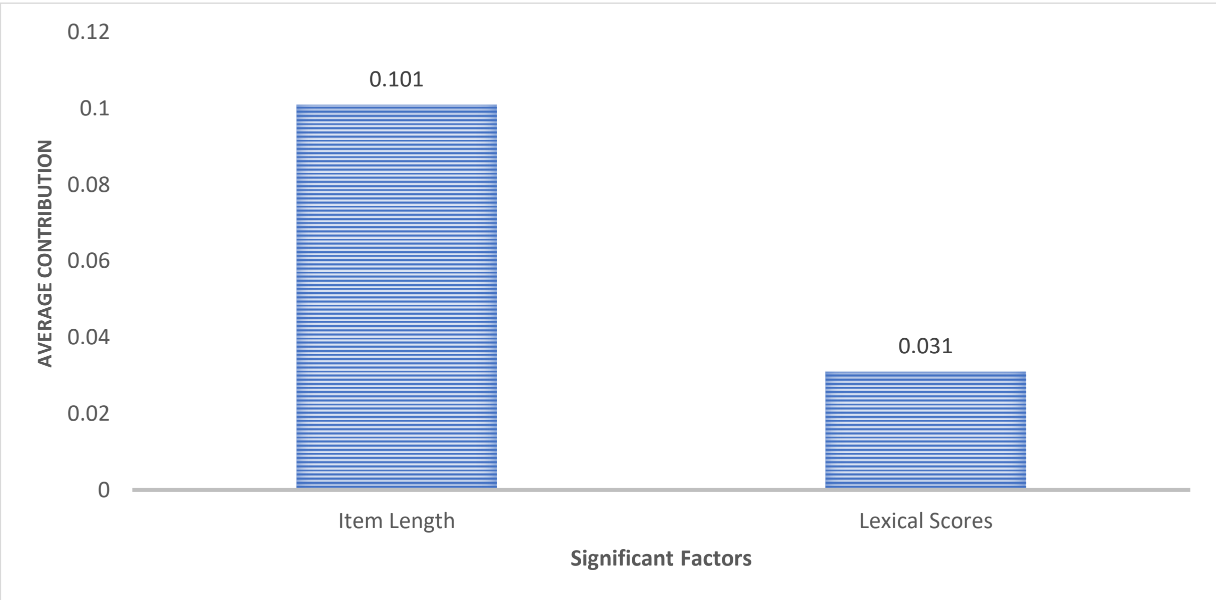

Figure 4.1 Average Contribution of Significant Factors in Predicting NWR Scores at the Whole Item Correct Level

Although the effect of language/clinical status was non-significant, the means of NWR scores at this stringent level were computed separately (Figure 4.2) to find out how children from different language groups performed. The results showed that the MajTD group obtained the highest scores, followed by EMTD and MajDLD (MajTD > EMTD > MajDLD).

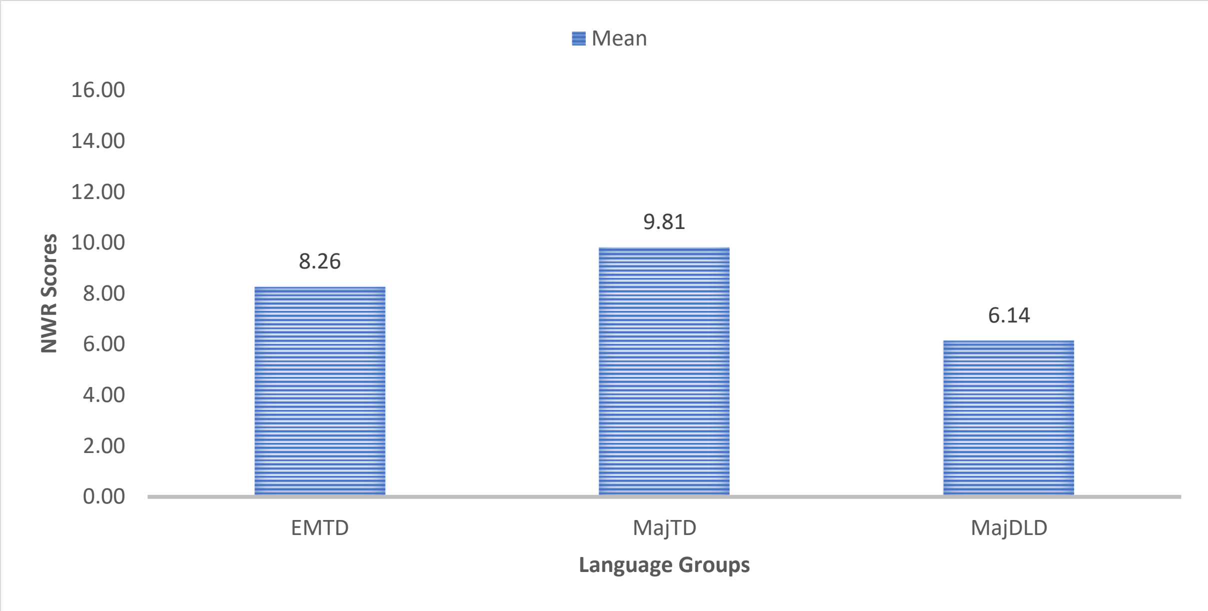

Figure 4.2 Mean NWR Accuracy Scores at the Whole Item Correct Level in Each Language Group *(Max. Score 16)*

4.4.2 NWR Accuracy at the Segment Correct Level

The second round of analyses again explored the effects of different participant-related (language/clinical status, age, gender, nonverbal IQ, lexical scores, and maternal education) and item-related (item length) factors on NWR performance at segment correct level. The NWR scores were included as an outcome variable. Language group (MajDLD vs. MajTD vs. EMTD), lexical scores, age (mean-centered), gender, maternal education (mean-centered), nonverbal IQ, item length (2 vs. 3 vs. 4 vs. 5 syllables), and their interactions were added as fixed effects, and participants and items as random effects. Variables were added to the final model using a stepwise forward selection approach, and the models were compared using the likelihood ratio tests. None of the interactions were significant. The final model was as follows: NWRScores ~ LanguageGroup + Gender + Length + (1|participant) + (1|Item).

Table 4.2 presents the fixed effects of the model. Language/clinical status was significantly affected at this scoring level (Figure 4.3). The MajTD group scored significantly worse than both TD groups (MajTD & EM). The post-hoc results (Table 4.4) calculated using pairwise comparisons further revealed that the difference between the TD and DLD groups was significant. On the other hand, the difference in the NWR accuracy between the TD children from majority and minority groups was non-significant. So, as far as the current sample of children and the current results are concerned, it can be assumed that for the Urdu CL-NWR test, the segment correct scoring scheme better differentiated between the TD and DLD children, and this new assessment test does not disadvantage children with limited language experience. Furthermore, item length had a significant negative effect, showing that the performance on the shorter items was better than on the longer ones. The effect of gender, on the other hand, diminished when other factors were considered. The average contribution of each significant predictor was also calculated. The analysis showed that language/clinical status and item length contributed equally to predicting NWR accuracy at this relatively lenient level (Figure 4.5).

Table 4.2 *GLMM Analysis Summary for Fixed Effects Predicting NWR Accuracy Scores at the Segment Correct Level (Max. Score 56)*

Fixed Effects	*B*	*SE*	*z*	*p*
(Intercept)	2.37	.603	3.93	<.0001***
Language/Clinical Status (MajDLD vs. MajTD)	.833	.197	4.25	<.0001***
Language/Clinical Status (MajDLD vs. EM)	.524	.193	2.71	.006**
Gender	-.145	.143	-1.01	.310
Item Length	-.339	.155	-2.19	.028*

Table 4.3 *Post-hoc Pairwise Comparison of NWR Accuracy at the Segment Correct Level in Children from Different Language Groups*

Contrasts	*β*	*SE*	*z*	*p*
MajDLD – MajTD	-.834	.196	-4.25	.0001***
MajDLD – EM	-.524	.193	-2.71	.018*
MajTD – EM	.309	.156	1.97	.177

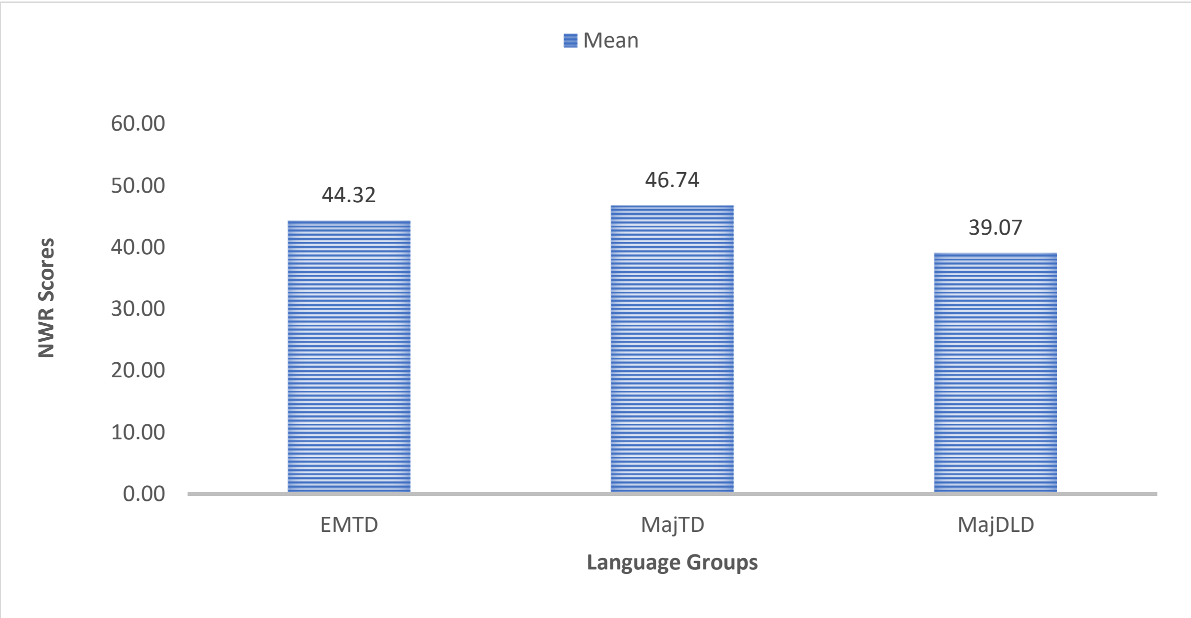

Figure 4.3 Mean NWR Accuracy Scores at the Segment Correct Level in Each Language Group (Max. Score 56)

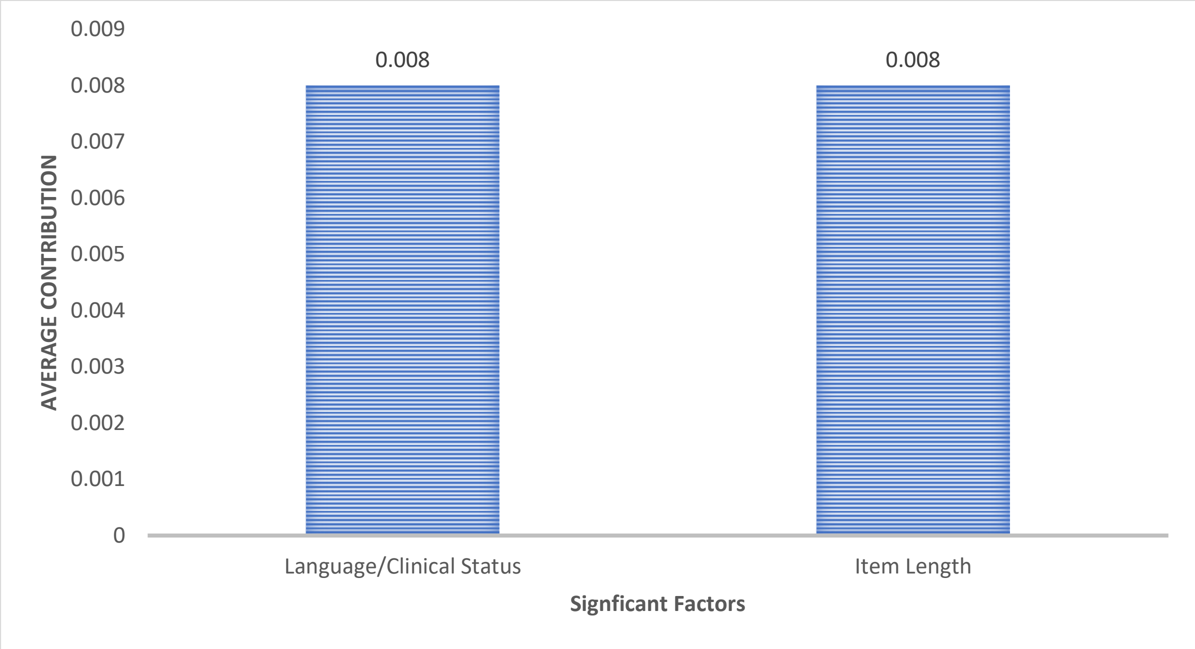

Figure 4.4 Average Contribution of Significant Factors in Predicting NWR Scores at the Segment Correct Level

4.5 Discussion

This study examined the diagnostic potential of the new Urdu CL-NWR in teasing apart the effects of limited language experience and DLD. In addition to analyzing the effects of clinical (TD vs. DLD) and language (minority vs. majority) status, the impact of item (length) and participant (age, gender, lexical scores, nonverbal IQ, and ME) factors on the overall NWR accuracy was also analyzed. NWR accuracy was analyzed at two different levels: whole item correct and segment correct. Overall, the findings revealed that item length affects NWR performance at both levels, but language/clinical status, lexical knowledge, and other factors have varying effects on these two levels.

The findings revealed a significant effect of language/clinical status only at the segment correct level and not at the whole item correct level. This indicates that the repetition of nonwords at the stringent (whole item) level was probably equally challenging for children from different language groups (MajDLD, MajTD, & EMTD). The significant effect of language/clinical status at the segment correct level was further probed using a post-hoc analysis. This helped to answer the first and second research questions.

The first question was: Does the new Urdu CL-NWR test disadvantage TD children with reduced language experience? The findings from the post-hoc analysis confirmed that the group difference between TD minority and majority groups was non-significant. This shows that language status, which is also correlated with language input conditions, does not affect the NWR accuracy in TD children. This finding supports the claim that NWR tasks do not disadvantage TD children acquiring language under restricted input conditions (Core et al., 2017; de Almeida et al., 2017; Lee et al., 2013; Farabolini et al., 2021). Overall, this set of findings shows that the effect of language impairment outweighs the impact of reduced language experience. This finding is also supported by the effect size of group differences: MajDLD vs. MajTD (d = -.834), EM vs. MajDLD (d = -.524), and MajTD vs. EM (d = .309).

The second question was: Does the new Urdu CL-NWR test differentiate between children with DLD and children with TD, even for TD children with reduced language experience? The results showed that the children with DLD performed significantly worse than TD peers from both majority and minority groups at the segment level. This indicates that children with DLD, in general, could repeat a significantly smaller number of nonword segments accurately, highlighting the diagnostic potential of this new NWR test. Significantly low NWR accuracy in children with DLD in comparison to their TD peers has also been reported in several previous studies (e.g., Chiat, 2015; Coady & Evans, 2008; Ortiz, 2021; Schwob et al., 2021).

The successful repetition of aurally presented novel words incorporates numerous skills, including speech perception, phonological encoding, phonological memory, phonological assembly, motor planning, and production. Different studies examining these skills in children with DLD have reported that these children demonstrate impairments in all of these auxiliary skills, leading to poor NWR accuracy (Coady & Evans, 2008).

It is worth mentioning that if this new tool is intended for clinical screening, then it is essential to identify which scoring method is more efficient. In contrast to previous claims (e.g., Dispaldro et al., 2013) where scoring at the whole item correct level has been identified as a better method for clinical screening, present results showed that scoring at the segment level (in the case of the new Urdu CL-NWR) might be more sensitive to the DLD identification. Boerma et al. (2015) also reported that scoring at both item and segment levels distinguished well between children with and without language disorders in both monolingual and multilingual groups. In the monolingual group, scoring at the item correct level gained a higher level of specificity and sensitivity. Meanwhile, in the multilingual group, findings from both scoring conventions were highly similar. Indeed, evidence from a larger sample size is required to verify these findings.

The following sections present some additional findings related to the effects of some participant and item-related factors.

Vocabulary/lexical knowledge is often claimed to be associated with NWR scores. Children with higher lexical scores are also found to score well during NWR tasks (e.g., Farabolini et al., 2021; S. E. Gathercole, 2006; Meziane & MacLeod, 2021; Szewczyk et al., 2018). The results from this study also coincide with these previous claims as there was a significant positive effect of lexical abilities and NWR accuracy at the whole item level, implying that children with higher lexical scores also demonstrated better NWR performance. This positive correlation between lexical knowledge and NWR accuracy has been explained in terms of the lexical redintegration process (S. E. Gathercole, 2006; Jones & Witherstone, 2011). This process is claimed to play a vital role in repeating the nonwords accurately, where individuals use their lexical knowledge to repair the degraded trace of information during the NWR task, enhancing their NWR precision.

Results from the additional analyses showed that children with DLD, in addition to poor NWR scores, also obtained significantly poorer lexical scores than both MajTD (*p* = <.0001) and EMTD (*p* = <.0001) peers. These significant differences in the lexical and NWR scores of children with typical development and DLD also provide support to the notion that vocabulary size/lexical knowledge is positively associated with NWR performance (e.g., Bowey, 1996; Coady & Evans, 2008; Farabolini et al., 2021; Metsala, 1999). Children with DLD do not possess strong enough linguistic or lexical skills, so they might not be able to use redintegration strategies to fetch support during NWR tasks (Graf-Estes et al., 2007; Stokes et al., 2006), thus resulting in significantly poor NWR accuracy.

Furthermore, age did not turn out to be a significant predictor of NWR abilities. The children in this study were aged between 6;0 to 11;0, and the findings revealed no significant age-related progress. Some previous studies focusing on NWR competence of school-aged children (6;5– 13;11 years) also reported non-significant age effects (e.g., Kohnert et al., 2006; Montgomery, 2004; Reid, 2019). This finding implies that children as young as six might have developed stable enough NWR skills that significant age-related progress might not be evident anymore. A different angle to this finding could be that the current NWR tasks might not be sensitive enough to capture developmental effects in older children. Hence, some more complicated, more challenging stimuli could be included to scrutinize age effects in older children.

The effect of nonverbal IQ was also examined. In line with the previous findings, the effect of nonverbal IQ was non-significant (Bishop et al., 1996; Botting & Conti-Ramsden, 2001; Stothard et al., 1998). This finding suggests that the new Urdu CL-NWR might be a suitable screening tool for identifying language impairment regardless of an individual's nonverbal abilities. Bishop et al. (1996) have also stated that a valuable parameter for a linguistic marker must be highly independent of nonverbal abilities because one is interested in identifying linguistic deficits in children and not in the overall learning capacity.

Additionally, the influence of some other individual characteristics was also examined. In line with some previous findings, the effects of other participant-related factors, including gender (e.g., Chiat & Roy, 2007; Roy & Chiat, 2004) and maternal education (e.g., Balladares et al., 2016; Farabolini et al., 2021) were non-significant. These findings imply that NWR tasks might be less weighed down by different individual characteristics, making it a more precise clinical tool for detecting genuine language impairment.

Next, the effect of item length was also examined. The length of the test items varied between 2 to 5 syllables. The findings from the present study concur with the previous literature reporting a decline in NWR accuracy with an increase in item length (e.g., L. M. D. Archibald & S. E. Gathercole, 2006; S. E. Gathercole & A. D. Baddeley, 1989). This shows that longer nonwords might place more burden on the segmentation and articulation abilities of the individuals than the shorter ones.

Finally, the interaction between language group and length was also non-significant. This suggests that despite language or clinical status, item length posits the same difficulty level across all language groups. Similar results were also reported by Weismer et al. (2000), who examined NWR accuracy in second graders with typical development, DLD, and any non-specific language impairment. They also reported a uniform significant main effect of item length across groups. Similar to this study, item length was inversely proportional to NWR accuracy in all groups.

4.5.1 Study Limitations

The present study also has some limitations. It is important to note that the current study assessed the school-aged (6–11 years) children only, so future work can recruit the younger children to verify whether these present findings still hold and that the new Urdu CL-NWR tool has the potential to differentiate between persistent language impairment and transient language problems arising from restricted language input in younger children as well. Validating the present findings in the younger population is highly vital as the development of clinical screening tools should aim for early identification, which could help open avenues for early intervention. Last but not least, all the data was collected online due to the COVID-19 pandemic. Future studies could further this line of research and verify these findings by collecting data in the in-person, face-to-face setting.

4.6 Conclusion

This study is the first to document the NWR abilities in Urdu-speaking children. Consistent with the crosslinguistic literature, the findings showed that NWR tasks have the potential to differentiate between Urdu-speaking children with and without DLD. Scoring at the segment level turned out to be more sensitive to DLD screening. The Urdu CL-NWR tool could detect DLD and appear unaffected by the limited language experience of EMTD multilingual children, as the TD children from both minority and majority backgrounds exhibited comparable NWR accuracy. The current findings also support the notion that lexical and NWR abilities are positively correlated and that individuals with higher lexical knowledge might also have higher NWR accuracy. The effects of other individual factors, including age, gender, nonverbal ability, and maternal education, were all non-significant. Furthermore, item length, as reported in several previous studies, was inversely related to the overall NWR accuracy in children. The performance declined as the item length of novel words increased, and this decline in performance was consistent in children with and without DLD.

Chapter 5 - Summary of the Major Findings

5.1 Introduction

This thesis studied the acquisition of heritage minority language (L1) in the ethnic minority (EM) Urdu-Cantonese multilingual children residing in Hong Kong. Minority heritage language refers to the language spoken at home by the EM groups and is not the dominant societal language (Rothman, 2009). The context of heritage language development based on its minority sociopolitical status is often characterized by reduced input (Montrul, 2016). This reduced language input could lead to vulnerabilities in certain linguistic domains, including lexicon (Bohnacker et al., 2016), morphosyntax (Haman et al., 2017), and expressive narrative microstructure (Rodina, 2017). This incessantly waning heritage language input and usage often leads to incomplete heritage language acquisition (Anderson, 2001) or attrition of the Aboriginal linguistic system, which involves the loss of linguistic skills in a multilingual scenario (Montrul, 2016; Polinsky, 2006).

The L1 attrition in a minority context may also lead to restricted progress across age (compared to L1 acquisition in the majority context, which would show more prominent progress across age). Furthermore, linguistic vulnerabilities associated with reduced input could resemble linguistic vulnerabilities related to Developmental Language Disorder (DLD; Blom et al., 2019). There can be apparent overlaps between the language profiles of typically developing (TD) multilingual children and monolingual children with DLD (Armon-Lotem & de Jong 2015). Given these overlaps and less awareness about DLD in the general public, there could be higher chances of over- or under-identification of DLD in multilingual children (Armon-Lotem & de Jong 2015). Thus, differentiating the effects of reduced input and the impact of DLD might be more challenging in a heritage minority language context among EM multilingual children.

The professionals working with these EM children often face practical constraints while conducting speech and language assessments and making diagnoses. This is because suitable assessment tools to evaluate the speech and language abilities of these multilingual children's heritage language (likely also the best for younger children) are usually lacking or absent. On the other hand, if available, the existing assessment tools/measures are usually normed for predominantly monolingual children. These monolingual norms should not be generalized to multilingual learners (Armon-Lotem, 2018; Bedore & Peña, 2008). Currently, we do not even have reference data to inform researchers and clinicians about the developmental expectations for the best language of these multilingual EM children. This Ph.D. project aimed to improve the assessment of multilingual children and the identification of DLD.

Therefore, this project wanted to capitalize on different assessment tools from the European COST Action LITMUS (Language Impairment Testing in Multilingual Settings) battery to facilitate the identification of multilingual Urdu-speaking children with and without DLD. Four of the widely used LITMUS tools were adapted into Urdu for this project: Crosslinguistic lexical tasks (CLT, Haman et al., 2015; Hamdani et al., 2020a), Sentence Repetition task (SRep, Marinis & Armon-Lotem, 2015; Hamdani et al., 2020b), Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2019a; Hamdani et al., 2020c), and Crosslinguistic Nonword Repetition Test (CL-NWR, Chiat, 2015; Hamdani et al., 2020d). In addition to the offline versions, the digital versions were also designed to enable remote online testing, as face-to-face testing was not feasible due to the COVID-19 upsurge.

This thesis includes three studies. Study 1 aimed to identify the areas of linguistic strength and vulnerabilities in the EM children acquiring language under reduced input conditions by assessing them in a range of linguistic domains (lexical, morphosyntactic, and narrative abilities). In doing so, their performance was also compared to the TD age- gender- and grade-matched peers from Pakistan acquiring L1 Urdu in relatively abundant language conditions. In addition to the effect of language status (minority vs. majority), the impact of age, amount of Urdu input, and some domain-specific factors were also explored. Study 2 demonstrated the feasibility of using the CATALISE diagnostic criteria in combination with LITMUS tools to identify multilingual children with DLD and collect reference data via remote online testing. Study 3 concentrated on NWR abilities in three groups of children, including children with and without DLD. This study investigated the diagnostic efficiency of the new Urdu CL-NWR test in disentangling limited language experience from genuine language impairment.

This chapter first summarizes the findings from these studies. Then, it highlights the empirical and clinical significance of this project and discusses some implications for future work. The final section concludes this thesis.

5.2 Summary of Findings

Study 1

This study documents the linguistic vulnerabilities associated with reduced input in multilingual children acquiring Urdu as their L1 and minority language. The EM children's linguistic competence was compared to their age- gender- and grade-matched control peers from Pakistan. The EM children living abroad acquire their L1 Urdu under reduced input conditions in their host countries. On the other hand, children from the majority group residing in Pakistan acquire their L1 Urdu under relatively abundant input conditions. Since heritage speakers are mostly raised under circumstances of subtractive multilingualism, the core features of their heritage language and their linguistic competence might be different from the linguistic abilities of their age-matched counterparts from their home countries or even parents (Montrul, 2010). Therefore, it was predicted that reduced input might affect certain linguistic abilities more than others. Moreover, reduced input associated with minority language acquisition context might lead to more restricted age-related progress in the first/heritage language (first language attrition).

Thirty-one EM children from Hong Kong and 31 age- gender- and grade-matched peers from Pakistan participated (mean age = 8;2 [years; months], range = 6;0–10;11) in this study. All the participants had normal nonverbal abilities confirmed via Raven's Progressive Matrices (Raven & Court, 1998). The information about participants, their language environment, and their parents was obtained via a parental questionnaire (LITMUS-PABIQ; Tuller, 2015). Lexical (comprehension and production abilities), morphosyntactic (production), and narrative (comprehension and production) abilities were assessed using the Urdu versions of CLT (Haman et al., 2015; Hamdani et al., 2020a), SRep (Marinis & Armon-Lotem, 2015; Hamdani et al., 2020b), and MAIN (Gagarina et al., 2019a; Hamdani et al., 2020c), respectively. The statistical analyses were conducted using Mixed-effects models via R.

Results showed that, on the one hand, the EM children obtained comparable scores to their control peers in narrative comprehension, expressive narrative macrostructure (story structure and story complexity), expressive narrative microstructure (length and productivity, syntactic complexity, lexical diversity), and use of internal state terms (ISTs, the internal states of story characters like their perceptions and thoughts). On the other hand, they scored significantly lower than the controls in the measures of narrative microstructure (proportion of grammatical C-units), morphosyntactic, and lexical competence. Moreover, the age effects were found to be uniform across both groups. The EM children exhibited age-related progress similar to their control peers in most linguistic areas except for two expressive narrative microstructure measures: mean length of C-units and verb accuracy (number of correct verbs). In general, these findings reveal that despite reduced input, the EM children for their L1 acquisition, like their control peers, might follow a similar developmental route. However, this pattern might be delayed to some extent.

In addition to language status (minority vs. majority), the correlations between the amount of Urdu input and different language measures, including lexicon, morphosyntax, and expressive narrative measures (syntactic complexity and grammaticality) were found to be highly significant. These language domains appeared to be more sensitive to the language input. In contrast, narrative comprehension, expressive narrative macrostructure, some expressive narrative macrostructure such as NDW and TNW, and ISTs appeared to be less affected by the amount of Urdu input, as the associations between the Urdu input and these measures were non-significant.

The influence of some domain-specific predictors was also assessed. For the lexical abilities, performance in production was significantly worse than comprehension, and nouns were easier than verbs. There was also a significant effect of sentence structure in the SRep task, and children found sentences with ORC, subordination, and wh-object questions to be the most difficult. The effect of elicitation modes (retelling vs. telling) on the narrative productions was also examined. The significant effect of the elicitation mode was evident for the proportion of subordinate C-units (syntactic complexity measure) only, and performance was better during the retelling. These TD children were generally found to produce stories of comparable quality in both conditions with and without support from the model story. Lastly, the error analyses showed that the EM children committed significantly more morphosyntactic (during narrative and SRep tasks) and lexical errors than their control peers.

Moreover, maternal education was found to have varying impacts across different language areas. The children with higher levels of maternal education produced lengthier narratives (productivity) and employed various types of ISTs. On the contrary, children with lower levels of maternal education gained significantly higher lexical scores and generated more grammatical utterances during their story productions. The effect was non-significant for the other outcome measures.

Overall, the findings from this study support the prediction that reduced input might affect specific linguistic abilities more than other linguistic abilities. Expressive narrative microstructure (grammaticality), morphosyntactic, and lexical competence turned out to be more sensitive to input effects. In contrast, narrative comprehension, expressive narrative macrostructure, microstructure (lexical diversity, syntactic complexity, and productivity), and use of ISTs were relatively less affected by specific language experience. In other words, despite restricted input conditions, narrative comprehension, expressive narrative macrostructure, syntactic complexity, lexical diversity, narrative length (microstructure), and ISTs emerged as the areas of linguistic strength in EM children. On the other hand, vocabulary and morphosyntax (in both narrative production and SRep) appeared as the areas of linguistic vulnerabilities in these children. There is a likelihood that as these EM children grow older and their exposure to the L2 (majority language) increases (Kohnert, 2004), this gap in performance between the two language groups might increase to the level that it transforms into incomplete acquisition or even language attrition in L1 (Montrul, 2008). In such scenarios, it could be inferred that for some language abilities, these EM children would either exhibit developmental delay or might not be able to attain the age-appropriate competence levels similar to their monolingual or even multilingual majority peers (Montrul, 2022).

Study 2

A long-standing issue in identifying DLD in multilingual children is differentiating between the effects of language experience and genuine impairment when clinicians often lack suitable norm-referenced assessments. This case study demonstrates that it is feasible to identify DLD in a multilingual child using the CATALISE diagnostic guidelines, LITMUS assessment tools, and remote online testing.

This case study features one six-year-old Urdu-Cantonese multilingual EM child suspected of DLD (S-DLD) and seven age and grade-matched multilingual TD peers. They were tested via Zoom using the Urdu versions of CLT (Haman et al., 2015; Hamdani et al., 2020a), SRep (Marinis & Armon-Lotem, 2015; Hamdani et al., 2020b), MAIN (Gagarina et al., 2019a; Hamdani et al., 2020c), and CL-NWR (Chiat, 2015; Hamdani et al. 2020d), . Nonverbal IQ was measured using Raven's Progressive Matrices (Raven & Court, 1998). All the children, including the S-DLD child, had normal non-verbal abilities. The information about children and their language environments was obtained using the LITMUS-PABIQ (Tuller, 2015). All the participants in this case study had comparable SES, were studying in the grade P1 in local Cantonese-medium schools, and had L1 Urdu as their best language.

Results showed that the S-DLD child obtained lower scores in most of the tests and sub-tests, including comprehension, story structure (telling mode), and story complexity (telling mode) in the MAIN, all the CLT-subtasks, the SRep, and the CL-NWR according to both scoring schemes, compared to her age- and grade-matched peers. Moreover, as Table 3.1 (Chapter 3) indicated, the amount of Urdu input to the S-DLD child was not less than that to her other TD peers. Still, the S-DLD child scored noticeably lower than the other TD peers across multiple language areas in the assessments. Some TD peers scored better even with less reported amount of Urdu exposure. These data collectively suggest that the S-DLD child is, in fact, not competent for her age in Urdu. As Urdu is this participant's L1 and best language, the results suggest language difficulties affecting multiple language domains. Therefore, together with the presence of negative functional impact, poor prognostic features, and absence of associated biomedical conditions, the results suggest this participant could be identified as having DLD using the CATALISE diagnostic criteria.

The results add further evidence to the literature that it is possible to identify DLD in multilingual children using remote online testing. The LITMUS tools are designed to be appropriate for identifying DLD in multilingual children from different cultures and can be administered online with video conferencing software. These tools can also be used to establish reference data in TD children to guide developmental expectations for specific populations where there are not yet appropriate tools and multilingual norms for assessing multilingual children.

Study 3

Nonword repetition (NWR) tasks require repeating novel words that have neither been heard before nor are part of one's lexical repertoire (Chiat, 2015). Several studies have acknowledged NWR as a tool that, on the one hand, is robust in identifying children with DLD (e.g., Bishop et al., 1996; Schwob et al., 2021) and, on the other hand, is immune to language input conditions/language experience (Roy & Chiat, 2013; Thordardottir & Brandeker, 2013). This study also aimed to examine the clinical utility of the new Urdu CL-NWR tool in differentiating between children with persistent language problems (clinical population) and those with transient language difficulties based on reduced input (multilingual population).

Three groups of children were tested. (i) TD children acquiring L1 Urdu as a majority language in Pakistan (MajTD group; N=31; mean age = 8;1, range = 6;1–10;10); (ii) age- gender- and grade-matched TD peers acquiring L1 Urdu as a minority heritage language in Hong Kong who have reduced experience of the language of assessment (EMTD group; N=31; mean age = 8;1, range = 6;1–10;11); and (iii) a group of age-matched peers with DLD acquiring L1 Urdu as majority language in Pakistan who do not have the issue of reduced experience of the language of assessment but suffer from language disorder (MajDLD group: N=14; mean age = 8;1, range = 6;3–10;11). NWR abilities were assessed using the Urdu version of CL-NWR (Chiat, 2015; Hamdani et al., 2020d).

The impact of some item (length) and participant (age, gender, lexical scores, nonverbal IQ, and maternal education) related factors on the overall NWR performance was also analyzed. Nonverbal IQ was measured using Raven's Progressive Matrices (Raven & Court, 1998), and the information related to children and their language environments was obtained using the LITMUS-PABIQ (Tuller, 2015). Lexical knowledge was assessed using the CLT.

NWR abilities were assessed at the whole item correct and segment correct levels. The statistical analyses were completed using mixed-effect models in R. Findings showed that this new Urdu CL-NWR tool was not only able to distinguish children with DLD but also seemed to be unaffected by the limited language experience of EM multilingual children, as the TD children from both minority and majority backgrounds exhibited comparable NWR accuracy at the segment correct level. Next, the effect of lexical scores was also significant, and children who scored higher during the lexical task also obtained better scores during the NWR task. This finding provides credence to the notion that vocabulary size/lexical knowledge is positively associated with NWR performance (e.g., Bowey, 1996; Coady & Evans, 2008; Farabolini et al., 2021; Metsala, 1999). This positive association between lexical knowledge and NWR accuracy has been explained in the literature in terms of a lexical redintegration process (S. E. Gathercole, 2006; Jones & Witherstone, 2011). Since children with DLD do not possess strong enough linguistic or lexical competence, they may be unable to use redintegration strategies to retrieve assistance during NWR tasks (Graf-Estes et al., 2007; Stokes et al., 2006). Furthermore, item length was also found to be inversely proportional to NWR accuracy. The accuracy declined as the item length increased. Adding more to this new tool's clinical utility was its robustness in the face of different participant-related factors, namely age, gender, nonverbal ability, and maternal education.

5.3 Empirical and Clinical Significance of This Thesis

The present thesis contributes findings to the field of heritage minority language acquisition from an understudied language, Urdu. This language is not only common in Hong Kong (spoken by 24% of the total South Asian people living in Hong Kong, Census and Statistics Department, Government of the Hong Kong SAR, 2021) but is also ranked as the 10th most spoken language in the world (Eberhard et al., 2022). Despite its prevalence, child language studies on Urdu have been scarce in the current child language acquisition literature. The present thesis assesses children acquiring L1 Urdu in minority vs. majority contexts in a wide range of linguistic domains (NWR, lexicon, morphosyntax, and narratives). Moreover, this thesis also presents an online diagnostic protocol to improve the identification of multilingual children with DLD, a condition that despite its high prevalence, stable trajectory (McKean et al., 2017), and enduring impact (Clegg et al., 2005) is still under-researched and underserved (McGregor, 2020). Below, the empirical and clinical significance of the three studies are presented sequentially.

Study 1 studied patterns of L1 Urdu development in multilingual children acquiring it under different acquisitional circumstances (i.e., minority vs. majority). It contributes knowledge related to the *L1 strengths and vulnerabilities of TD Urdu-speaking EM children*. Given their minority status, these EM children are acquiring their L1 (Urdu) under reduced input conditions. In such scenarios, it is not uncommon that certain L1 linguistic abilities might undergo delay, incomplete acquisition, or even attrition. The main objective of this study was to elucidate the comprehensive linguistic profiles of these EM children by testing them in a *range of domains, including lexical (comprehension and production), morphosyntactic (production), and narrative (comprehension and production)* abilities.

Furthermore, a group of TD age- gender- and grade-matched control peers acquiring L1 Urdu as a majority language in Pakistan was also assessed in the same tasks to specify similarities and differences in linguistic competencies and vulnerabilities between Urdu-speaking children acquiring L1 under minority vs. majority contexts. These findings bear on conceptual issues in heritage language acquisition, such as maintenance and incomplete acquisition, which have gained increasing attention in the international research community. The set of findings from this study also provides *some useful reference data* that are anticipated to be practical for practitioners and parents working with EM multilingual children in terms of highlighting their linguistic weaknesses for focused intervention.

Study 2 presents a unique case study highlighting the promise of using the CATALISE framework (Bishop et al., 2017) and LITMUS tools in tandem to identify children with language disorders using remote online testing. This study used Urdu-Cantonese multilingual EM children in Hong Kong as an illustrative example to demonstrate *how the CATALISE framework, with the support of the LITMUS tools, can be used to identify DLD in multilingual children and establish reference data*. Importantly, the online versions of these four LITMUS tests, including the CLT (Haman et al., 2015; Hamdani et al., 2020a), SRep (Marinis & Armon-Lotem, 2015; Hamdani et al., 2020b), MAIN (Gagarina et al., 2019a; Hamdani et al., 2020c), and CL-NWR (Chiat, 2015; Hamdani et al., 2020d) are likely to have the potential to support the identification of multilingual DLD in any target language in terms of global impact.

To the best of my knowledge, this study is likely the *first to have adapted the LITMUS tools into Urdu*. So far, *no published research has documented the diagnostic potential/accuracy of these Urdu LITMUS tools*. This study is likely the first to address the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children in this major but understudied world language. This tutorial also presents the first published data on the Urdu materials as far as the LITMUS-Cross-linguistic Lexical Task (LITMUS-CLT), the Cross-linguistic Nonword Repetition (CL-NWR) and the LITMUS-Sentence Repetition (LITMUS-SRep) data are concerned. As for the Urdu MAIN, (Chan et al., 2023) published some TD data featuring another group of older elementary school children on their production of Urdu narratives using the Urdu MAIN, examining the relationship between narrative macrostructure and microstructure. Therefore, if we are referring to the Urdu MAIN data from (suspected) DLD children and new data addressing the clinical utility of the Urdu LITMUS tools in identifying DLD in Urdu-speaking children, this study also presents the first published data using the Urdu MAIN in this regard.

Study 3 documents the NWR abilities in Urdu-speaking children. In congruence with the crosslinguistic literature (e.g., Dutch: de Bree et al., 2007; French: Thordardottir & Brandeker, 2013; Italian: Dispaldro et al., 2013; Russian: Kavitskaya et al., 2011), the findings revealed that the *new Urdu CL-NWR tasks have the potential to distinguish between the Urdu-speaking children with and without DLD*. Adding more to the Urdu CL-NWR's clinical utility were the findings demonstrating its resistance against not only language input contexts (minority vs. majority) but other individual factors like nonverbal ability, gender, and maternal education.

It is essential to mention that till this point, there are virtually *no assessment tools and reference data* for assessing speech and language abilities (e.g., NWR, lexical, morphosyntactic, and narrative skills) of Urdu-speaking children that are widely accessible by the local and international research and clinical communities. This lack of suitable assessment tools and appropriate multilingual norms often leads to over- and under-identification of language impairment in EM multilingual children (Armon-Lotem, 2018; Bedore & Peña, 2008). The current studies tried to address these long-standing issues by providing professionals with a *set of assessment tools and useful reference data* from both EM and majority groups. This reference data based on multiple linguistic domains was gathered using Urdu versions of four different language assessment tools from the LITMUS battery developed under COST Action (mentioned above). These LITMUS tools, which are *available for free*, are uniquely designed to facilitate cross-cultural and cross-linguistic language assessments of multilingual children (Armon-Lotem & de Jong, 2015) and have already been adapted into numerous languages. These new Urdu LITMUS tools, together with some reference data, might help practitioners and researchers carry out appropriate speech and language assessments of Urdu-speaking children around the globe.

Furthermore, due to the global COVID-19 pandemic, all the testing was moved online. To make online testing feasible and well-controlled, my chief supervisor and I worked closely with the original authors of all four LITMUS tools. As a result, we created the digital versions of all the assessment tools mentioned above using PowerPoint with embedded audio and videos. These assessment tools might be useful to many professionals, as the materials developed to support online testing are free and can be flexibly adapted to suit any target language and ethnic group. The MAIN materials for online testing can be accessed for free via the official MAIN homepage. These *digital versions are new valuable resources* that can be used by SLTs and researchers to assess the speech and language abilities of Urdu-speaking children even when face-to-face in-person testing is not possible. Overall, online testing has the merit of being *less disrupted by pandemic rebounds or other uncontrollable/unforeseeable factors* in the future as well.

The studies in this thesis also assessed different participant-related variables (age, gender, nonverbal ability, amount of Urdu input, and maternal education) and domain-specific (e.g., elicitation modes (narrative), word category (lexical tasks), modality (lexical tasks), sentence structures (SRep), item length (NWR)) factors. The varying effects of these factors across multiple linguistic domains are discussed in the corresponding chapters of this thesis. Furthermore, all three studies recruited participants with normal nonverbal abilities. The intention was to ascertain that all the children were in the normal range and the variations in their language outcomes were at least not due to their low nonverbal cognitive abilities (assessed using analytical reasoning abilities via Raven's Progressive Matrices).

Furthermore, different studies in this thesis present preliminary findings showcasing which scoring schemes could better differentiate children with and without DLD. For instance, as far as this sample is concerned, scoring NWR performance at the segment correct level turned out to be more efficient in discerning children with DLD (Chapter 4). For assessing story complexity (narrative macrostructure), scoring schemes by Gagarina et al. (2019) and Maviş et al. (2016) appeared to be more efficient than Sheng et al. (2020) in differentiating multilingual children with and without DLD (Chapter 3).

Lastly, all three studies in this thesis have delimited their scope of investigation to elementary school-aged children from first to fourth grades. It has laid a promising foundation to extend to younger children studying in kindergartens for future research to promote early identification of language disorders and language learning needs in young children. Overall, the outcomes from the above-mentioned studies have provided promising groundwork for future research. The following section also outlines some of the implications for future research.

5.4 Implications for Future Work

This section discusses how future studies could extend this work.

Diagnostic Accuracy of Assessment Measures

The diagnostic accuracy of these new Urdu LITMUS assessment tools in identifying Urdu-speaking multilingual children with and without LD or DLD still needs to be established. For example, in Study 2 (Chapter 3), which presents a case study, there was only one child with DLD and seven age- and grade-matched TD peers. Even though this sample size is relatively small, this is not uncommon in language acquisition studies on multilingual EM children (e.g., Farndale et al., 2016; Hu et al., 2014; Serrano-Hidalgo, 2018). However, in terms of diagnostic accuracy measures for the Urdu-speaking multilingual children suspected of LD or DLD, further studies are needed to examine and compare the diagnostic accuracy of newly adapted Urdu LITMUS tools with larger sample sizes with different age bands, especially younger ages. Extending work to younger children is vital in promoting early identification of language disorders and language learning needs in younger children.

Due to the COVID-19 upsurge, data for all three studies were collected remotely using Urdu online versions of different LITMUS tools. Recently, there have been some efforts to evaluate the efficacy of telepractice for conducting diagnostic assessments to identify children with and without LD/DLD, and the results are encouraging. For example, Nelson and Plante (2022) compared the Test of Integrated Language and Literacy Skills (TILLS) results when administered in-person and online. Findings showed 96% agreement between the two testing modes in detecting children with and without language or literacy disorders. Eikerling et al. (2023) also validated the remote screening of DLD in multilingual Spanish-Italian-speaking children using a new web-based application called MuLiMi. Their results provided supporting evidence that the Spanish-Italian MuLiMi application has the diagnostic potential to discern children at risk for DLD. Pratt et al. (2022) also compared online and face-to-face narrative comprehension in Spanish and English using the MAIN (Gagarina et al., 2019a). The findings indicated a significant correlation between the online and offline narrative macrostructure comprehension scores.

However, the data in the present thesis were collected using remote online testing only. Therefore, future studies could collect data using face-to-face mode and compare the results from online and offline modes. This would help corroborate the present findings and the efficiency of digital and offline versions of the Urdu MAIN, CLT, SRep, and NWR tests in identifying children with LD/DLD and collecting reference data.

Scoring Schemes

Based on the findings from Study 2 (Chapter 3), some preliminary advice could be offered that the scoring schemes of Maviş et al. (2016) and Gagarina et al. (2019) used to assess story complexity (narrative macrostructure) appeared to be better than the scoring scheme of Sheng et al. (2020) in differentiating DLDs/TDs. However, it is hard to conclude from these findings whether the scoring scheme of Gagarina et al. (2019) is better than that of Maviş et al. (2016). Similarly, in the NWR study (Chapter 4), scoring at the segment correct level turned out to be more efficient in discriminating multilingual children with and without DLD. This finding, on the one hand, is in line with the results of Boerma et al. (2015). On the other hand, this contrasts with Dispaldro et al.'s (2013) claims, concluding that scoring at the whole item correct level was a better clinical screening method. Given the relatively smaller sample sizes of the present studies, it is hard to make any firm recommendations. Thus, further investigation of these observations, including larger sample sizes and examining different age ranges and diagnostic/classification accuracy, such as sensitivity and specificity, is warranted.

Moreover, practicality also needs to be considered if recommendations for clinicians are required to be made. While a scoring scheme with the finest level of granularity (for story complexity) and segment level (for NWR) are likely the most differentiating, they could also be slower/harder to score. More informative data in the future comparing the scoring schemes would allow us to consider both diagnostic accuracy and practicality, to make more concrete recommendations on which scoring system is ideal or good enough for clinicians in speech and language therapy clinics.

Intervention

Apart from improving assessment, future research could also pilot intervention approaches that may be suitable for these multilingual EM children with LD or DLD, a direction that requires years to develop down the road. Future studies could investigate further how the areas of linguistic vulnerabilities identified in this thesis, namely vocabulary (noun and verb comprehension and production) and grammar (gender and number marking, word order, case/postpositions), could inform the design of better interventions for EM children.

Adding Quality of Input as a Basic External Factor

Studies 1 and 3 analyzed the impact of the quantity of Urdu language input on the different language abilities of children. In addition to quantity, the quality of input is also signified as a crucial factor in heritage minority language acquisition and is suggested to be a positive predictor of overall language growth in multilingual children (Hoff, 2006; Huttenlocher et al., 1991). Quality of input refers to the richness and diversity of the content to which multilinguals are exposed, including the use of diverse and precise vocabularies, the use of diverse and complex sentence structures, and the use of different types of discourse depending on the context/topic in their daily routines (Jia & Paradis, 2015). It is suggested that future studies may also include some other fundamental external factors like quality of input as a predictor variable in order to evaluate how input richness and diversity shape multilingual children's language outcomes across various linguistic domains.

Including Findings from the other Languages of these Multilingual Children

This thesis focused on heritage minority language acquisition. Thus, it features multilingual children who happened to have a clear language dominance profile, with their first language, Urdu, being the strongest language and Cantonese and other languages being the second/weaker languages. Bearing the focus in mind, this thesis only concentrated on findings from L1 Urdu. Assessing the other languages of these multilingual children was outside the scope of the present thesis. However, in the future, it would be interesting to explore what is going on in the children's other languages, especially in children with more balanced language dominance profiles.

5.5 Conclusion

Despite being the 10th most spoken language in the world (Eberhard et al., 2022), Urdu has remained an understudied language in terms of language acquisition. This thesis has tried to address this gap by not only assessing the patterns of L1 Urdu development across multiple linguistic domains (NWR, lexicon, morphosyntax, narrative) in TD EM children (acquiring language under reduced input conditions) but also in their TD age- gender- and grade-matched peers from Pakistan (acquiring language under relatively abundant input conditions). Linguistic competence was assessed remotely at both expressive and comprehensive levels. The findings revealed that children from the TD EM group performed significantly worse than their TD Maj peers in tasks assessing expressive narrative microstructure (grammaticality: proportion of grammatical C-units), lexical comprehension and production, and sentence repetition targeting morphosyntactic competence. In general, results support the idea that some language abilities are more/less affected by input conditions: morphosyntactic (during both narrative production and SRep) and lexical competence were revealed as more sensitive to input effects. Whereas narrative comprehension, expressive narrative macrostructure (story structure and story complexity), expressive narrative microstructure (narrative length, lexical diversity, and syntactic complexity), use of ISTs, and NWR accuracy were relatively less dependent on language-specific experience. The findings showed that children acquiring the same language under different acquisition contexts exhibit unique linguistic profiles. Therefore, it is suggested that reference or normative data for children with different language acquisition backgrounds should be generated exclusively to avoid over- and under-identification of language impairment, especially in EM multilingual children acquiring both/all languages under reduced input conditions.

Moreover, the comparison between children with and without DLD demonstrated that children with DLD lagged behind their TD peers across multiple linguistic domains, including receptive vocabulary, expressive vocabulary, NWR, morphosyntax, narrative comprehension, and narrative macrostructure production (telling mode). These results demonstrate the feasibility of identifying children with DLD and generating reference data using the CATALISE diagnostic criteria in combination with the LITMUS tools via remote online testing. Overall, the findings from the above-mentioned studies could pave the way to high-impact and more wide-ranging research that would have strong and positive societal implications across the globe.

References

Altman, C., Armon-Lotem, S., Fichman, S. & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics, 37,* 165–193. https://doi.org/10.1017/S0142716415000466

Altman, C., Goldstein, T., & Armon-Lotem, S. (2017). Quantitative and qualitative differences in the lexical knowledge of monolingual and bilingual children on the LITMUS-CLT task. *Clinical Linguistics & Phonetics, 31*(11–12), 931–954. https://doi.org/10.1080/02699206.2017.1312533

Ambreen, S. (2023). *The development of speech sounds and phonetic complexity in Urdu-speaking children.* [Unpublished Doctoral thesis]. The University of Hong Kong.

American Speech-Language-Hearing Association. (2004). *Knowledge and skills needed by speech-language pathologists and audiologists to provide culturally and linguistically appropriate services.* Retrieved from www.asha.org/policy

Anderson, R. T. (2001). Lexical morphology and verb use in child first language loss: A preliminary case study investigation. *International Journal of Bilingualism, 5*(4)*,* 377 – 401. https://doi.org/10.1177%2F13670069010050040101

Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition task. *Clinical Linguistics & Phonetics*, *34*(4), 293–311. https://doi.org/10.1080/02699206.2019.1637458

Antonova Ünlü, E., & Li, W. (2016). Aspect acquisition in Russian as the weaker language: evidence from a Turkish–Russian child*. International Journal of Bilingualism, 20*(2), 210–228. http://doi.org/10.1177/1367006914552138

Antonova Ünlü, E., & Li, W. (2018). Examining the effect of reduced input on language development: the case of gender acquisition in Russian as a nondominant and dispreferred language by a bilingual Turkish–Russian child. *International Journal of Bilingualism, 22*(2), 215–233. http://doi.org/10.1177/1367006916666390

Applebee, A. N. (1978). *The child's concept of a story*. Chicago University Press.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49*(5), 970–983. http://dx.doi.org/10.1044/1092-4388(2006/070)

Armon-Lotem, S. (2018). SLI in bilingual development: How do we approach assessment. In A. Bar-On, D. Ravid, & E. Dattner (Eds.), *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives (*pp. 617-641). De Gruyter Mouton. https://doi.org/10.1515/9781614514909

Armon-Lotem, S., & de Jong, J. (2015). Introduction. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Methods for Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 1–24). Multilingual Matters.

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

Balčiūnienė, I., & Dabašinskienė, I. (2019). Language dominance in bilingual acquisition: A case study of narrative production in Lithuanian. *Eesti Rakenduslingvistika Ühingu Aastaraamat, 15*, 5–19. https://doi.org/10.5128/ERYa15.01

Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Language*, *36*, 338–351. https://doi.org/10.1177%2F0142723715626067

Bamberg, M., & Damrad-Frye, R. (1991). On the ability to provide evaluative comments: further explorations of children's narrative competencies. *Journal of Child Language*, *18*(3), 689–710. https://doi.org/10.1017/s0305000900011314

Bari, S., & Ajmal, S. (2016). The first flight: Acquisition of Urdu Consonants in Native Children. *International Journal of English and Education, 5*(4), 56–71.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001heipl

Bartsch, K., & Wellman, H.M. (1995). *Children talk about the mind*. Oxford University Press.

Bates, D., Mächler, M., Bolker, B., &Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children's imitative productions. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology* (pp. 207–217). Cambridge University Press.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–29. https://doi.org/10.2167/beb392.0

Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., & Lancee, W. (1996). Long-term consistency in speech/language profiles: Developmental and academic outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(6), 804–814. https://doi.org/10.1097/00004583-199606000-00021

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3-4), 129–181. https://doi.org/10.1515/tl-2013-0009

Berman, R. A., & Katzenberger, I. (2004). Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes*, *38*(1), 57–94. https://doi.org/10.1207/s15326950dp3801_3

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Routledge.

Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and codemixing in the weaker language: the Ivy hypothesis. *Bilingualism, 7,* 49–69. https://doi.org/10.1017/S1366728904001270

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13,* 525–531. https://doi.org/10.1017/S1366728909990423

Bishop, D. V. M., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology*, *23*(1), 25–46. https://doi.org/10.1348/026151004X20685

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*(2), 156–173. https://doi.org/ 10.1044/jshd.5202.156

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium 2 (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. https://doi.org/10.1111/jcpp.12721

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The Journal of Child Psychology and Psychiatry, 37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Blankenstijn, C., & Scheper, A. (2003). *Language impairment in Dutch-speaking children with psychiatric impairment.* Paper presented at the meeting of Society for Research in Child Development, Tampa, FL.

Blažienė A. (2015). Narrative of bilingual children: General productivity and lexical diversity. *Taikomoji Kalbotyra, 7*, 1–25. https://doi.org/10.15388/TK.2015.17492

Blom, E., Boerma, T., & De Jong, J. (2019). First language attrition and Developmental Language Disorder. In M. S. Schmid and B. Kopke (Eds.), *The Oxford handbook of language attrition* (pp. 108–120). Oxford University Press.

Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, *66*, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

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. https://10.1044/2015_JSLHR-L-15-0058

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, *51*(6), 626–638. https://doi.org/10.1111/1460-6984.12234

Boeschoten, H., & Verhoeven, L. (1986). Turkish language development of Turkish children in the Netherlands. In A. Aksu Koç & E. Erguvanli Taylan (Eds.), *Proceedings of the Turkish linguistics conference, August 9-10, 1984* (pp. 269–80). Boğaziçi University Publications No. 400.

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, *37*(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U., & Gagarina, N. (Eds.). (2020). *Developing narrative comprehension: Multilingual assessment instrument for narratives* (vol. 61). John Benjamins Publishing Company.

Bohnacker, U., Haddad, R., & Öberg, L. (2021). Arabic-Swedish-Speaking Children Living in Sweden: Vocabulary Skills in Relation to Age, SES and Language Exposure. *Journal of Home Language Research, 4*(1, 1–18. https://doi.org/10.16993/jhlr.37

Bohnacker, U., Lindgren, J., & Öztekin, B. (2016). Turkish-and German-speaking bilingual 4-to-6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research*, *1*, 17–41. http://hdl.handle.net/10092/12907

Bolonyai, A. (2007). (In)vulnerable agreement in incomplete bilingual L1 learners. *International Journal of Bilingualism, 11*(1), 3–23. https://doi.org/10.1177/13670069070110010201

Bonifacci, P., Barbieri, M., Tomassini, M., & Roch, M. (2018). In few words: linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language, 45*(1), 120–147. https://doi.org/10.1017/S0305000917000149

Bortolini, U., Caselli, M. C., & Leonard, L. B. (1997). Grammatical deficits in Italian-speaking children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 40*(4), 809–820. http://dx.doi.org/10.1044/jslhr.4004.809

Bosch, L., & N. Sebastián-Gallés. 2003. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46,* 217–243. https://1 0.1177/00238309030460020801

Bosma, E., & Blom, E. (2019). Language activities in a minority-majority language context: book reading at home is more important for Frisian than for Dutch. *Journal of Child Language, 47*(2), 289–308. https://doi.org/1 – 20. 10.1017/S0305000919000023

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21. https://doi.org/10.1191%2F0265659002ct224oa

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International journal of language & communication disorders*, *36*(4), 421–432. https://doi.org/10.1080/13682820110074971

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63*(1), 44–78. http://dx.doi.org/10.1006/jecp.1996.0042.

Bruner, J. (1986). *Actual Minds, Possible Worlds.* Harvard University Press.

Burns, F. A., de Villiers, P. A., Pearson, B. Z., & Champion, T. B. (2012). Dialect-neutral indices of narrative cohesion and evaluation. *Language, Speech, and Hearing Services in Schools*, *43*(2), 132–152. https://doi.org/10.1044/0161-1461(2011/10-01

Burris, S. E., & Brown, D. D. (2014). When all children comprehend: increasing the external validity of narrative comprehension development research. *Frontiers in Psychology*, *5*, 168. https://doi.org/10.3389/fpsyg.2014.00168

Bustos, C. (2020). Dominance Analysis. *R package version 2.0.0.* https://rdocumentation.org/packages/dominanceanalysis/versions/2.0.0

Butt, M. (2006). *Theories of case.* Cambridge University Press.

Cahill, P., Cleave, P., Asp, E., Squires, B., & Bird, E. K.-R. (2020). Measuring the complex syntax of school-aged children in language sample analysis: A known-groups validation study. *International Journal of Language & Communication Disorders*, *55*(5), 765–776. https://doi.org/10.1111/1460-6984.12562

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40*(3), 519–525. https://doi.org/10.1044/jslhr.4003.519

Census and Statistics Department, Government of the Hong Kong Special Administrative Region (2021). *2021 Population census – Summary Results*. Hong Kong Census and Statistics Department. https://www.census2021.gov.hk/en/publications.html

Chan, A., Chen, S., Hamdani, S., Z., Tse, B., & Cheng, K. (2023). Story telling in bilingual Urdu-Cantonese ethnic minority children in Hong Kong: macrostructure and its relation to microstructural linguistic skills. *Frontiers in Psychology*, *14*(2023). https://doi.org/10.3389/fpsyg.2023.924056

Chen, L., & Yan, R. (2011). Development and use of English evaluative expressions in narratives of Chinese–English bilinguals. *Bilingualism: Language and Cognition, 14*, 570–578. https://doi.org/10.1017/S1366728910000362

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. L. de Jong., & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 123–148). Multilingual Matters.

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests. *Journal of Speech, Language, and Hearing Research, 59*(5), 1179–1189. https://doi.org/10.1044/2016_JSLHRL-15-0293

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

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, morphology and complex syntax in successive bilingual children. *Linguistic Approaches to Bilingualism, 1*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism, 15*, 5–21. https://doi.org/10.1017/S1366728911000368

Chondrogianni, V., Butcher, M., & Garaffa, M. (2021). Supporting children with Developmental Language Disorder in Gaelic-medium primary education. In W. McLeod, A. Gunderloch, & R. Dunbar (Eds.), *Rannsachadh na Gàidhlig 2018 proceedings* (pp. 205–2018). Aberdeen University Press.

Clark, A., O'Hare, A., Watson, J., Cohen, W., Cowie, H., Elton, R., & Seckl, J. (2007). Severe receptive language disorder in childhood- familial aspects and long-term outcomes: results from a Scottish study. *Archives of Disease in Childhood*, *92*(7), 614–619. https://doi.org/10.1136/adc.2006.101758

Clark, E. V. (2009). *First language acquisition* (2nd ed.). Cambridge University Press.

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

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology, 89*(3), 183–213. http://dx.doi.org/10.1016/j.jecp.2004.07.004.

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

Coleman, J. J., Frymark, T., Franceschini, N. M., & Theodoros, D. G. (2015). Assessment and treatment of cognition and communication skills in adults with acquired brain injury via telepractice: A systematic review. *American Journal of Speech Language Pathology, 24*(2), 295–315. https://doi.org/10.1044/2015_AJSLP-14-0028

Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: cross sectional and longitudinal studies of bilingual toddlers. *Child Development, 77*, 712–735. http://dx.doi.org/10.1111/j.1467-8624.2006.00899.x

Core, C., Chaturvedy, S., & Martinez-Nadramia, D. (2017). The role of language experience in nonword repetition tasks in young bilingual Spanish-English speaking children. *Proceedings of the 41st Annual Boston University Conference on Language Development*. Cascadilla Press. Retrieved from http://www.lingref.com/bucld/41/BUCLD41-14.pdf.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research, 56*(5), 1637–1649. https://doi.org/10.1044/1092-4388(2013/11-0044)

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire.* Multilingual Matters.

Cummins, J. (2005). A proposal for action: strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *The Modern Language Journal, 89*, 585–592. http://www.jstor.org/stable/3588628

de Almeida, L., Ferrré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., Zebib, R., & Barthez, M. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism, 7*(3/4), 331–358. https://doi.org/10.1075/lab.15019.alm

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

De Houwer, A. (2005). Early bilingual acquisition: focus on morphosyntax and the separate development hypothesis, In F. Kroll & A. M. B. De Groot (Eds.), *The handbook of bilingualism* (pp. 30-48). Oxford University Press.

De Houwer, A. (2015). Harmonious bilingual development: young families' wellbeing in language contact situations. *International Journal of Bilingualism,* 19, 169–184. https://doi.org/10.1177/1367006913489202

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

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. https://doi.org/10.1017/S1366728915000012

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*(1), 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)

Dixon, C., Thomson, J., & Fricke, S. (2020). Language and reading development in children learning English as an additional language in primary school in England. *Journal of Research in Reading, 43*(3), 309–328. https://doi.org/10.1111/1467-9817.12305

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research, 42,* 1432–1443. https://doi.org/10.1044/jslhr.4206.1432

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

Dollaghan, C., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16,* 211–222. https://doi.org/10.1017/S0142716400007098

Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, *25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Eberhard, D. M., Gary, F. S., & Charles, D. F. (Eds.). (2022). Ethnologue: Languages of the world (25th ed.). SIL International. Retrieved from https://bit.ly/3Bwie5p

Ebert, K. D., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language Teaching*, *49*(3), 301–338. https://doi.org/10.1017/S0261444816000070

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. http://dx.doi.org/10.1044/1092-4388(2004/034).

Eikerling, M. R., Bloder, T. S., & Lorusso, M. L. (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-non-specific nonwords. *Frontiers in Psychology, 13,* 826540. https://doi.org/10.3389/fpsyg.2022

Eikerling, M., Andreoletti, M., Secco, M., Luculli, B., Cha, G., Castro, S., Gazzola, S., Sarti, D., Garzotto, F., Guasti, M. T., & Lorusso, M. L. (2023). Remote Screening for Developmental Language Disorder in Bilingual Children: Preliminary Validation in Spanish–Italian Speaking Preschool Children. *Applied Sciences, 13*(3), 1442. http://dx.doi.org/10.3390/app13031442

Elliott, L. L., Hammer, M. A., & Scholl, M. E. (1989). Fine-grained auditory discrimination in normal children and children with language-learning problems. *Journal of Speech, Language, and Hearing Research, 32*(1), 112–119. http://dx.doi.org/10.1044/jshr.3201.112

European COST Action. (2009–2013). *IS0804: Language impairment in a multilingual society: Linguistic patterns and the road to assessment.* Retrieved from https://www.cost.eu/actions/IS0804/

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism, 13,* 81–97. https://doi.org/10.1080/13670050902783528

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 https://doi.org/2050571X.2021.1879609

Farndale, A., Harris, P., de Courcy, M. (2016). Social, emotional, and linguistic experiences of developing bilingual preschoolers as they learn English as an Additional Language (EAL). *The International Journal of Learning, 23*, 41–58. https://doi.org/10.18848/1447-9494/CGP/41-58

Fichman, S., & Altman, C. (2019). Referential Cohesion in the Narratives of Bilingual and Monolingual Children with Typically Developing Language and With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 62*(1), 123–142. https://doi.org/10.1044/2018_JSLHR-L-18-0054

Fichman, S., Altman, C., Voloskovich, A., Armon-Lotem, S., & Walters, J. (2017). Story grammar elements and causal relations in the narratives of Russian-Hebrew bilingual children with SLI and typical language development. *Journal of Communication Disorders, 69,* 72–93. https://doi.org/10.1016/j.jcomdis.2017.08.001

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: language and task effects. *Language, Speech, and Hearing Services in Schools, 35*(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016

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

Flores, C. (2012). Differential effects of language attrition in the domains of verb placement and object expression. *Bilingualism: Language and Cognition, 13*, 533–546. https://doi.org/10.1017/s1366728911000666

Friend, M., Lopez, O., De Anda, S., Abreu-Mendoza, R. A., & Arias-Trejo, N. (2022). Maternal education revisited: Vocabulary growth in English and Spanish from 16 to 30 months of age. *Infant Behavior & Development, 66,* 101685. https://doi.org/10.1016/j.infbeh.2021.101685

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

Fu, N.C., Chen, S., Polišenská, K., Chan, A., Kan, R., & Chiat S. (2024). *Nonword Repetition in Children with Developmental Language Disorder: Revisiting the Case of Cantonese* [manuscript under review]. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong.

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, *37*(1), 91–122.

Gagarina, N., & Klassert, A. (2018). Input dominance and home language development in Russian-German bilinguals. *Frontiers in Communication, 3*(40), 1–14. https://doi.org/10.3389/fcomm.2018.00040

Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., Golcher, F., & Walters, J. (2014). Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian Hebrew preschool children. In R. K. Silbereisen, P. F. Titzmann, & Y. Shovit (Eds.), *The challenges of diaspora migration: Interdisciplinary perspectives on Israel and Germany* (pp. 63-82). Ashgate.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. *ZAS Papers in Linguistics*, *56*, 155.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U. & Walters, J. (2019a). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics, 63*(20). https://doi.org/10.21248/zaspil.63.2019.516

Gagarina, N., Gey, S., & Sürmeli, N. (2019b). Identifying early preschool bilinguals at risk of DLD: A composite profile of narrative and sentence repetition skills. *ZAS Papers in Linguistics, 62,* 168–189. https://doi.org/10.21248/zaspil.62.2019.448

Gagarina, N., Klop, D., Tsimpli, I., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics, 37*(1), 11–17. https://doi.org/10.1017/S0142716415000399

Gámez, P. B., & González, D. A. (2019). Comparison of narrative skill in Spanish-English bilinguals and their functionally monolingual Spanish-speaking and English-only peers. *International Journal of Bilingualism, 23*(1), 329–348. https://doi.org/10.1177/1367006917728391

Ganuza, N., & Hedman, C. (2019). The impact of mother tongue instruction on the development of biliteracy: Evidence from Somali–Swedish bilinguals. *Applied Linguistics,* 40(1), 108–131. DOI: https://doi.org/10.1093/applin/amx010

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

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419. https://doi.org/10.1016/S13646613(99)01388-1

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

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200–213. http://dx.doi.org/10.1016/0749-596X(89)90044-2

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

Gathercole, V. C. M. (2002). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 207–219). Multilingual Matters.

Gathercole, V. C. M. (Ed.). (2007). Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism, 10*(3), 224–247. https://doi.org/10.2167/beb442.0

Gathercole, V. C. M. (Ed.). (2013). *Issues in the Assessment of Bilinguals*. Multilingual Matters.

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, *12*(2), 213–237. https://doi.org/10.1017/S1366728909004015

Gathercole, V. C. M., Thomas, E. M., Roberts, E. J., Hughes, C. O., & Hughes, E. K. (2013). Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. In V. C. M. Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 20–55). Multilingual Matters.

Genesee, F., & Nicoladis, E. (2007). Bilingual first language acquisition. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 324–342). Blackwell.

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), 551–60. https://doi.org/10.3109/13682820903198058

Goldfield, B. A. (2000). Nouns before verbs in comprehension vs. production: The view from pragmatics. *Journal of Child Language, 27,* 501–520. https://doi.org/10.1017/s0305000900004244

Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 63–80). Lawrence Erlbaum Associates, Inc.

Graf-Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: *A meta-analysis. Journal of Speech, Language, and Hearing Research, 50*(1), 177–195. http://dx.doi.org/10.1044/1092-4388(2007/015)

Grimm, A. (2022). The use of the LITMUS quasi-universal nonword repetition task to identify DLD in monolingual and early second language learners aged 8 to 10. *Languages, 7*(3), 218. https://doi.org/10.3390/languages7030218

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology, 23*, 574–586.

Gu, M. M., & Patkin, J. (2013). Heritage and identity: Ethnic minority students from South Asia in Hong Kong. *Linguistics and Education, 24*(2), 131–141. https://doi.org/10.1016/j.linged.2012.12.008

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: a comparison with monolinguals and second language learners. *Applied Psycholinguistics, 29,* 3–19. https://doi.org/10.1017/S0142716408080016

Guven, C., & Islam, A. (2015). Age at migration, language proficiency, and socioeconomic outcomes: evidence from Australia. *Demography, 52*, 513–542. https://doi.org/10.1007/s13524-015-0373-6

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 196–240). Multilingual Matters.

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., & Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in Psychology, 8,* 1444. https://doi.org/10.3389/fpsyg.2017.01444

Hamann, C., Chilla, S., Ibrahim, L. A., & Fekete, I. (2020). Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Applied Psycholinguistics*, *41*(6), 1375–1414. https://doi.org/10.1017/S0142716420000399

Hamdani, S., Kan, R., & Chan, A. (2020a). Crosslinguistic Lexical Tasks: Urdu version (CLT-UR). *A part of LITMUS COST IS0804 Battery* [Unpublished material]. Chinese and Bilingual Studies, The Hong Kong Polytechnic University.

Hamdani, S., Chan, A., Kan, R., Armon-Lotem, S. (2020b). *Sentence Repetition Task: Urdu Version* [Unpublished material]. Chinese and Bilingual Studies, The Hong Kong Polytechnic University.

Hamdani, S., Kan, R., Chan, A. & Gagarina, N. (2020c). The Multilingual Assessment Instrument for Narratives (MAIN): Adding Urdu to MAIN. *ZAS Papers in Linguistics, 64,* 257–261. https://10.21248/zaspil.64.2020.580

Hamdani, S., Chan, A., Kan, R., Chiat, S., & Polišenská, K. (2020d). *Urdu Crosslinguistic Nonword Repetition Test* [Unpublished material]. Chinese and Bilingual Studies, The Hong Kong Polytechnic University.

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology*, *19*(2), 154–166.

Heritage. (2022). *In The American Heritage College Dictionary*. Retrieved from https://www.ahdictionary.com/

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2014). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism, 19*(6), 677– 692. https://doi.org/10.1177/1367006914534330

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E. (2020). Lessons from the study of input effects on bilingual development. *International Journal of Bilingualism, 24*(1), 82–88. https://doi.org/10.1177/1367006918768370

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*, 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E., Quinn, J. M., & Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, *21*(2), e12536. https://doi.org/10.1111/desc.12536

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive Vocabulary Development in Children from Bilingual and Monolingual Homes: A Longitudinal Study from Two to Four Years. *Early Childhood Research Quarterly, 29*, 433–444. https://doi.org/10.1016/J.Ecresq.2014.04.012

Hothorn, T. (2022). Simultaneous Inference in General Parametric Models. *R package version 1.4–19.* https://cran.r-project.org/packages/multcomp

Hreich, E. K., & Messarra, C. (2013). *Assessment of L2 children in the Lebanese multilingual context: Using questionnaire for parents of bilingual children (PABIQ) scores as diagnostic indicators for language impairment* [Unpublished Manuscript]. Saint Joseph's University, Beirut.

Hu, J., Torr, J., & Whiteman, P. (2014). Australian Chinese parents' language attitudes and practices relating to their children's bilingual development prior to school. *Journal of Early Childhood Research, 12*(2), 139–153. https://doi.org/10.1177/1476718X13515429

Huls, S.K. (2017). *Bilingual children's performance on three nonword repetition tasks: the role of language experience and ability* [Doctoral dissertation]. University of Kansa, USA. Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/26333/Huls_ku_0099D_15325_DATA_1.pdf

Hulsen, M. (2000). *Language Loss and Language Processing: Three Generations of Dutch Migrants in New Zealand* [Unpublished doctoral dissertation]. University of Nijmegen.

Humayoun, M., Hammarström, H., & Ranta, A. (2007). Urdu Morphology, Orthography and Lexicon Extraction. In *CAASL-2: The Second Workshop on Computational Approaches to Arabic Script-based Languages.* LSA 2007 Linguistic Institute, Stanford University.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://psycnet.apa.org/doi/10.1037/0012-1649.27.2.236

International Expert Panel on Multilingual Children's Speech. (2012). *Multilingual children with speech sound disorders: Position paper.* Research Institute for Professional Practice, Learning and Education, Charles Sturt University, Australia.

Isurin, L., & Ivanova-Sullivan, T. (2008). Lost in between: the case of Russian heritage speakers. *Heritage Language Journal, 6,* 72–104. https://doi.org/10.46538/hlj.6.1.4

Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, *18*(4), 737–752. https://doi.org/10.1017/S1366728914000728

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 Journal of Speech-Language Pathology, 19 (1),* 51–65. https://doi.org/10.1044/1058-0360(2009/08-0083)

Jones, G., & Witherstone, H. L. (2011). Lexical and sublexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition, 39*(4), 588–599. http://dx.doi.org/10.3758/s13421-010-0045-0

Kapalková, S., & Slančová, D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics & Phonetics*, *31*(11–12), 893–909. https://doi.org/10.1080/02699206.2017.1308016

Kauschke, C., Lee, H.-W., & Pae, S. (2007). Similarities and variation in noun and verb acquisition: A crosslinguistic study of children learning German, Korean, and Turkish. *Language and Cognitive Processes, 22*(7), 1045–1072. https://doi.org/10.1080/01690960701307348

Kavitskaya, D., Babyonyshev, M., Walls, T., & Grigorenko, E. (2011). Investigating the effects of syllable complexity in Russian-speaking children with SLI. *Journal of Child Language, 38*(5), 979–998. https://doi.org/10.1017/S0305000910000413

Kehoe, M., & Havy, M. (2019). Bilingual phonological acquisition: the influence of language-internal, language-external, and lexical factors. *Journal of Child Language, 46*(2), 292–333 https://dx.doi.org/S0305000918000478

Kim, J.-H., Montrul, S., & Yoon, J. (2009). Binding interpretation of anaphors in Korean heritage speakers *Language Acquisition, 16,* 3–35. https://www.jstor.org/stable/20462541

Kohnert, K. (2004). Processing skills in early sequential bilinguals. In B. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (pp. 53–76). Brookes Publishing Company.

Kohnert, K., Windsor, J., & Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals? *Learning Disabilities Research and Practice, 21*(1), 19–29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Komeili, M., & Marshall, C. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical Linguistics and Phonetics*, *27*, 152–162.

Komeili, M., Marinis, T., Tavakoli, P., & Kazemi, Y. (2020). Sentence Repetition in Farsi-English Bilingual Children. *Journal of the European Second Language Association, 4*(1), 1–12. http://doi.org/10.22599/jesla.55

Kraljević, J.K., Hržica, G., & Gorup, I.V. (2020). A comparative macrostructural analysis of narrative discourse in children with typical language development and children with Developmental Language Disorder. *Drustvena Istrazivanja, 29*, 453–470. https://doi.org/10.5559/di.29.3.06

Kulkarni, A. A., Chadd, K. E., Lambert, S. B., Earl, G., Longhurst, L. M., McKean, C., Hulme, C., McGregor, K. K., Cunniff, A., Pagnamenta, E., Joffe, V., Ebbels, S. E., Bangera, S., Wallinger, J., & Norbury, C. F. (2022). Editorial perspective: Speaking up for developmental language disorder – the top 10 priorities for research. *Journal of Child Psychology and Psychiatry, 63*(8), 957–960. https://doi.org/10.1111/jcpp.13592

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, *37*(1), 123–144. https://doi.org/10.1017/S0142716415000442

Labov, W. (1972). *Language in the inner city: studies in the black English vernacular*. University of Pennsylvania Press.

Labov, W. (2013). *The language of life and death: The transformation of experience in oral narrative.* Cambridge University Press.

Lee, H. J., Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International journal of speech-language pathology*, *15*(4), 375–382. https://doi.org/10.3109/17549507.2012.752866

Lenneberg, E. (1967), *Biological Foundations of Language.* Wiley.

Lenth, R. (2022). Estimated marginal means, aka least-squares means. *R package version 1.7.5.* https://cran.r-project.org/package=emmeans

Leseman, P. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Multilingual and Multicultural Development, 21*(2), 93–112. https://doi.org/10.1080/01434630008666396

Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT Press.

Li'el, N., Williams, C., & Kane, R. (2019). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology, 21*(6), 613–622. https://doi.org/10.1080/17549507.2018.1513073

Liles, B. Z. (1993). Narrative discourse in children with language disorder and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, *36*(5), 868–882. https://doi.org/10.1044/jshr.3605.868

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6* [Doctoral dissertation]. Uppsala Universiteit. Uppsala. Retrieved from https://bit.ly/3pQHmNs

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely-related languages: Age, word type and cognate facilitation effects in bilingual Swedish-German preschool children. *Linguistic Approaches to Bilingualism, 10*(5), 587–622. https://doi.org/10.1075/lab.18041.lin

Loban, W. (1976). *Language development: kindergarten through grade twelve.* National Council of Teachers of English.

Lynch, J. S., Van Den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, *29*(4), 327–365. https://doi.org/10.1080/02702710802165416

MacLeod, A. A. N., Castellanos-Ryan, N., Parent, S., Jacques, S., & Séguin, J. R. (2019). Modelling vocabulary development among multilingual children prior to and following the transition to school entry. *International Journal of Bilingual Education and Bilingualism, 22*(4), 473–492. https://doi.org/10.1080/13670050.2016.1269718

Manhardt, J., & Rescorla, L. (2002). Oral narrative skills of late talkers at ages 8 and 9. *Applied Psycholinguistics, 23*(1), 1–21. https://doi.org/10.1017/S0142716402000012

Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research, 60*(12), 3462–3473. https://doi.org/10.1044/ 2017_JSLHR-L-15-0415

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. L. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 116–143). Multilingual Matters.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. https://doi.org/oi:10.3758/s13428-011-0168-7

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, *37*(1), 69–89. https://doi.org/10.1017/S0142716415000429

McGregor, K.K. (2020). How we fail children with developmental language disorder. *Language, Speech & Hearing Services in Schools, 51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017). Subgroups in language trajectories from 4 to 11 years: The nature and predictors of stable, improving and decreasing language trajectory groups. *Journal of Child Psychology and Psychiatry, 58*(10), 1081–1091. https://doi.org/10.1111/jcpp.12790

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology, 8*, 1442. https://dx.doi.org/10.3389%2Ffpsyg.2017.01442

Meir, N., & Janssen, B. (2021). Child Heritage Language Development: An Interplay Between Cross-Linguistic Influence and Language-External Factors. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.651730

Meir, N., Walters, J., & Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition task: the impact of L1 and L2 properties. *International Journal of Bilingualism, 20*(4), 421–452. https://doi.org/10.1177/1367006915609240

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105*(4), 306–323. http://dx.doi.org/10.1016/j.jecp.2009.12.006

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3–19. http://dx.doi.org/10.1037/0022-0663.91.1.3.

Meziane, R. S., & MacLeod, A. A. (2021). Evidence of phonological transfer in bilingual preschoolers who speak Arabic and French. *International Journal of Bilingualism, 25*(6), 1680–1695. https://doi.org/10.1177/13670069211034232

Miękisz, A., Haman, E., Łuniewska, M., Kuś, K., O'Toole, C. & Katsos, N. (2017). Impact of first-generation immigrant environment on the heritage language: Productive vocabulary of Polish toddlers living in the UK and Ireland. *International Journal of Bilingual Education and Bilingualism, 20*(2), 183–200. https://doi.org/10.1080/13670050.2016.1179259

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L. & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*(3), 30-43. https://doi.org/10.1111/j.1540-5826.2006.00205.x

Miller, P. H., & Aloise, P. A. (1989). Young children's understanding of the psychological causes of behaviour: A review. *Child Development, 60* (2), 257–285. https://doi.org/10.2307/1130975

Montgomery J. (2004). Sentence comprehension in children with specific language impairment: effects of input rate and phonological working memory. *International journal of language & communication disorders, 39*(1), 115–133. https://doi.org/10.1080/13682820310001616985

Montgomery, J. W. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics, 16*(4), 355–378. https://doi.org/10.1017/S0142716400065991

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. John Benjamins.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, 2–23. https://doi.org/10.1017/S0267190510000103

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Montrul, S. (2018). Heritage language development: Connecting the dots. *International Journal of Bilingualism, 22*(5), 530–546. https://doi.org/10.1177/1367006916654368

Montrul, S. (2022). The development of the heritage language in childhood bi-/multilingualism. In A. Stavans & U. Jessner (Eds.), *The Cambridge Handbook of Childhood Multilingualism* (pp. 537–555). Cambridge University Press.

Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English Bilingual Children. *International Journal of Bilingualism, 11*(3), 301–328. https://doi.org/10.1177/13670069070110030301

Munson, B. (2001). Phonological Pattern Frequency and Speech Production in Adults and Children. *Journal of Speech, Language, and Hearing Research, 44*(4), 778–792. https://doi.org/10.1044/1092-4388(2001/061

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Nelson, N. W., & Plante, E. (2022). Evaluating the Equivalence of Telepractice and Traditional Administration of the Test of Integrated Language and Literacy Skills. *Language, Speech, and Hearing Services in Schools, 53*(2), 376–390. https://doi.org/10.1044/2022_LSHSS-21-00056

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

O'Toole, C., Gatt, D., Hickey, T. M., Haman, E., Mi˛ekisz, A., & Armon-Lotem, S. (2017). Parent report of early lexical production in bilingual children across varied contexts: a CDI study. *International Journal of Bilingual Education and Bilingualism*, *20*(2), 124–145. https://doi.org/10.1080/13670050.2016.1179258

Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Language Learning, 70*, 15–53. https://doi.org/10.1111/lang.12347

Ortiz J. A. (2021). Using nonword repetition to identify language impairment in bilingual children: A meta-analysis of diagnostic accuracy. *American Journal of Speech-Language Pathology*, *30*(5), 2275–2295.

https://doi.org/10.1044/2021_AJSLP-20-00237

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M. & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, *23*(9), 1083–1107. https://doi.org/10.1080/13670050.2018.1434124

Öztekin, B. (2019). Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7. *Studia Linguistica Upsaliensia 25*. Uppsala University.

Paradis, J. (2009). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 387–405). Blackwell Publishing.

Paradis, J. (2010). Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning, 60*(3), 651–680. https://doi.org/10.1111/j.1467-9922.2010.00567.x

Paradis, J. (2016). The development of English as a second language with and without specific language impairment: Clinical implications. Journal of Speech, Language, and Hearing Research, 59(1), 171–182. https://doi.org/10.1044/2015_JSLHR-L-15-0008.

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language,* 1–25.

Paradis, J., & Kirova, A. (2014). English second-language learners in preschool: profile effects in their English abilities and the role of home language environment. *International Journal of Behavioural Development, 38*, 342–349. https://doi.org/10.1177/0165025414530630

Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, *41*(6), 1255–1281. https://doi.org/10.1017/S014271642000017X

Paradis, J., Soto-Corominas, A., Daskalaki, E., Chen, X., & Gottardo, A. (2021). Morphosyntactic Development in First Generation Arabic—English Children: The Effect of Cognitive, Age, and Input Factors over Time and across Languages. *Languages, 6*(1), 51. MDPI AG. Retrieved from https://doi.org/10.3390/languages6010051

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish–English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*(1), 113–25. https://dx.doi.org/10.1016%2Fj.jecp.2010.07.011

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Multilingual Matters.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58. https://doi.org/10.1017/S0142716400009863

Peña, E. D., Bedore, L. M., Lugo-Neris, M. J., & Albudoor, N. (2020). Identifying Developmental Language Disorder in School Age Bilinguals: Semantics, Grammar, and Narratives. *Language Assessment Quarterly, 17*(5), 541–558. https://doi.org/10.1080/15434303.2020.1827258

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English–Spanish Assessment (BESA)*. Brookes.

Penfield, W. (1953). A consideration of the neurophysiological mechanisms of speech and some educational consequences. *Proceedings of the American Academy of Arts and Sciences, 82,* 199 – 214. https://doi.org/10.2307/20023716

Peristeri, E., Andreou, M., Tsimpli, I. M., & Durrleman, S. (2020). Bilingualism effects in the narrative comprehension of children with Developmental Language Disorder and L2-Greek: Links with language, executive function and Theory of Mind. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives.* (pp. 297–330). John Benjamins. https://doi.org/10.1075/sibil.61.05kun

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. Plenum Press.

Polinsky, M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics, 14*, 191–262. http://www.jstor.org/stable/24599616

Polinsky, M. (2008). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal*, *6*, 40–71. http://nrs.harvard.edu/urn-3:HUL.InstRepos:3382970

Polinsky, M. (2011). Reanalysis in adult heritage language: New evidence in support of attrition. *Studies in Second Language Acquisition*, *33*(2), 305–328. https://doi.org/10.1017/S027226311000077X

Polinsky, M. (2016). Bilingual children and adult heritage speakers: the range of comparison. *International Journal of Bilingualism, 22*, 547–563. https://doi.org/10.1177/1367006916656048

Polinsky, M. (2018). *Heritage Languages and Their Speakers.* Cambridge University Press.

Polišenská, K., & Kapalková, S. (2014). Improving Child Compliance on a Computer-Administered Nonword Repetition Task. *Journal of Speech, Language & Hearing Research, 57*(3), 1060–1068. https://doi.org/10.1044/1092-4388(2013/13-0014

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure? *International Journal of Language & Communication Disorders, 50*(1), 106–118. https://doi.org/10.1111/1460-6984.12126

Poon, A. Y. K. (2010). Language use, and language policy and planning in Hong Kong. Current *Issues in Language Planning, 11(1)*, 1-66. http://dx.doi.org/10.1080/14664201003682327

Potowski, K. (2013). Language maintenance and shift. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics* (pp. 321–339). Oxford University Press.

Pratt, A. S., Anaya, J. B., Ramos, M. N., Pham, G., Muñoz, M., Bedore, L. M., & Peña, E. D. (2022). From a distance: comparison of in-person and virtual assessments with adult-child dyads from linguistically diverse backgrounds. *Language, Speech, And Hearing Services in Schools, 53*(2), 360–375. https://doi.org/10.1044/2021_LSHSS-21-00070

Prevoo, M. J., Malda, M., Mesman, J., Emmen, R. A., Yeniad, N., Van Ijzendoorn, M. H., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: different pathways for host and ethnic language. *Journal of Child Language*, *41*(5), 963–984. https://doi.org/10.1017/S0305000913000299

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/

Raven, J. C., & Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales* (vol. 759). Oxford Psychologists Press.

Reid, H. (2019). *Nonword repetition and sequential memory in children aged 7–11 and adults – differential effects of bilingualism, developmental language disorders, and the role of development* [Master's thesis, McGill University, Montreal]. Retrieved from https://escholarship.mcgill.ca/downloads/sb397d584

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2; 6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, *58*(2), 345–359. https://doi.org/ 10.1044/2015_JSLHR-L-14-0150

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders, 47*(5), 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Rizvi, S. M. J. (2007). *Development of algorithms and computational grammar for Urdu* [Doctoral dissertation]. Institute of Engineering and Applied Sciences, Islamabad, Pakistan. Retrieved from https://doi.org/10.13140/RG.2.1.4129.3846

Roch, M., & Hržica, G. (2020). Narrative comprehension by Croatian-Italian bilingual children 5–7 years old. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (pp. 171–196). Amsterdam: John Benjamins.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, *37*(1), 49–67.

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism, 21*(5), 617–635. https://doi.org/10.1177/F1367006916643528

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism, 13*, 155–163. https://doi.org/10.1177/1367006909339814

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47*(1), 223–234. https://doi.org/10.1044/1092-4388(2004/019)

Roy, P., & Chiat, S. (2013). Teasing apart disadvantage from disorder: The case of poor language. In C.R. Marshall (Ed.), *Current Issues in Developmental Disorders* (pp.125-150). Psychology Press.

Saliby, C. K. A., dos Santos, C., Hreich, E. K., & Messarra, C. (2017). Assessing Lebanese bilingual children: The use of cross-linguistic lexical tasks in Lebanese Arabic. *Clinical Linguistics & Phonetics, 31*(11–12), 874–892. https://doi.org/10.1080/02699206.2017.1308016

Scherger, A. (2022). The role of age and timing in bilingual assessment: non-word repetition, subject-verb agreement and case marking in L1 and eL2 children with and without SLI. *Clinical Linguistics & Phonetics, 36*(1), 54–74. https://doi.org/10.1080/02699206.2021.1885497

Schmidt, R. L. (1999). *Urdu: An essential grammar.* Routledge.

Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., & Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *International Journal of Bilingualism, 19*(6), 726–752. https://doi.org/10.1177/1367006914544989

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*(2), 168–175. http://dx.doi.org/10.3758/BF03202729

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Scontras, G., Fuchs, Z., & Polinsky, M. (2015). Heritage language and linguistic theory. *Frontiers in Psychology,* 6, 01545. https://doi.org/10.3389/fpsyg.2015.01545

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders, 45*(6), 691–702. https://doi.org/10.3109/13682820903509432

Serrano-Hidalgo, E. (2018). A case study of the experiences and perspectives of Hispanic immigrant parents on heritage language maintenance and bilingual education in the rural community of Brandon, Manitoba. *BU Journal of Graduate Studies in Education, 10*(2), 4–7. https://eric.ed.gov/?id=EJ1230328

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, *27*(6), 960–974. http://dx.doi.org/10.1037/0012-1649.27.6.960

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. https://doi.org/10.1177/0265659012465208

Sheng, L., Lu, Y., & Kan, P. F. (2011). Lexical development in Mandarin-English bilingual children. *Bilingualism: Language and Cognition, 14(4),* 579–587. http://dx.doi.org/10.1017/S1366728910000647

Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative Production in Mandarin-Speaking Children: Effects of Language Ability and Elicitation Method. *Journal of Speech, Language, and Hearing Research, 63*(3), 774–792. https://doi.org/10.1044/2019_JSLHR-19-00087

Shields, M. A., & Price, S. W. (2002). The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics*, *15*(1), 137–160. https://doi.org/10.1007/PL00003836

Silva, A. C., Ferreira, A. A., & de Queirgo, B. A. (2014). Development of oral narrative and education level of mother. *Revista CEFAC, 16*(1)*,* 174–186. http://dx.doi.org/10.1590/1982-021620146412

Simon-Cereijido, G., & Gutierrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *30*, 315–337. http://dx.doi.org/10.1017/S0142716409090134

Simonsen, H. G., & Haman, E. (2017). LITMUS-CLT: A new way to assess bilingual lexicons. *Clinical Linguistics & Phonetics, 31*(11-12), 811-817. http://dx.doi.org/10.1080/02699206.2017.1307454

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12*(3), 369–373. http://dx.doi.org/10.1017/S0142716400009279

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gilliam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders, 49*(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Ablex.

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech and Hearing Research, 49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, *41*(2), 407–418. https://doi.org/10.1044/jslhr.4102.407

Strand, S., Malmberg, L., & Hall, J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. http://hdl.handle.net/10871/23323

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, & Communication Disorders, 45*(4), 480–493. https://doi.org/10.3109/13682820903198058

Sutherland, R., Hodge, A., Trembath, D., Drevensek, S., & Roberts, J. (2016). Overcoming barriers to using telehealth for standardized language assessments. *Perspectives of the ASHA Special Interest Groups, 1(*18), 41–50. https://doi.org/10.1044/persp1.SIG18.41

Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare, 23*(1), 106–115. https://doi.org/10.1177/1357633X15623921

Syed, A. Z. (2013). *Redefining Urdu morphology and grammar for the development of an integrated sentiment analysis framework* [Doctoral dissertation]. University of Engineering and Technology, Lahore, Pakistan. Retrieved from http://prr.hec.gov.pk/jspui/bitstream/123456789/2223/1/2773S.pdf

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition, 179,* 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

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

Taylor, H. G., Lean, D., & Schwartz, S. (1989). Pseudoword repetition ability in learning disabled children. *Applied Psycholinguistics, 10,* 203–219. https://doi.org/10.1017/S0142716400008523

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*(4), 922–937. https://doi.org/10.1044/1092-4388(2008/068)

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. https://doi.org/10.1016/j.jcomdis.2012.08.002

Thordardottir, E., & Reid, H. (2022). Nonword repetition to identify DLD in older school-age children: A French nonword repetition test with word-like and nonword-like items from 2 to 8 syllables. *Enfance, 1*(1), 59–79. https://doi.org/10.3917/enf2.221.0059

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research*, *40*(6), 1245–1260. https://10.1044/jslhr.4006.1245

Trabasso, T., & Nickels, M. (1992). The development of goal plan action in the narration of a picture story. *Discourse Processes, 15*(3), 294–275. https://doi.org/10.1080/01638539209544812

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goal/plans: A conceptual basis for narrating "Frog where are you? In R. A. Berman & D. I. Slobin (Eds.), *Relating Events in Narrative: A Crosslinguistic Developmental Study* (pp. 85–106). Lawrence Erlbaum.

Trauner, D., Wulfeck, B., Tallal, P., & Hesselink, J. (2000). Neurological and MRI profiles of children with developmental language impairment. *Developmental Medicine & Child Neurology, 42*(7), 470–475. http://dx.doi.org/10.1111/j.1469-8749.2000.tb00350.x.

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics, 37(1),* 195–216. https://doi.org/10.1017/s0142716415000478

Tuller, L. (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. In S. Armon-Lotem, J. de Jong, N. de Meir (Eds.), *Assessing Multilingual Children* (pp. 301–330). Multilingual Matters.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders, 53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*(3), 225–236. https://doi.org/10.1044/0161-1461(2007/024)

van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. J. Bauer & T. Bourg (Eds.), *Developmental spans in event comprehension and representation* (pp. 321–342). Lawrence Erlbaum Associates.

Verhoeven, L., Steenge, J., van Weerdenburg, M., & van Balkom, H. (2011). Assessment of second language proficiency in bilingual children with specific language impairment: A clinical perspective. *Research in Developmental Disabilities*, *32*(5), 1798–1807. https://doi.org/10.1016/j.ridd.2011.03.010

Vettori, G., Bigozzi, B., Incognito, O., & Pinto, G. (2022). Relationship between oral narrative and vocabulary skills among bilingual language-minority children and their monolingual peers in primary school. *International Journal of Bilingualism.* https://doi.org/10.1177/13670069221089064

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language, 52*(2), 193–204. http://dx.doi.org/10.1016/j.jml.2004.10.003

Washington, J. A., & Craig, H. K. (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology, 13*(4), 329–340. https://doi.org/10.1044/1058-0360(2004/033)

Weidner, K., & Lowman, J. (2020). Telepractice for adult speech language pathology services: A systematic review. *Perspectives of the ASHA Special Interest Groups, 5*(1), 326–338. https://doi.org/10.1044/2019_PERSP-19-00146

Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language and Hearing Research, 43,* 865–878. https://doi.org/10.1044/jslhr.4304.865

Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and Reading Disabilities* (pp. 157–232). Allyn & Bacon.

Westby, C. E. (2012). Assessing and remediating text comprehension problems. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities*. (3rd ed., pp. 163–225). Pearson.

Wiley, T. (2008). Chinese "dialect" speakers as heritage language learners: A case study. In D. Brinton, O. Kagan and S. Bauckus (Eds.), *Heritage Language Education: A New Field Emerging* (pp. 91–106). Routledge.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American journal of speech-language pathology, 19*(4), 298–310. https://doi.org/10.1044/1058-0360(2010/09-0064)

Winsler, A., Díaz, R. M., Espinosa, L., & Rodriguez, J. L. (1999). When learning a second language does not mean losing the first: Bilingual language development in low-income, Spanish-speaking children attending bilingual preschool. *Child Development, 70,* 349–362. https://doi.org/10.1111/1467-8624.t01-1-00026

Witkowska, D., Lucas, L., Jelen, M., Kin, H., & Norbury, C. (2022). Development of complex syntax in the narratives of children with English as an Additional Language and their monolingual peers. *PsyArXiv.* https://doi.org/10.31234/osf.io/96dx5

Wright, A. J. (2018). Equivalence of remote, online administration and traditional, face-to-face administration of the Woodcock-Johnson IV cognitive and achievement tests. *Archives of Assessment Psychology, 8*(1), 23–35.

Yu, B. (2013). Issues in bilingualism and heritage language maintenance: Perspectives of minority-language mothers of children with autism spectrum disorders. *American Journal of Speech Language Pathology, 22,* 10–24. https://doi.org/10.1044/1058-0360(2012/10-0078)

Zahra, N. (2016). *Acquisition of Urdu Consonants in Typically Developing Children.* [Unpublished Master's thesis]. Centre for Clinical Psychology, University of the Punjab, Lahore, Pakistan.

Zamuner, T. S. (2009). Phonotactic probabilities at the onset of language development: Speech production and word position. *Journal of Speech, Language, and Hearing Research, 52*(1), 49–60. http://dx.doi.org/10.1044/1092-4388(2008/07-0138)

Supplementary Materials

Table A *ISTs in MAIN*

IST Types	English Examples	Urdu Examples
Perceptual state terms	see, hear	دیکھا، سنا
Physiological state terms	thirsty, hungry	پیاسا، بھوکا
Consciousness terms	awake, asleep	جاگنا، سونا
Emotion terms	glad, happy, sad, worried, surprised, brave	خوش، اداس، پریشان، حیران، بہادر
Mental verbs	want, think, believe	چاہتا، سوچا، فیصلہ کیا
Linguistic verbs	say, ask, call	کہا، بولا، بلایا

Table B *Structure and Stimuli of the LITMUS-SRep Test*

Block	Number of structures ($N = 10$)	Number of items ($N = 35$)	Number of words in each sentence	Number of syllables in each sentence
1	4	17	6–8	8–12
2	4	12	6–9	9–13
3	2	6	9–10	12–15

Table C *The Three Scoring Schemes for Structural Complexity*

Studies	Scoring methods	Complete episodes	Incomplete episodes		Action or reaction sequence	Isolated description	
		GAO	G	GA, GO	AO	A, O	none
Maviş, et al. 2016)	Weighting system was employed. The maximum score for each episode is 3 points.	3 points	2 points		1 point	0 point	
Sheng et al. (2020)	Dichotomized scores were adopted for each episode. Only episodes that included a goal statement can receive 1 point.	1 point	1 point		0 point	0 point	
Gagarina et al. (2019)	Weighting system was employed. The maximum score for each episode is 6 points.	6 points	3 points	4 points	2 points	1 point	0 point

Note. G = Goal; A = Attempt; O = Outcome

Table D *Narrative Text to Elaborate Scoring (Participant TD-6: Baby Bird Story-Retelling)*

C-Units	Original Urdu Story	English Translation	SS	Episodes	IST Tokens	SC-M	SC-Sh	SC-Ga
1	دو بچے تھے۔	There were two babies.						
2	ان کی ماں ان کو دیکھ رہی تھی کہ ان کو بھوک لگی ہے۔	Their mother was looking at them that they were hungry.	IST as IE (1)		2			
3	تو ماں ان کو ان کے لئے کھانا ڈھونڈنے چلی گئی۔	Then mother went to find food for them them.	A (1)	AO		1		2
4	پھر ایک کتا دیکھ رہا تھا کہ میں اس میں کیا دیکھ رہا ہوں۔	Then a dog was looking that what am I seeing in there.	IST as IE (2)		2			
5	فر اس اس ان کی ماں ان کے لئے ایک کیڑا پکڑ کے لائی۔	Then him him their mother brought a worm for them.	O (1)					
6	اس بلی نے بلی نے ایک بچے کو پکڑ لیا۔	That cat cat grabbed one baby.	O (2)	O				1
7	فر ایک کتا دیکھ رہا تھا وہ اس پرندے کو۔ غصے والا۔	Then a dog was looking at that bird. Angrily.	IST as IE (3)		2			
8	اور اس نے فیصلہ کیا کہ میں بچا لیتا ہوں۔	And he decided to rescue.	G (3)	GAO	1	3	1	6
9	فر اس کتے نے اس کی ٹیل پکڑ لی۔	Then that dog caught his tail.	A (3)					
10	فر اس کے پیچھے بھاگا۔	Then ran after him.						
11	فر پھر وہ کتا اس کے پیچھے لگ جاتا ہے۔	Then Then that dog follows him.						
12	فر فر وہ وہاں سے چلا جاتا ہے۔	Then then he goes from there.	O (3)					
Total Scores			9		7	4	1	9

Note. SS = Story Structure; G = Goal; A = Attempt; O = Outcome; IST = Internal State Terms; IE = Initiating Event; SC-M = Story Complexity by Maviş et al. (2016); SC-Sh = Story Complexity by Sheng et al. (2020); SC-Ga = Story Complexity by Gagarina et al. (2019); Mazes; IST Tokens.

Figure A The Cat Story

Figure B The Dog Story

Figure C The Baby Birds Story

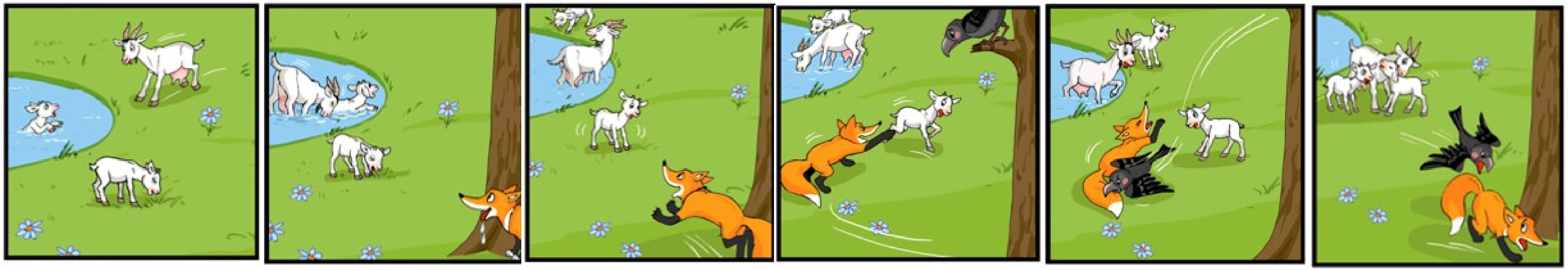

Figure D The Baby Goats Story

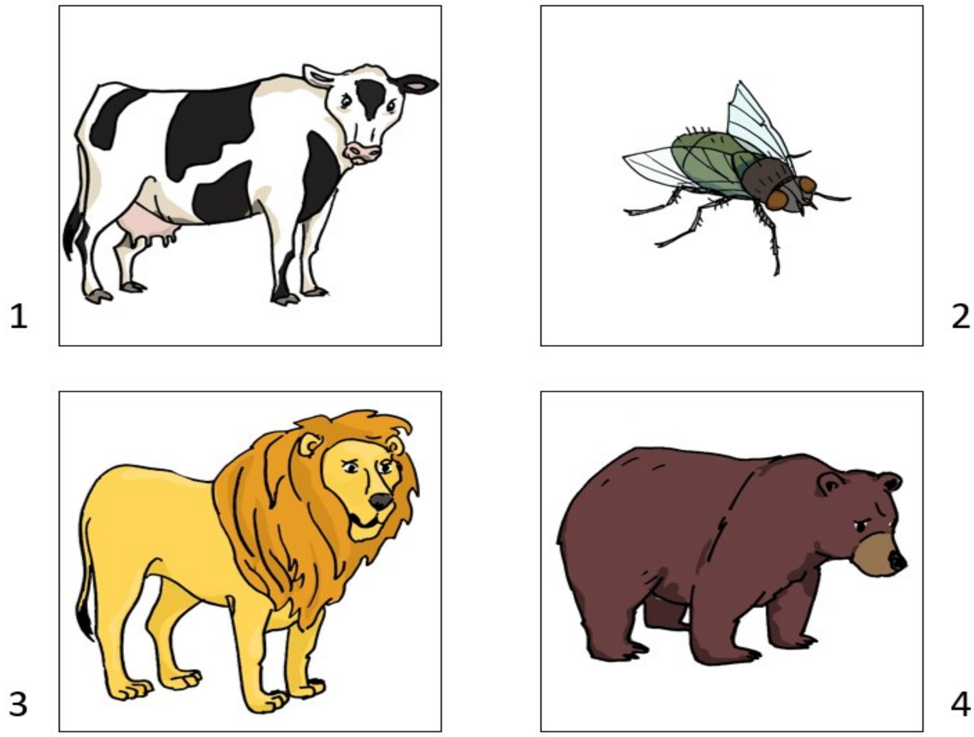

Figure E Noun Comprehension Task in CLT (Selecting the Correct Object)

Figure F Verb Comprehension Task in CLT (Selecting the Correct Action)

Figure G Noun Production Task in CLT (Naming Object)

Figure H Verb Production Task in CLT (Naming Action)

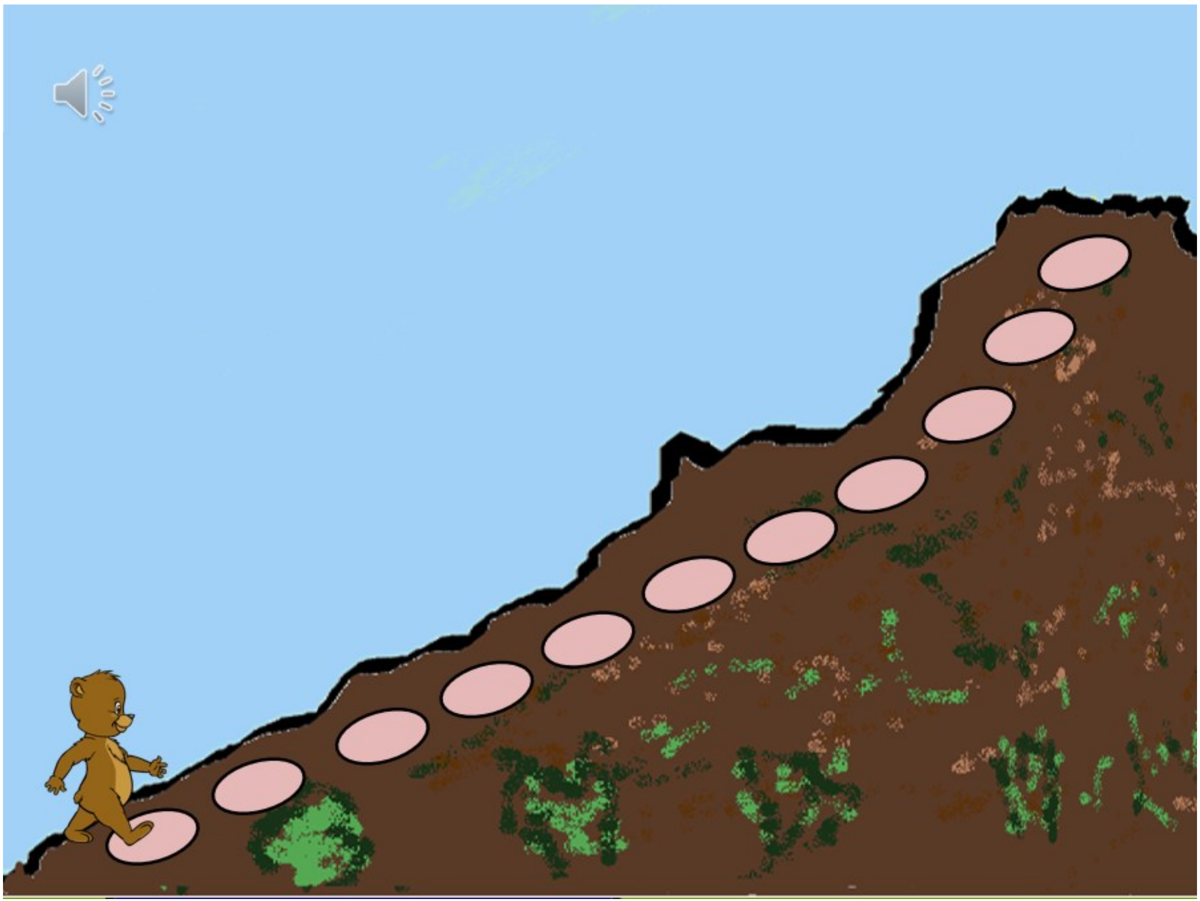

Figure I LITMUS-SRep Task

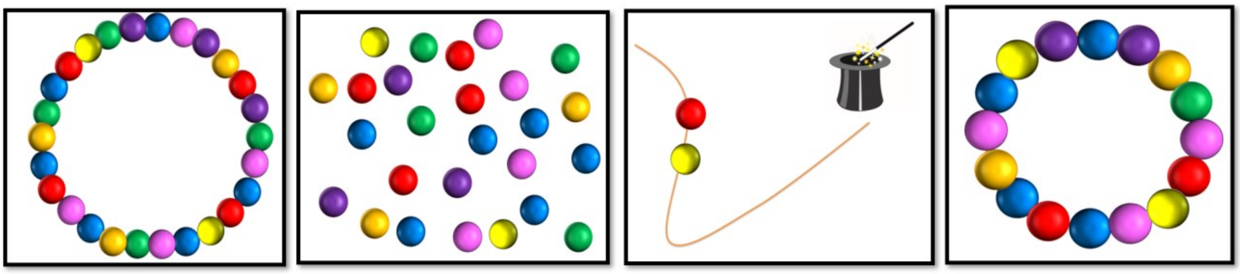

Figure J CL-NWR Task (Polišenská & Kapalková, 2014)