

Article

Special Needs Assessment in Bilingual School-Age Children in Germany

Irina Hertel , Solveig Chilla  and Lina Abed Ibrahim 

Institute for Special Education, Europa-Universität Flensburg, 24943 Flensburg, Germany; solveig.chilla@uni-flensburg.de (S.C.); lina.abed-ibrahim@uni-flensburg.de (L.A.I.)

* Correspondence: irina.hertel@uni-flensburg.de

Abstract: Educational and (psycho-)linguistic research on L1 and L2 acquisition in bilingual children sketches them as a group of language learners varying in many aspects. However, most studies to date have based evaluations of language proficiency or new assessment tools on data from heritage children, while studies on the appropriateness of assessment tools for school-age refugee children remain a notable exception. This study focuses on the standardized assessment tool BUEGA for primary school children, which is, among others, a widespread tool for the assessment of pedagogical support or special needs (SN) in Germany. We compare the performance of 12 typically developing monolinguals (MoTD: 7;3–12;1), 14 heritage-bilinguals (BiTD: 7;1–13;4, L1 Turkish and Arabic), 12 refugee- students (BiTD: 8;7–13;1, L1 Arabic), and 7 children with developmental language disorders (DLD: 7;7–13;9) on the subtests of grammar, word-reading, and spelling. Overall results show that refugee-BiTDs perform in the (monolingual) pathology range. No significant differences emerged between students with DLD and typically developing (TD) refugee students. Considering the assessment of school-related language performance, bilingual refugees are at risk of misdiagnosis, along with the well-known effects of educational disadvantage. This particularly applies to children with low socioeconomic status (SES). Looking beyond oral language competencies and using test combinations can help exclude language disorders in school-age children with limited L2 proficiency.

Keywords: bilingualism; special needs; refugee and heritage speakers; late child L2



Citation: Hertel, Irina, Solveig Chilla, and Lina Abed Ibrahim. 2022. Special Needs Assessment in Bilingual School-Age Children in Germany.

Languages 7: 4. <https://doi.org/10.3390/languages7010004>

Academic Editors: Dina Tsagari and Henrik Bohn

Received: 30 September 2021

Accepted: 16 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In Germany, the assessment of language skills plays an important role across all stages and sectors of education. Not only does it play a key role in determining whether a child is eligible for special needs categories¹ (SN) and enables language support recommendations, but it also informs decisions affecting the crucial transition from kindergarten to elementary and from elementary to secondary school. In most Federal States of Germany, students with SN visit special schools or courses, with negative implications for career opportunities in adulthood. In most states, students are referred to a suitable school type² by teachers after grade four, mainly based on their abilities in German and math (see Ahrenholz et al. 2016). This transition is viewed as a critical event in each student's educational path, since a referral to the *Hauptschule* or the *Realschule* rules out or at least complicates the possibility of obtaining a higher-level school leaving certificate, e.g., the general university entrance qualification usually obtained at the *Gymnasium*. In particular, children with SN are significantly less likely to be recommended school types with subsequent access to higher education. Such an early allocation of students based on school performance is viewed with criticism by the research community since it contributes to strengthening the link between social background and educational outcomes in Germany (Ditton 2011).

In Germany, 40% of children under the age of 18 have a migrant background, a significant number of whom grow up acquiring German as their second language (L2) (BPB 2021). Disparities in literacy skills and academic language proficiency in German have been

argued to contribute to the educational inequalities that children from migrant backgrounds encounter. Academic language skills are generally considered cognitively more demanding than everyday language since they are context-reduced and characterized by conceptual literality, i.e., sharing specific lexical and morphosyntactic properties with written discourse. Hence, they tend to show a significantly slower acquisition rate relative to oral language skills (Cummins 2008; Feilke 2012; Gogolin and Duarte 2016). Two decades of large-scale assessments and numerous studies (e.g., Klieme et al. 2010) show that, in addition to having to acquire a second language, people with a migrant background often have low socioeconomic status (SES). This puts their children at a higher risk of educational failure. They are more likely to have a delayed school entry (Autorengruppe Bildungsberichterstattung 2020), repeat a class grade, and/or obtain overall lower qualifications (Olczyk et al. 2016). Furthermore, bilingual children appear to be over-represented in schools for students with SN and support networks (cf. Powell and Wagner 2014), and a substantial proportion of them exit school without obtaining a school-leaving certificate, particularly those with low SES (Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration 2019; Bertelsmann Stiftung 2015).

German teachers employ a range of standardized language tests to assess the L2 abilities of bilingual children. However, most of these tests are not tailored to the linguistic diversity in German schools; in other words, these tests do not include appropriate bilingual norms that consider the language experience of the bilingual child (Elsner 2015). The very few available exceptions, e.g., the German LiSe-DaZ (Schulz and Tracy 2011) and the Russian SKRUK (Gagarina et al. 2010), were normed with simultaneous bilinguals, who acquired both languages from birth or shortly after (2L1), or with early successive child bilinguals (eL2)³, who started acquiring their L2 at kindergarten age (Meisel 1990; 2009). However, they do not consider school-age children with a considerably later age of onset (AoO) for exposure to the L2 (>5;0), i.e., late successive child bilinguals (IL2) (Chilla 2008). The latter scenario particularly applies to a significant proportion of the refugee children and adolescents who recently entered the German school system as asylum seekers from Syria, Afghanistan, and Iraq, 17–27% of whom are minors aged 6–17 (Statistisches Bundesamt 2021).

In light of the lack of adequate language assessment tools and the absence of uniform and specific recommendations for the formative language assessment of recently immigrated school-age children (KMK 2016), teachers have no choice but to resort to the standard assessment tools used to examine monolingual students to assess heritage (2L1 and eL2) and refugee children (Elsner 2015). Yet, unlike heritage children who have been exposed to the societal language for an extended period of time and are thus very likely to be dominant in the L2 upon school entry, IL2 child refugees face the challenge of learning academic content and acquiring literacy skills at school at the same time as they are still learning the societal language and might still demonstrate weaker L2 oral proficiency relative to their monolingual and heritage peers. Consequently, refugee children might be at a particular disadvantage when it comes to school performance since oral language proficiency in one language supports the development of literacy and academic skills in the same language (cf. Quigley et al. 2020). Hence, the ultimate goal of this study is to provide practical recommendations for the assessment of L2 abilities which capitalize on the strengths of typically developing late-L2 learners while maximizing the gap for children who are true candidates for Special Needs in the field of speech and language education (*Sonderpädagogischer Förderbedarf Sprache*), most of whom are children with developmental language disorders (DLD). By investigating not only grammar but also word-reading (word decoding) and spelling abilities that are crucial for the development of literacy and reading comprehension skills (Gough and Tunmer 1986; Nation and Norbury 2005), we want to explore whether looking beyond oral language competencies, as well as using test combinations, can help exclude language disorders in school-age children with limited L2 proficiency. The latter area is of particular interest since language-impaired individuals have been reported to show word decoding problems and impaired reading abilities that

persist into adolescence (Botting 2020; Palikara et al. 2011). Conversely, typically developing IL2 children are likely to profit from their existing L1 knowledge and metalinguistic awareness and demonstrate better word-reading abilities than their language-impaired peers (Durgunoğlu 2002). Moreover, IL2 children are speculated to benefit from their older AoO and the associated greater cognitive resources in acquiring certain linguistic phenomena and reading skills, especially if they are literate in their L1 (Gottardo et al. 2020; Rothman et al. 2016).

1.1. Identification of Special Language Needs in Heritage and Refugee Bilingual Children

1.1.1. Sources of Individual Variation in Child L2 Acquisition

Research on bilingual language acquisition has identified multiple mutually non-exclusive factors contributing to individual differences in child L2 acquisition (see Armon-Lotem and Meir 2019; Chondrogianni 2018; Paradis 2011; Unsworth 2016 for overviews). In addition to factors known to influence language development in monolingual acquisition, e.g., age, working memory, and cognitive capacities, bilingual children experience significantly more variation in their language environment, resulting in differences in individual outcomes in each of their languages (Kohnert 2013; Unsworth 2016). Age and input factors influencing bilingual language development include the age of onset (AoO) for systematic (sustained) exposure to the L2 (see Birdsong 2018 for a discussion), the length of exposure (LoE) to the L2, and quantitative and qualitative aspects of linguistic input (Unsworth 2019), i.e., amount of exposure and linguistic richness. Another relevant factor is socioeconomic status, a broad language environment factor that modulates the quantitative and qualitative aspects of the linguistic input the children receive (Paradis 2011; Prevoo et al. 2014). In studies on bilingual language acquisition, SES is frequently operationalized as years of maternal education (e.g., Armon-Lotem et al. 2011; Paradis and Jia 2017; Duncan and Paradis 2020). This is because mothers serve as the primary caregivers and thus constitute the main source of language input at home, especially in German migrant settings SES has been found to particularly impact lexical diversity and the acquisition of (complex) morphosyntax (Cztinglar et al. 2015; Hoff et al. 2002; Hoff 2003; Vernon-Feagans et al. 2019). For example, Paradis and Jia (2017) found that environmental factors, such as language exposure, the mother's education, the mother's English fluency, the child's use of English in the home, and the richness/quality of the English input outside of school, differentially predicted outcomes of L2 proficiency in both younger and older children (8;6–10;6). The interplay of these factors makes it notoriously difficult to establish what is typical for bilingual language development (Tuller et al. 2018).

1.1.2. Developmental Language Disorder (DLD)

Developmental language disorder (also known as Specific Language Impairment, SLI; Bishop et al. 2017) is a life-long neurodevelopmental disorder affecting the process of language acquisition that cannot be ascribed to primary deficits such as hearing loss, cognitive disability, neurological deficits, or autism spectrum disorder. With a prevalence rate of 7.4%, DLD constitutes one of the most common developmental communication disorders affecting both monolingual and bilingual children (Leonard 2014a; Norbury et al. 2016; Tomblin et al. 1997). DLD is associated with a range of expressive and receptive language deficits, and clinical markers could manifest differently in different languages (Leonard 2014b). Although DLD primarily affects morphosyntactic (Marinis 2011) and phonological development (dos Santos and Ferré 2018), children with DLD also evince lexical/semantic deficits (Novogrodsky and Kreiser 2015; Schulz and Grimm 2020) as well as deficits in aspects of pragmatic and narrative abilities (Davies et al. 2016; Tsimpli et al. 2016). Moreover, a child's diagnosis with DLD is associated with a significant risk for developing reading impairments and lower literacy levels (Catts et al. 2014), putting affected children at an increased risk of school failure and developing socio-emotional and behavioral problems (Yew and O'Kearney 2013). Thus, early identification and therapeutic intervention are crucial for ameliorating the long-term outcomes of DLD.

1.1.3. Bilingualism and DLD

When it comes to language assessment in bilingual children, much caution should be taken due to the great variability in their (typical) language development, which is determined by a myriad of the child's internal and external factors (de Grüter and Paradis 2014; Hamann 2012). The assessment of bilingual children is further complicated by the (temporary) overlap of bilingual error patterns and errors serving as clinical markers for DLD in a particular language (Chilla 2008; Paradis 2010). This leads to both overdiagnosis and underdiagnosis with developmental language disorder (DLD) (Bedore and Peña 2008; Genesee et al. 2004; Rothweiler 2006) in bilingual children. The assessment of child refugees is even more complicated compared to their heritage age-peers. In addition to the well-documented sources of individual variation in child bilinguals, such as AoO to the L2 and quantitative and qualitative aspects of language input, refugee children are subject to unique risk factors that could have adverse effects on their "overall development, including language development" (Paradis et al. 2021, p. 2). The latter include interrupted schooling and factors affecting their socio-emotional wellbeing, such as exposure to violence, poverty, frequent transitions, trauma, and difficulties adapting to the new schooling system as well as to the linguistic and cultural environment (Graham et al. 2017; Hadfield et al. 2017; Kaplan et al. 2016).

On the other hand, given their later AoO to the L2, IL2 refugee children are likely to benefit from their previously acquired L1 knowledge, literacy skills, and metalinguistic awareness (phonological and morphological awareness). Moreover, their more advanced cognitive resources (working memory and analytic reasoning abilities) could play a facilitative role in the acquisition of certain L2 linguistic phenomena, such as complex constructions. As a result, they are expected to show better L2 word-reading abilities than children with DLD (Durgunoğlu 2002; Schiff and Saiegh-Haddad 2018), especially if they are literate in the L1 (Gottardo et al. 2020).

Within the Canadian refugee context, Al-Janaideh et al. (2020) reported very poor reading and oral performance in first-generation Syrian refugees (6;0–13;0) in both their L1 and L2, which they attributed to low levels of SES, insufficient language exposure (<3 years) and richness in the home environment, as well as signs of emotional trauma. In another study on the latter sample, Paradis et al. (2021) examined, in a longitudinal approach, the morphosyntactic development in Arabic-L1 and English-L2 using an English sentence-repetition task (SRT). Their results indicated that age and input factors, among other factors such as cognition, have differential effects on both the L1 and L2. A third study by Gottardo et al. (2020), within the same project, investigated L1 and L2 literacy skills in Syrian child refugees (ages 6;0–13;0) with short LoE to English (<18 months) and found effects of phonological and morphological awareness on reading skills within and across languages. Their findings suggest "that the learners' L1 linguistic and metalinguistic skills, which are the linguistic skills most accessible to beginner L2 learners", play a facilitative role in acquiring L2 literacy skills. The first studies investigating L2 development in school-age refugee children within the German school context highlight the "difficulties in choosing and applying language assessment tools in the absence of valid norms or of comparable L2 reference groups" (Hamann et al. 2020, p. 1377). In a study with primary school-age refugees (ages 10;0–17;0), Montanari and Abel (2017) registered a significant gap in the performance of refugees relative to their heritage peers on measures assessing vocabulary development and (picture-based) essay-writing. In the same vein, Abed Ibrahim et al. (2020) and Hamann et al. (2020) investigated a group of younger school-age refugees (ages 6;6–12;8) and found significant deficits in the performance of refugees on measures of vocabulary and morphosyntax relative to younger heritage bilinguals, even on the LiSe-DaZ test (Schulz and Tracy 2011), which offers bilingual norms for eL2 children.

1.2. Attempting Solutions to the Assessment Challenges

Even though an increasing number of publications mention the educational disadvantages for heritage and refugee children with German as an L2, it remains an open

question of how to effectively support bilingual children, with academic language skills in particular (cf. Edele et al. 2020). The first step towards providing targeted intervention and support for children with SN is to determine whether poor L2 language skills result from inborn language impairment, i.e., developmental language disorder (DLD), or are the consequence of emerging bilingualism. Yet, the lack of assessment tools normed for bilingual populations, paired with the partially overlapping linguistic profiles of bilingual children and monolingual children with developmental language disorder (DLD), make it very difficult to disentangle genuine language impairment from low language proficiency as a result of insufficient exposure to the language of assessment. Since all children are expected to partake in standardized assessment procedures, much caution should be taken in the case of bilinguals, particularly when interpreting the results of refugee populations since their language deficits might arise from input factors rather than from DLD (Andreou and Garyfallia 2020). Given the crucial role of early identification and intervention for ameliorating the long-term outcomes of DLD, there is an urgent need for effective assessment tools that do not involve waiting until the child has had sufficient exposure to the L2, especially in the case of late-L2 bilinguals.

For excluding language disorders in bilingual children, it is considered best practice to assess the child in both languages (IALP 2011) or at least in the dominant language (Fredman 2006). Especially in the case of late successive bilinguals with limited exposure to the L2, L1 assessment could become crucial for excluding DLD. IL2 children are likely to retain superior abilities in their L1 given that they experienced a more extended period of monolingual exposure in a qualitatively rich language environment and probably had access to formal language registers via schooling (Paradis et al. 2021; Montrul 2016). However, assessment in the L1 is not always feasible due to the lack of L1-speaking practitioners and adequate tests for bilingual children acquiring their first language (L1) in heritage contexts, i.e., in contexts where the L1 is not the majority language of the society (Rothman 2009). This leaves educators with no other options but to resort to direct assessment measures in the L2.

Standardized test procedures designed for monolingual children can be informative within the school context when the aim of the assessment is to identify specific areas of language difficulty in need of additional in-school support, e.g., certain linguistic structures. However, if the assessment goal is to rule out language disorder, the child's overall language skills need to be compared to other bilingual children with similar language acquisition conditions. Different proposals were made to cope with diagnostic challenges in bilingual populations. One suggestion for utilizing monolingual tools is adjusting the norms for bilingualism according to the status of the language being tested as the dominant or weaker language; see the recommendations by Thordardottir (2015). However, the latter approach was proposed for simultaneous child bilinguals and might thus be unsuitable for use with other bilingual populations, including IL2 refugee children. For example, Hamann et al. (2020) showed that despite applying dominance-adjusted cut-off scores, a significant proportion of their refugee sample scored in the pathology range on standardized tests assessing L2 vocabulary (WWT, Glück 2011) and L2 morphosyntax (TROG-D Fox 2009).

During COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment" (<https://www.bi-sli.org>, accessed on 21 September 2021), cross-linguistically valid tools were developed in an attempt to cope with the diagnostic challenges in bilingual populations. These tools are known as the LITMUS tools (see Armon-Lotem et al. 2015) and were devised to minimize the effects of factors related to bilingualism on task performance so that DLD can be reliably identified in bilingual contexts. Of particular relevance to this study are sentence repetition (LITMUS-SRT, and Marinis and Armon-Lotem 2015; Hamann et al. 2013) and quasi-universal nonword repetition tasks (LITMUS-QU-NWR, Grimm et al. 2014). Such tasks have been shown to reliably identify DLD in monolingual children (Conti-Ramsden et al. 2001) and are frequently part of standardized assessment measures.

Multiple studies featuring simultaneous-bilingual and early-L2 heritage bilinguals have recently reported good to excellent diagnostic accuracy for nonword and sentence repetition tasks in bilingual children with diverse L1 backgrounds, especially when used in combination (a.o. [Abed Ibrahim and Fekete 2019](#); [Armon-Lotem and Meir 2016](#); [Chiat and Polišenská 2016](#); [Chilla et al. 2021](#)). These results are encouraging since these tools are easy and fast to administer and could be employed as first screening tools in bilingual contexts, including in schools. However, little research has been conducted on the efficacy of LITMUS repetition tools for identifying DLD in children with a later AoO to the L2 (>5;0 years), as is the case in the refugee population. In an attempt to bridge this gap, [Hamann et al. \(2020\)](#) compared the performance of 15 school-age Syrian refugees (6;6–12;8) with an LoE > 18 months to 12 2L1 and eL2 heritage speakers (6;0–12;9) on the German LITMUS quasi-universal NWRT (LITMUS-QU-NWRT, [Grimm et al. 2014](#)) and LITMUS-SRT ([Hamann et al. 2013](#)). Results showed comparable performance between heritage and refugee bilingual children only for LITMUS-QU-NWRT, regardless of age and input factors. In contrast, only heritage bilinguals showed adequate performance on the German LITMUS-SRT. Refugee children with fewer than 24 months of exposure to German performed below the cut-offs for DLD that were previously established for younger heritage bilinguals (cf. [Hamann and Abed Ibrahim 2017](#)). Since morphosyntactic competence “does not develop as fast as many schooling and integration models presuppose”, [Hamann et al. \(2020, p. 1405\)](#) recommend complementing the SRT with the LITMUS-QU-NWRT, given its small linguistic load and robustness against the influence of L2 experience.

1.3. The Present Study

The present study compares the performance of typically developing late-L2 refugee bilinguals (henceforth refugee-BiTD) and 2L1/eL2 heritage bilinguals (henceforth heritage-BiTD) to age-matched, typically developing monolingual children (MoTD), and to a control group of children with DLD, on three subtests of the BUEGA ([Esser et al. 2008](#)). The BUEGA test battery is of particular interest. It is often recommended in the guidelines on procedures for SN assessment in the field of speech and language education provided by the States (e.g., [Senatsverwaltung für Bildung, Jugend und Familie 2017](#), p. 9). The following research questions are addressed in this paper:

RQ1: Can the BUEGA subtests, assessing expressive grammar, word-reading, and spelling, be used as reliable measures for excluding language impairment in school-age heritage and bilingual refugee children with special attention to subtests assessing literacy? In particular, we ask whether adapting the monolingual cut-off points for performance in the pathology range according to the child’s language dominance (as recommended by [Thordardottir 2015](#)) helps enhance diagnostic accuracy and avoid misdiagnosis cases in bilingual children.

RQ2: Which age, input, and language environment variables influence performance on the BUEGA subtests?

RQ3: Does combining the BUEGA, particularly subtests assessing literacy skills, with newly developed LITMUS experimental tools (nonword and sentence repetition tasks), enhance diagnostic accuracy and help to avoid misdiagnosis?

RQ4: Can qualitative error analyses, especially in tasks assessing literacy skills, better discriminate BiTD children from children with DLD than mere quantitative measures and thus help to avoid misdiagnosis and provide recommendations for targeted special needs support?

2. Materials and Methods

2.1. Participants

This study presents data from Wave 2 of the longitudinal research project BiliSAT⁴ (Bilingual Language Development in School-age Children with/without Language Impairment with Arabic and Turkish as first languages). See [Hamann et al. \(2020\)](#) for results on Wave 1 and a detailed description of the refugee sample. For Wave 2, the refugee

participants had 12 additional months of exposure to the L2. It is worth noting that this sample contains an overlapping but not identical participant sample to Hamann et al. 2020, and the overlap only concerns the Arabic-speaking participant subset. In addition to the bilingual Arabic-German refugees and heritage children, this sample includes heritage bilinguals with Turkish as an L1, a monolingual control group, and a group of children with DLD. Importantly, this study expands upon previous research conducted within the BiliSAT project by investigating word-reading abilities alongside L2 oral abilities.

The current participant sample includes 45 monolingual and bilingual school-age children (ages 7;1–13;9) with and without DLD (see Table 1). The participants were divided into four groups: a control group of 12 monolingual, typically developing German-speaking children (MoTD, age range 7;3–12;1); a group of 7 children with DLD (4 Turkish, 2 German, 1 Arabic, age range 7;7–13;9); a group of 14 typically developing heritage bilinguals with either Turkish or Arabic as an L1 (heritage-BiTD, age range 7;1–13;4); and a group of 12 typically developing Syrian refugee bilinguals (refugee-BiTD, age range 8;7–13;1). All refugee-BiTDs were first-generation speakers of Syrian Arabic and were first exposed to German upon their arrival at an age of >6;0⁵, with an average of 46 months of exposure to German at the time of assessment. In the case of the refugee-BiTDs, the LoE corresponded to the length of schooling in the L2. Nine out of twelve had L1 schooling before arrival in Germany (1–3 years). Still, only 6 of them had literacy skills in Arabic due to interruptions in formal education either in Syria or during transit. Although the majority of the refugee children were initially assigned to age-appropriate grades at school, 2 of them had to repeat a grade at Wave 2, and another was downgraded in terms of school form, i.e., from *Realschule* to *Hauptschule*. The heritage children, on the other hand, had no literacy skills in their L1, started obtaining L2 literacy skills with school entry, and had a comparable length of L2 schooling to their refugee counterparts. Interestingly, the levels of maternal education in years (as a proxy for SES, see Hoff et al. 2002; Hoff 2003) was slightly higher in the refugee group than in the heritage group, with an average of 15 years vs. 11;6 years, respectively.

Table 1. Participant overview (mean, SD, range).

Variable	Refugee-BiTD	Heritage-BiTD	DLD	MoTD
Age (in years)	10.5 (1.9) 8.7–13.1	11.4 (1.6) 7.1–13.4	9.6 (2.2) 7.7–13.9	9.7 (1.6) 7.3–12.1
Age of Onset (in years)	6.8 (2.5) 5.5–10.0	2.3 (0.7) 1.0–3.5	2.2 (0.1) 0.0–4.9	-
Length of Exposure (in years)	3.1 (0.4) 2.3–3.6	9.2 (1.7) 4.5–11.7	7.5 (3.5) 3.3–13.7	-
Language Dominance Index	−25.58 (8.48) −36–(−10.50)	−3.42 (10.41) −21–15.50	3.64 (32.21) −26–50	50 (0.00) 50–50
Socioeconomic Status (years of maternal education)	15.00 (5.1) 8–22	11.6 (4.4) 4–17	10.6 (3.9) 4–16	16.7 (2.6) 13.5–20

Verification of clinical status as TD or DLD was conducted at Wave 1 via a comprehensive assessment procedure using standardized tests in the L1 and L2, considering dominance effects on test performance as recommended by Thordardottir (2015); see Hamann and Abed Ibrahim (2017); Hamann et al. (2020) for particulars. The following standardized tests were employed: L1-Arabic (ELO-L, Zebib et al. 2017) and for L2 (WWT, Glück 2011; LiSe-DaZ, Schulz and Tracy 2011; TROG-D, Fox 2009 and PLAKSS, Fox-Boyer

2014). In the case of the refugee group, clinical classification was only based on their L1 performance given their limited L2 proficiency.

2.2. Assessment Tool and Procedures

2.2.1. BUEGA Subtests

For assessing German abilities, we used parts of the standardized BUEGA test battery (Esser et al. 2008). The aim of the tool is to diagnose performance in different areas of language development and differentiate children with weak global performance (potential cases of DLD) from those with partial language deficits. The BUEGA covers the age range for most of our participants as it is normed for children in classes one to five; however, the test manual only provides norms for monolingual children.

Our investigation is centered around the subtests of grammar, word-reading, and spelling. Grammar targets morphological awareness and is normed for ages 6;0–11;5. The test includes 57 items targeting aspects of nominal morphology (plural formation, (ir)regular comparative and superlative forms of adjectives, and the production of accusative, dative, and genitive case forms) and verbal morphology (formation of regular and irregular past tense forms (both preterit and participial forms)). We used the oldest norms available for our older children.

Word-reading abilities are assessed with two lists of 56 words in total. The first list contains 32 short and commonly used nouns, verbs, articles, and pronouns (e.g., *in*, *never*, *know*). The second list consists of 24 words of greater length and complexity (e.g., *glittering stone*, *observe*, *conjure*).

The subtest spelling investigates children's abilities to correctly produce real written words by dictating 16–18 words with increasing complexity. Four lists are available depending on the child's grade. Test evaluation can be executed by counting the correctly written whole words or graphemes. In this study, we chose evaluation based on graphemes, as recommended in the manual, because it allows a more precise investigation of progress in spelling abilities. The test provides fixed specifications on how many letters in a word are counted as a grapheme for each of the four lists, setting a maximum number of potential wrong graphemes in one word. As in subtest reading, norms are available from grades one to five.

2.2.2. LITMUS Repetition Tasks

The German LITMUS-SRT

The German LITMUS-SRT used in this study, first introduced by Hamann et al. (2013) within COST Action IS0408, was constructed according to the LITMUS principles (Marinis and Armon-Lotem 2015). It thus contains complex structures known to be difficult for children with DLD cross-linguistically, e.g., relative clauses, finite complement clauses, and passive constructions, in addition to structures that represent crucial milestones in the acquisition of the properties of German morphosyntaxes, such as topicalization and the sentence bracket. The new-version of the German LITMUS-SRT investigated in this study was devised for use with older children and contains 60 sentences with three levels of increasing complexity controlling for number of syllables within each level (five conditions per level, four items per condition). Stimuli are presented in a pseudorandomized order via a child-friendly PPT. The task takes about 10 min to administer. The test sentences can be rated by "identical repetition", only disregarding phonological errors, or they can be scored by "target structure", which aims at ascertaining mastery of a structure compensating for typical L2 errors such as lexical substitutions and systematic, recurrent case errors, as well as gender errors, as long as they do not affect the realization of the targeted structure.

German LITMUS-QU-NWRT

The quasi-universal NWRT used in this study (Grimm et al. 2014) relies on increasing phonological complexity, not increasing the numbers of syllables (Grimm and Hübner in press). It consists of 66 one-, two-, and three-syllable nonwords built using vowels and

consonants attested in most of the world's languages. At the same time, it targets aspects of phonological complexity shown to be challenging for children with DLD (cf. [Ferré et al. 2012](#)), i.e., complex onsets and codas; see [Abed Ibrahim and Fekete \(2019\)](#) as well as [Schulz and Grimm \(2020\)](#) for a detailed description of the task's properties. The stimuli are presented via an appealing PPT in a pseudorandomized order through headphones. Task administration takes about 5–10 min, and the task is scored according to whole-item accuracy. In order to not disadvantage bilingual children, minimally different vowels, e.g., /o/ vs. /u/, and errors pertaining to the voicing of consonants, e.g., /b/ instead of /p/ are disregarded.

The Parental Questionnaire PaBiQ

Background information on the participants was collected using the Questionnaire for Parents of Bilingual Children (PaBiQ, [Tuller 2015](#)), which was augmented with questions about the refugee situation, such as L1 schooling, access to language courses, transit itineraries, means of transportation, as well as past and present living conditions. Relevant variables for this study were selected based on previous research and include chronological age, SES (operationalized as years of maternal education), AoO, LoE, current L2 use (the relative amount of L2 use within the immediate family), and the richness of the L2 environment. Moreover, the PaBiQ also allows for the calculation of an experiential index for language dominance based on early and current exposure patterns of the L2 relative to the L1; see [Abed Ibrahim and Fekete \(2019\)](#) for particulars.

2.3. Data Analysis

All standardized tests, the German LITMUS-SRT, and the LITMUS-QU-NWRT were administered according to test instructions (cf. [Hamann et al. 2020](#), 1384 for particulars). Children's oral responses were recorded using special recorders. Data transcription, verification, and coding were carried out offline by two independent linguistically trained raters. As for the two repetition tasks, the percentage of correct responses was calculated for each repetition measure. In the case of the LITMUS-SRT, this study only considers the scoring measure "correct target structure" since it has been shown to be a fairer measure for the assessment of bilinguals as opposed to "identical repetition" (cf. [Hamann and Abed Ibrahim 2017](#)).

Statistical analyses were conducted using [IBM SPSS 27 \(2020\)](#). Nonparametric tests were used for group comparisons, due to unequal and small sample sizes and the violation of normality assumptions, checked by the Shapiro–Wilk test. To explore which age and input variables predicted performance on the BUEGA subtests, we first ran a Spearman nonparametric correlational analysis between BUEGA subtests and relevant age/input variables (chronological age, AoO, LoE, SES, current L2 use, and L2 richness), then followed that with a hierarchical regression analysis using only variables yielding significant strong correlations with the BUEGA subtests as potential predictors. In order to examine the diagnostic potential of the BUEGA, we first calculated the diagnostic accuracy following the test manual, which only offers monolingual norms. In this case, a child is considered at risk of impairment in a particular domain, i.e., grammar, reading, or spelling, if she scored below a t-value of 35 ($-1.5 SD$) on the respective subtest. In the second step, we wanted to see whether following [Thordardottir's \(2015\)](#) recommendations by adapting cut-off scores to the degree of the child's language dominance in the L2, as estimated by the PaBiQ, would help avoid cases of misdiagnosis. Depending on whether the L2 was the child's dominant or weaker language, test results were interpreted differently. Children tested in their non-dominant language were allowed to score up to $-2.25 SD$ below the group mean before they were diagnosed with DLD. In the case of balanced and L2-dominant bilinguals, the cut-off was set at $-1.75 SD$ and $-1.5 SD$, respectively. In line with previous research recommending the use of combinations of tools for the assessment of bilinguals (e.g., [Chilla et al. 2021](#); [Tuller et al. 2018](#)), we explored whether combining subtests of the BUEGA with the German LITMUS-SRT and/or the LITMUS-QU-NWRT would enhance diagnostic

accuracy, especially in the case of the refugee group. Lastly, a qualitative error analysis was carried out on the BUEGA subtests to investigate whether heritage- and refugee-BiTDs show error patterns distinct from children with DLD.

3. Results

3.1. Overall Results on the BUEGA Subtests

Kruskal–Wallis tests comparing standardized group performance scores (z-scores) of refugee-BiTDs, heritage-BiTDs, MoTDs, and DLDs yielded significant results for grammar ($(\chi^2(3), N = 44) = 24.2, p = 0.000$), reading accuracy ($(\chi^2(3), N = 43) = 17.5, p = 0.001$), and spelling ($(\chi^2(3), N = 43) = 10.0, p = 0.018$), but not for reading pace ($(\chi^2(3), N = 43) = 75.2, p = 0.057$). Subsequent pairwise comparisons using Mann–Whitney U tests (see Appendix B and Figure 1) showed significant differences between MoTDs and all groups on the subtests of grammar and reading accuracy. Whereas heritage-BiTDs outperformed DLDs on the subtest of grammar, no significant differences were found between refugee-BiTDs and DLDs, or between refugee- and heritage-BiTDs. Interestingly, neither heritage-BiTDs nor refugee-BiTDs differed from DLDs on the subtest assessing reading accuracy. Concerning spelling, MoTDs performed significantly better than the DLD group. No significant differences emerged between MoTDs and heritage-BiTDs, and only a marginally significant difference emerged between MoTDs and refugee-BiTDs. While heritage-BiTDs performed significantly better than DLDs, refugee-BiTDs appeared to perform as poorly as DLDs. At the same time, no significant differences emerged between heritage and refugee-BiTDs; see Appendix C for an overview of group means and SDs on the individual subtests.

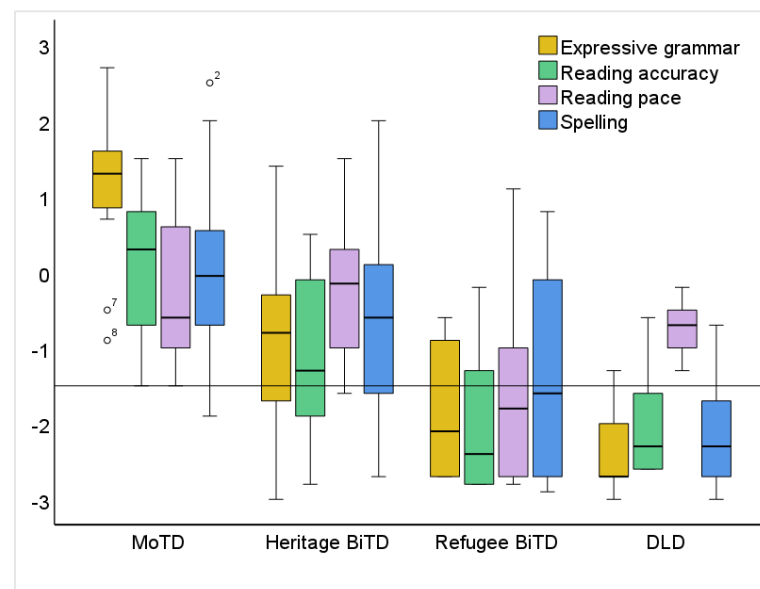


Figure 1. Between-group comparisons on BUEGA subtests (standardized z-scores) with cut-off point specified by the BUEGA manual (-1.5 SD).

3.2. Predictors of Performance on the BUEGA Subtests

To determine factors predicting performance in the BiTD group (refugees and heritage collapsed), we first ran nonparametric Spearman correlations between the BUEGA subtests (grammar, reading, and spelling) and age and input variables known to influence performance in language measures. Concerning grammar, strong positive correlations emerged with current L2 use ($r = 0.746, p = 0.000$), L2 richness ($r = 0.695, p = 0.000$), and SES ($r = 0.693, p = 0.000$), whereas no significant correlations were found between age, AoO, or LoE and grammar. Reading pace did not correlate with any age/input variables, while reading accuracy was significantly correlated with L2 richness ($r = 0.533, p = 0.007$) and

current L2 use ($r = 0.461$, $p = 0.023$). Unlike grammar and reading accuracy, no significant correlations emerged between the subtest of spelling and any of the age or input factors.

Next, hierarchical regression modeling was conducted to investigate which age and input factors explained the variance in the performance of the BiTDs on the BUEGA subtests. Since spelling was not correlated with any of the age or input variables, regression analyses were only done for the subtests of reading accuracy and grammar. Modeling was performed using the z-scores of each subtest as the dependent variable and background variables as independent variables. Only age and input variables yielding significant correlations were considered for the regression analyses. The latter were current L2 use, L2 richness, and SES (only for grammar), and were entered into the respective models in order of the strength of their correlation with the respective BUEGA subtests. As shown in Table 2, the primary predictor for performance on grammar was current L2 use, accounting for 57.4% of the variance in step 1. The addition of L2 richness in step 2 did not explain any additional variance, whereas adding SES in step 3 explained a further 11.0% of the variance. As for the subtest of reading accuracy, L2 richness emerged as a single significant predictor; however, it only explained 27.4% of the variance, suggesting that factors other than exposure variables are likely to be involved. The addition of current L2 use to the model at step 2 did not show any further significant contribution.

Table 2. Summary of hierarchical regression analyses for age/input variables predicting performance on BUEGA subtests grammar and reading accuracy.

Subtest	Step		<i>b</i>	<i>SE b</i>	β
Grammar ¹	1	Constant	2.417	−0.295	
		Current L2 use	0.251	0.045	0.758
	2	Constant	−3.072	0.563	
		Current L2 use	0.175	0.071	0.530
		L2 richness	0.139	0.102	0.291
	3	Constant	−3.633	0.526	
		Current L2 use	0.162	0.062	0.492
		L2 richness	0.062	0.093	0.130
		SES	0.094	0.033	0.384
Reading accuracy ²	1	Constant	−3.150	0.636	
		L2 richness	0.234	0.081	0.523
	2	Constant	−3.073	0.706	
		L2 richness	0.208	0.124	0.466
		Current L2 use	0.022	0.080	0.077

¹ $R^2 = 0.574$ for Step 1; $\Delta R^2 = 0.033$ for Step 2 ($p = 0.188$); $\Delta R^2 = 0.110$ for Step 3 ($p = 0.009$). ² $R^2 = 0.274$ for Step 1; $\Delta R^2 = 0.003$ for Step 2 ($p = 0.784$).

3.3. Diagnostic Accuracy of Single Measures and Combinations Thereof

Next, we calculated the diagnostic accuracy of the BUEGA subtests, examining different diagnostic cut-off criteria, and examined whether combining BUEGA subtests with recently developed LITMUS repetition tasks enhances diagnostic accuracy and helps avoid cases of misdiagnosis. We first used monolingual cut-off scores in accordance with the test manual, i.e., a child was viewed as having language disorder if she scored -1.5 SD below the respective group mean. Subsequently, we recalculated the diagnostic accuracy of the BUEGA subtests, applying cut-off scores adjusted to the child's language dominance as estimated by the PaBiQ. Following Thordardottir (2015), a cut-off score of -2.25 SD was used in the case of L1-dominant children (e.g., all refugee-BiTDs). For balanced and L2-dominant bilinguals, the cut-off scores were set at -1.75 SD and -1.5 SD, respectively. As can be seen in Table 3, applying monolingual cut-offs results in a significant proportion of overidentification as language impaired, not only in the refugee group but also in the heritage-BiTD group on all three BUEGA subtests. While applying dominance-adjusted cut-offs results in a substantial improvement in diagnostic accuracy in the heritage-BiTD

group, especially for the subtests of grammar (only one instead of 5/14 misdiagnosed) and reading accuracy (only two instead of 6/14 misdiagnosed), only a slight improvement is observed for the refugee group, especially for the subtest of reading accuracy, where more than a third of the sample would still be overidentified as having a language disorder.

Table 3. Proportion of BiTD children identified as language impaired using single (without and with dominance adjustment) and combined measures (total Ns and percent).

Group	Evaluation Procedure	BUEGA Subtest					
		Grammar		Reading Accuracy		Spelling	
		N	%	N	%	N	%
Heritage-BiTDs	BUEGA monolingual cut-off	5/14	36	6/14	43	4/12	29
	BUEGA LD ¹ adjusted cut-off	1/14	7	2/14	14	3/12	21
	BUEGA combined with SR	1/14	7	1/14	7	1/12	7
	BUEGA combined with NWRT	0/14	0	0/14	0	0/12	0
Refugee-BiTDs	BUEGA monolingual cut-off	6/12	50	6/12	50	6/12	50
	BUEGA LD adjusted cut-off	4/12	33	5/12	42	4/12	33
	BUEGA combined with SRT	3/12	25	3/12	25	3/12	25
	BUEGA combined with NWRT	0/12	0	0/12	0	0/12	0

¹ LD: language dominance.

As a next step, we explored whether combining BUEGA subtests with LITMUS-SRT and LITMUS-QU-NWRT would result in better diagnostic accuracy. Prior to this step, we wanted to verify whether refugee-BiTDs perform on par with their heritage peers and whether they significantly differ from the DLD group. Significant group effects were found for both tasks: LITMUS-SRT ($\chi^2(3, N = 45) = 12.649, p = 0.005$), and LITMUS-NWRT ($\chi^2(3, N = 45) = 15.787, p = 0.001$). Mann–Whitney U comparisons revealed significant differences between heritage- and refugee-BiTDs and DLDs for the LITMUS-SRT (refugee-BiTD vs. DLD: $U = 16.5, p = 0.036, r = 0.495$ and heritage-BiTD vs. DLD: $U = 16.0, p = 0.012, r = 0.538$) as well as the LITMUS-NWRT (refugee-BiTD vs. DLD: $U = 2.00, p = 0.000, r = 0.766$ and heritage-BiTD vs. DLD: $U = 5.50, p = 0.000, r = 0.709$) with medium to high effect sizes, while no significant differences emerged between heritage- and refugee-BiTDs, see Figure 2. Although using LITMUS-SRT increases diagnostic accuracy on almost all BUEGA subtests in both groups, 3/12 refugee children and one heritage-BiTD child were still overidentified as having a language impairment relative to the cut-off score established by heritage populations in previous research by the task authors (cf. Hamann and Abed Ibrahim 2017). Finally, we combined the latter measures with LITMUS-QU-NWRT, which is known to be relatively robust against the effects of SES on limited L2 exposure. Indeed, once NWRT was considered, there were no cases of overidentification in either group.

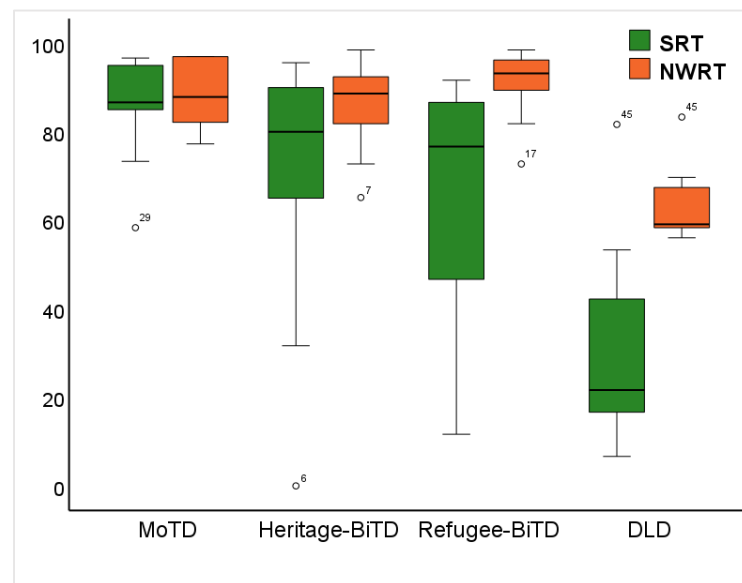


Figure 2. Group performance on LITMUS-SRT (scored by target structure) and NWRT with cut-off points according to Hamann and Abed Ibrahim (2017).

3.4. Qualitative Error Analyses

In the last step, we carried out qualitative analyses for the error categories enlisted in the BUEGA test manual for each of the three subtests. Since the primary goal of the qualitative error analysis was to investigate whether the individual error types can distinguish between diverse groups of typically developing bilingual children and children with DLD, we only focused on contrasts between the latter groups. As for grammar, Kruskal–Wallis tests for the distribution of errors between groups revealed significant group effects for all of the error categories provided by the manual (excluding passive)⁶; see Table 4. The latter include case (accusative, dative, genitive), past tense, comparative and superlative forms of adjectives, and plural forms. As demonstrated in Table 4, significant differences emerged between heritage-BiTDs and DLDs on all categories, whereas refugee-BiTDs scored as poorly as the DLDs on all categories.

Table 4. Group effects (Kruskal–Wallis tests) and pairwise comparisons (Mann–Whitney U tests) between groups and grammatical categories of the subtest of grammar. Significant values are given in bold.

	Case	Past Tense	Comparative	Superlative	Plural
	$\chi^2 = (3, n = 44) = 18.0, p = 0.000$	$\chi^2 = (3, n = 44) = 18.0, p = 0.000$	$\chi^2 = (3, n = 44) = 19.3, p = 0.000$	$\chi^2 = (3, n = 44) = 15.3, p = 0.002$	$\chi^2 = (3, n = 44) = 27.8, p = 0.000$
Refugee-BiTD vs. DLD	$U = 21.0, p = \text{n.s.}, r = 0.397$	$U = 18.0, p = \text{n.s.}, r = 0.450$	$U = 19.0, p = \text{n.s.}, r = 0.420$	$U = 16.0, p = \text{n.s.}, r = 0.485$	$U = 24.5, p = \text{n.s.}, r = 0.302$
Heritage-BiTD vs. DLD	$U = 11.5, p = 0.009, r = 0.631$	$U = 15.0, p = 0.030, r = 0.614$	$U = 12.5, p = 0.012, r = 0.610$	$U = 10.5, p = 0.006, r = 0.633$	$U = 11.50, p = 0.009, r = 0.618$

As for the subtest of word-reading accuracy, BUEGA specifies the following error types: (1) the omission of sounds, (2) the addition of sounds, (3) the substitution of sounds, and (4) unrecognizable words (Esser et al. 2008, p. 88). Qualitative analyses revealed significant differences with high effect sizes between both BiTD groups and their DLD peers for the categories “addition of sounds” (refugee-BiTD vs. DLD: $U = 9.5, p = 0.010$,

$r = 0.629$; heritage-BiTD vs. DLD: $U = 20.0$, $p = 0.031$, $r = 0.514$) and “unrecognizable words” (refugee-BiTD vs. DLD: $U = 6.0$, $p = 0.003$, $r = 0.769$; heritage-BiTD vs. DLD: $U = 11.5$, $p = 0.003$, $r = 0.677$); see Figure 3. On the other hand, no significant differences emerged between the heritage-BiTDs and the refugee-BiTDs.

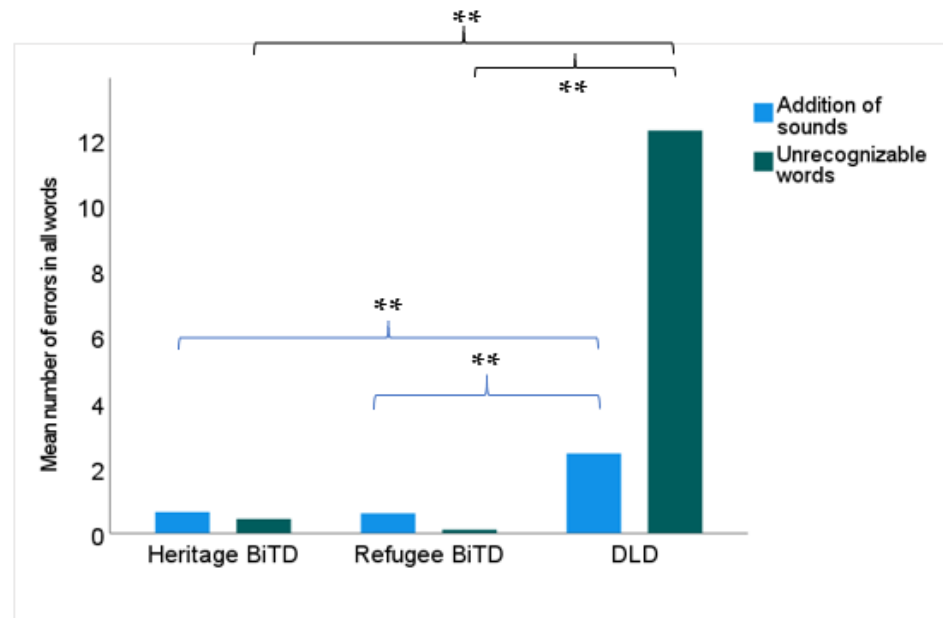


Figure 3. Qualitative analysis of subtest reading accuracy. ** Significance level < 0.01 .

The subtest of spelling provides five error types for qualitative analysis: (1) incorrect graphemes, (2) unrecognizable graphemes, (3) missing graphemes, (4) missing dots in umlauts (e.g., $\langle \ddot{u} \rangle \rightarrow \langle u \rangle$), and (5) upper and lower case, for children in the second term of grade two (Esser et al. 2008, p. 90). As shown in Table 5, qualitative analyses revealed no significant differences between refugee-BiTDs and DLDs for any of the aforementioned error types. Comparisons between heritage-BiTDs and DLDs yielded only one significant difference, namely for missing graphemes ($U = 8.50$, $p = 0.010$, $r = 0.569$), and no significant differences emerged between heritage- and refugee-BiTDs.

Table 5. Group comparisons for qualitative analysis of the subtest of spelling (Kruskal–Wallis tests and Mann–Whitney U comparisons).

	Incorrect Phonemes	Unrecognizable Phonemes	Missing Letters	Missing Dots in Umlaut	Lower and Upper Case
	$\chi^2 = (3, n = 42)$ $= 8.28$, $p = 0.040$	$\chi^2 = (3, n = 42)$ $= 4.66$, $p = 0.198$	$\chi^2 = (3, n = 40)$ $= 7.89$, $p = 0.048$	$\chi^2 = (3, n = 40)$ $= 5.00$, $p = 0.172$	$\chi^2 = (3, n = 40)$ $= 9.36$, $p = 0.025$
Refugee-BiTD vs. DLD	$U = 25.5$ $p = 0.646$ $r = 0.115$	$U = 29.5$ $p = 0.959$ $r = 0.019$	$U = 19.0$ $p = 0.513$ $r = 0.191$	$U = 18.0$ $p = 0.521$ $r = 0.200$	$U = 20.0$ $p = 0.684$ $r = 0.122$
Heritage-BiTD vs. DLD	$U = 17.0$ $p = 0.107$ $r = 0.383$	$U = 28.0$ $p = 0.559$ $r = 0.383$	$U = 8.5$ $p = 0.010$ $r = 0.569$	$U = 22.0$ $p = 0.703$ $r = 0.130$	$U = 19.0$ $p = 0.477$ $r = 0.194$

4. Discussion

Considering the far-reaching consequences of diagnosis in the German education system for determining further school and professional paths, standardized language

assessment tools for linguistically diverse groups of students are noticeably lacking. Hence, the main purpose of this study was to establish an adequate means of assessment and evaluation of language abilities in heterogeneous populations that are simultaneously able to exclude language disorders. Most importantly, we looked beyond oral language competencies and took L2 word-reading abilities into consideration as a possible resource for older L2 learners, as in the case of our refugee group. For this purpose, we examined the diagnostic potential of the standardized test battery BUEGA, which is a frequently recommended instrument in SN guidelines provided by the Federal State's Ministries of Education for the assessment of school performance in grades one to five. Furthermore, we wanted to explore whether qualitative error analyses could help disentangle DLD from typical language development in late successive bilinguals with limited exposure to the L2.

RQ1 asked whether the BUEGA subtests assessing grammar, word-reading, and spelling can be used as reliable measures for excluding language impairment in school-age heritage and bilingual refugee children, with a special focus on subtests assessing literacy. Our findings align with former research showing that standardized procedures designed for and based on monolingual populations are not automatically transferable to bilinguals and should be used with caution in bilingual contexts. This is evident in the high number of children overdiagnosed with DLD when the monolingual cut-off score ($-1.5 SD$) is applied based on the test's manual. Up to 50% of typically developing bilinguals across groups and subtests scored in the pathology range. This finding is particularly alarming since not only recently immigrated late-L2 bilingual children appear to be at risk of being misidentified as having DLD, but also simultaneous and early successive bilinguals who were born in Germany and have been exposed to the L2 for a significantly longer period of time. Following the recommendations of [Thordardottir \(2015\)](#) for norm adjustments according to the participants' language dominance, we found that only heritage-BiTDs slightly benefited from the procedure on measures of expressive grammar and reading accuracy, while no meaningful changes were observed for the IL2 group on either BUEGA measure. At the same time, several heritage bilingual children remained overdiagnosed across subtests, especially in the subtest assessing spelling. In the case of refugee-BiTDs, the small beneficial effect is practically absent. Only two children were classified as TD based on a cut-off considering their language dominance, whereas most children remained in the pathology range. However, this is not surprising given that Thordardottir's recommendations were based on simultaneous bilinguals and are unlikely to be suited for late successive child bilinguals. See [Hamann et al. \(2020\)](#) for similar findings on performance on L2 standardized assessment tools assessing vocabulary and morphosyntax.

In addition to their variable AoO to the L2, bilingual children experience significantly more sources of variation in their language environment than their monolingual peers, such as the amount of L2 (and L1) exposure/use inside and outside the home and linguistic richness. The latter is further affected by distal input factors, especially by SES (measured by years of maternal education), which plays a crucial role in modulating qualitative and quantitative aspects of language input ([Paradis and Jia 2017](#)). Hence, RQ2 asked which age, input, and language environment variables accounted for the variance in performance on the BUEGA subtests assessing expressive grammar and literacy skills. Since the subtest of spelling did not yield any significant correlations with any of the background variables, it was not considered for subsequent regression analyses.

Our results showed that a significant proportion of variance on the subtest of grammar, which mainly targets morphosyntax, was predicted by current L2 use (57%), with SES adding an additional 11%. In contrast, neither the AoO nor LoE had significant contributions. L2 use, as estimated by the parental questionnaire (PaBiQ), includes communication scenarios inside the home with caregivers, siblings, or other relatives. This implies that children with greater L2 exposure at home develop better morphosyntactic abilities than those who communicate in their L1 at home more often. This echoes recent findings of the IGLU survey reporting a performance gap of half a standard deviation between children who rarely speak German at home and those who frequently use German, with higher

numbers if both parents are first-generation immigrants (Hußmann et al. 2017). Similarly, in a study by Hamann et al. (2020), a combination of current L2 use and SES explained the sheer amount of performance variance on the German LITMUS-SRT, which also targets morphosyntactic abilities. Our results are also in line with Unsworth (2016), where current L2 use was the sole predictor for the performance of English-Dutch bilinguals with an AoO of four to seven years on measures assessing morphosyntax, vocabulary, and syntax-semantics, while the LoE did not show any significant contribution.

Moreover, SES adding to the variance corroborates the findings of a large body of research on language development. Children from households with higher SES are likely to experience quantitatively and qualitatively more enriched language input resulting in larger vocabulary size and better command of morphologically complex structures (Fernald et al. 2013; Hoff 2003; Hoff et al. 2002). Furthermore, a higher level of education often coincides with higher L2-proficiency levels, which modulates the quantity and quality of the L2 input to the child (e.g., Prevoo et al. 2014; Duncan and Paradis 2020).

Reading accuracy, on the other hand, appeared to be more robust against the influence of language environment factors, given that only L2 richness accounted for 27% of the variance. At the same time, the AoO, LoE, and current L2 use did not explain any of the variances. Language richness, according to the PaBiQ, gives an estimate of L2 input through extracurricular activities such as reading, watching movies, socializing with friends, or interactions with teachers or school peers. The richness of L2 input and the diversity of interaction contexts, especially exchanges with native speakers, thus seem to be of greater relevance for developing L2 reading skills than L2 use within the immediate home environment. Former studies found similar effects of L2 richness on L2 vocabulary, morphology, syntax, and narrative skills (Jia and Fuse 2007; Paradis and Jia 2017). For example, Jia and Fuse (2007) found that the acquisition of English grammatical morphology by native Mandarin speakers was predicted by language richness but not by LoE after five years of residency. Moreover, similar to their monolingual peers, bilingual children start to acquire reading skills upon school entry, which marks the crucial shift towards a qualitatively enhanced (academic) L2 input. Hence, it is no surprise that the quantity of L2 input at home was not a significant predictor of performance on word-reading. However, since 70% of variance remained unaccounted for, other factors must be considered for predicting reading abilities, e.g., L1 richness, morphological awareness, and working memory capacities.

Even though L1 richness was not investigated as a predictor in this study, multiple findings assume that L1 richness can positively contribute to L2 development in older learners who prefer to use the L1 over L2 at home and exhibit better L1 abilities (Pham and Tipton 2018). Research conducted with native speakers of English has shown that morphological awareness is significantly associated with various aspects of literacy skills, including word-reading (Carlisle 2020; Deacon et al. 2013; Kirby et al. 2012; Nagy et al. 2006). Studies also reported that children in elementary grades differ significantly in their ability to manipulate morphologically complex words, and these variances reflect children's differences. Hence, L2 learners with higher proficiency in their L1 are likely to profit from their previously acquired L1 knowledge and metalinguistic phonological and morphological awareness.

RQ3 examined whether combining BUEGA subtests with recently developed LITMUS repetition tasks could enhance diagnostic accuracy and help avoid cases of misdiagnosis. This method proved to be very promising since applying the LITMUS-NWRT, which is less reliant on previous language knowledge, eliminated all cases of misdiagnoses with DLD in both heritage-BiTDs and, more importantly, in IL2 refugee-BiTDs with limited L2 proficiency. This finding resonates with previous studies demonstrating the robustness of this tool against exposure variables, given its small linguistic load (i.e., Hamann and Abed Ibrahim 2017; Abed Ibrahim and Fekete 2019). However, it must be kept in mind that selective impairments do exist (Friedmann and Novogrodsky 2008) and that some

children have impairments in domains other than phonology. Thus, the task should be complemented with other tasks.

Since none of the BUEGA subtests distinguished between refugee-BiTDs and DLDs when global scores were considered, and even heritage-BiTDs were misdiagnosed using the global performance of the word-reading accuracy subtest, in RQ4 we wanted to explore whether employing qualitative error analyses could help differentiate between BiTDs and children with DLD. Unlike subtests assessing expressive grammar and spelling, qualitative analyses for the subtest “reading accuracy” yielded significant differences with high effect sizes between refugee-BiTDs and DLDs on the error types phoneme addition and unrecognizable words. This is an encouraging result since the latter error categories can be relatively quickly assessed by schoolteachers. The last error type is of particular interest since it might reflect weaker word decoding abilities in children with DLD. These deficits are likely to be associated with deficits in phonological and morphological representations that are assumed to be intact in typically developing bilinguals, even in cases of limited oral L2 proficiency (see [Gottardo et al. 2020](#)).

5. Conclusions and Implications for Language Assessment in Bilingual Contexts

Our study provides important insights into language and SN assessment in multilingual populations in Germany. In line with previous research, our results demonstrated that standardized language tests normed for monolingual populations, such as the BUEGA ([Esser et al. 2008](#)), should be used with caution with bilingual children, especially in the case of IL2 learners. Adjusting monolingual cut-off points for language pathology according to the child’s language dominance helped to avoid misdiagnosis cases in simultaneous and early successive bilinguals. However, it did not help to avoid overdiagnosis with DLD in late successive bilingual children. This stresses the need for alternative assessment procedures that go beyond quantitative analyses and focus on potential areas of strength in IL2 learners, such as word-reading abilities. Indeed, supplementing quantitative analyses with a qualitative evaluation of reading performance proved to be promising. One of the error types that stood out in the DLD group was the production of unrecognizable words, which can serve as a quickly identifiable marker of impairment. We have also shown that in order to avoid cases of misdiagnosis, established assessment procedures for formative and status diagnostics should possibly be combined with LITMUS tools for the assessment of language impairment, to cover both the nature and the nurture of underachievement in school-related language and the teachers’ demand for sensitive and easy-to-administer tools.

We further demonstrated that bilingual children’s performance on linguistic tasks is influenced by variations in the quantity and quality of language input within and outside the home environment. Hence, it is essential that language assessment procedures include parental questionnaires, which enable gathering relevant background information that serves as a basis for interpreting performance on language tasks. Employing parental surveys would also enable collecting specific information about the refugee population, such as interrupted schooling and trauma, which contribute to the socio-emotional wellbeing of the latter group. In addition, questionnaires can provide a subjective estimate of the child’s development in her L1, which is of great relevance for IL2 children with limited exposure to the L2, especially in cases where L1 formative assessment is not possible. Despite the importance of parental questionnaires, they are rarely employed in pedagogical practice due to time and resources limitations. Thus, it is essential to raise educators’ and clinical practitioners’ awareness of the existence of empirically researched questionnaires, which are not only accessible but also fast and easy to administer. Prominent examples that are being used already are the PaBiQ ([Tuller 2015](#)), which exists in different languages, and the questionnaire developed by the [BiSS-Trägerkonsortium \(2020\)](#) for use by educators working with newcomers at schools and daycare facilities in Germany.

Our results have further implications with regard to academic language skills. Literacy is central to academic success, as written products still serve as the primary means of

assessment and grading in school. They are, hence, both the prerequisite for and birthplace of school achievement and failure. Most oral language support interventions compensating for linguistic deficits have been shown to be moderately effective at best (see [Kempert et al. 2016](#)). The finding that word-reading abilities appear to be less sensitive to effects of the amount of exposure than expressive grammar should be taken as a reason to encourage considering word-reading abilities as an early indicator of typical language development in IL2 children. Hence, concepts with regard to literacy transfer should receive significantly more attention. As for policymakers within the educational sector, higher priority should be assigned to evidence-based assessment in L2 populations in future teacher training. The Federal States need to expand the range of qualifications available in their studies, traineeships, and further education. It is also necessary to ensure that educational institutions employ sufficiently qualified staff and provide professional support. These include, in particular, interdisciplinary teams of specialists and special pedagogues (trained for SN language).

A limitation of this study is that the examination of sources of individual variation in performance on the BUEGA subtests was confined to age and input factors. Given the age range of the participants in our sample, it is worth exploring in future studies whether cognitive abilities, e.g., working memory capacities, can account for differences in performance on linguistic tasks (e.g., [Paradis et al. 2017](#)). Moreover, future research should take a closer look at the interdependence between L1 and L2 abilities. A growing body of literature focusing on literacy skills and metalinguistic skills which support reading suggests a positive impact of the L1 on L2 performance ([Hammer et al. 2007](#); [Tabors et al. 2003](#)). Furthermore, when it comes to the vulnerable group of refugee children, recent research suggests that socio-emotional wellbeing factors are likely to affect linguistic performance and should thus be considered in future studies. For example, [Soto-Corominas et al. \(2020\)](#) found an association between language skills and socioemotional wellbeing in Syrian refugee children (see also [Chen et al. 2019](#)).

In sum, future research should pursue the goal of developing a fully integrative language assessment for linguistically diverse populations where significant impact factors are considered.

Author Contributions: Conceptualization, S.C. and I.H.; methodology, S.C. and L.A.I.; software, I.H.; validation, I.H., S.C. and L.A.I.; formal analysis, I.H.; investigation, I.H. and L.A.I.; resources, S.C.; data curation, L.A.I.; writing—original draft preparation, I.H., S.C. and L.A.I.; writing—review and editing, I.H., S.C. and L.A.I.; visualization, I.H.; supervision, S.C.; project administration, S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: The BiliSAT project was funded by DFG Grants to CH 1112/4-1S (to Chilla) and HA 2335/7-1C (to Hamann).

Institutional Review Board Statement: The present study was conducted in line with the compliance form, transaction number 20120416505890730506, of the German Science Foundation and the recommendation of the “Kommission für Forschungsfolgenabschätzung und Ethik” (commission for the evaluation of research consequences and ethics) of the Carl-von-Ossietzky University of Oldenburg (rf. Drs. 21/16/2013). Parents or legal guardians of all participating minors provided written informed consent for both data collection and analysis. The research protocol was approved by the “Kommission für Forschungsfolgenabschätzung und Ethik” of the Carl-von-Ossietzky University of Oldenburg.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank all the investigators involved in the project for their continued advice and support, with special thanks to Cornelia Hamann. Special thanks also go to Angela Grimm for sharing her German LITMUS NWRT. The authors thank all the parents, educators, teachers, and speech language therapists for their cooperation and, last but not least, the authors particularly thank the children for their participation and their patience with us in completing the tasks.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

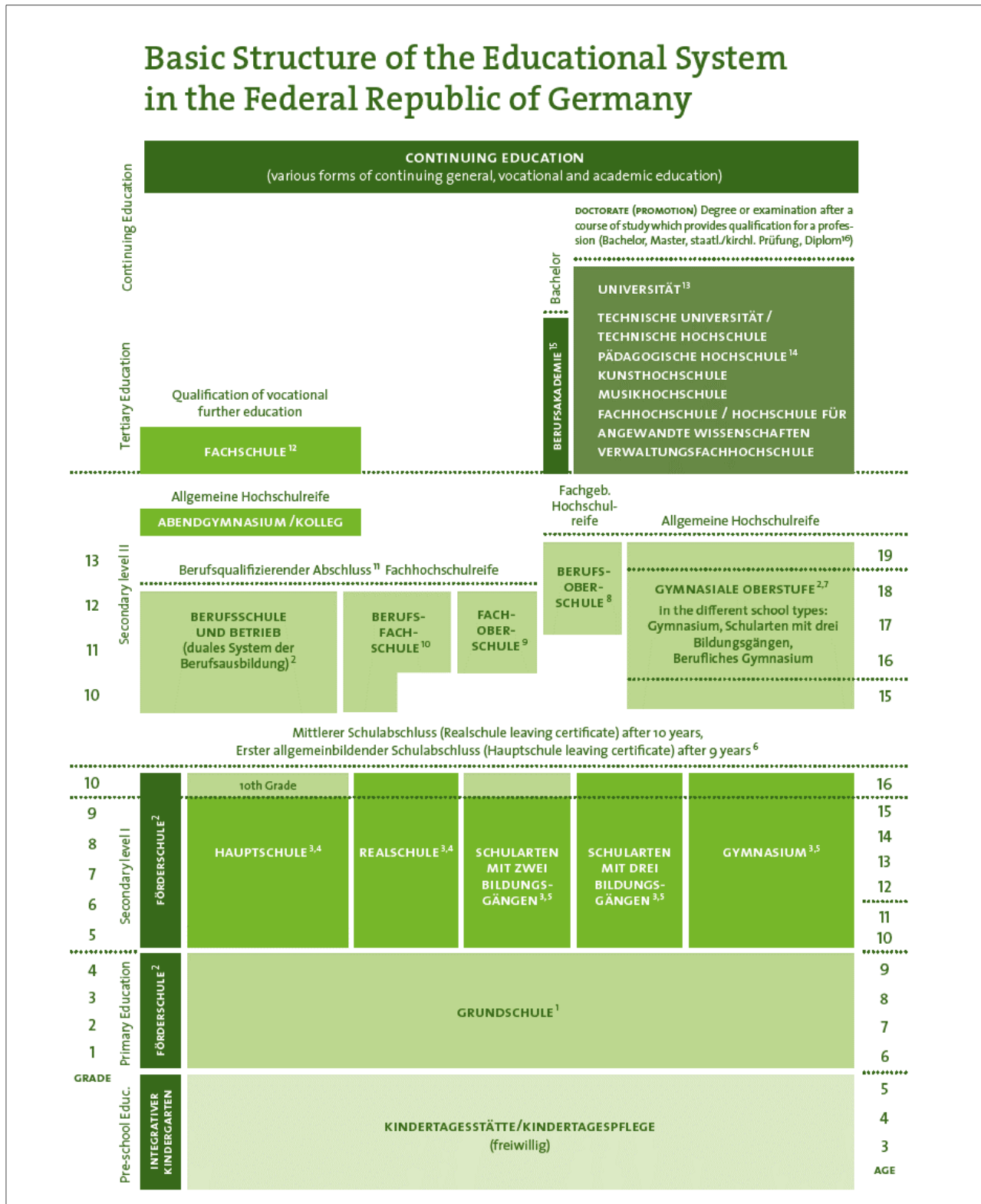


Figure A1. Basic structure of the German educational system.

Appendix B

Table A1. Group performances on BUEGA subtests in z-values (Mean, SD, range)¹.

Subtest	MoTD	Heritage-BiTD	Refugee-BiTD	DLD
Grammar	1.12	−0.80	−1.35	−2.27
	−1	−1.14	−1.13	−0.58
	−0.90–2.70	−3.00–1.40	−2.70–0.90	−3.00–(−1.30)
Reading pace	−0.27	−0.16	−1.49	−0.54
	−0.96	−0.92	−1.32	−0.51
	−1.50–1.50	−1.60–1.5	−2.80–1.10	−1.30–0.20
Reading accuracy	0.11	−1.07	−1.88	−1.80
	−0.92	−1.11	−0.94	−0.8
	−1.50–1.50	−2.8–0.50	−2.80–(−0.20)	−2.60–(−0.60)
Spelling	0.05	−0.64	−1.22	−2.08
	−1.26	−1.19	−1.33	−0.91
	−1.90–2.50	−2.7–2.00	−2.90–0.80	−3.00–(−0.70)

¹ Reported z values.

Appendix C

Table A2. Performance on BUEGA subtests: between-group comparisons (Mann–Whitney U tests), significant comparisons are given in bold.

Group	Grammar	Reading Accuracy	Spelling
Refugee-BiTD vs. heritage-BiTD	$U = 58.0$ $p = 0.317$ $r = 0.208$	$U = 41.0$ $p = 0.096$ $r = 0.348$	$U = 67.0$ $p = 0.403$ $r = 0.171$
Refugee-BiTD vs. DLD	$U = 23.0$ $p = 0.179$ $r = 0.280$	$U = 29.0$ $p = 0.601$ $r = 0.185$	$U = 17.0$ $p = 0.195$ $r = 0.284$
Heritage-BiTD vs. DLD	$U = 10.5$ $p = 0.002$ $r = 0.370$	$U = 28.5$ $p = 0.128$ $r = 0.270$	$U = 10.0$ $p = 0.019$ $r = 0.349$
MoTD vs. DLD	$U = 0.000$ $p = 0.000$ $r = 0.432$	$U = 5.0$ $p = 0.001$ $r = 0.405$	$U = 4.0$ $p = 0.004$ $r = 0.401$
Refugee-BiTD vs. MoTD	$U = 7.5$ $p = 0.000$ $r = 0.396$	$U = 8.5$ $p = 0.000$ $r = 0.393$	$U = 38.0$ $p = 0.052$ $r = 0.286$
Heritage-BiTD vs. MoTD	$U = 18.0$ $p = 0.000$ $r = 0.361$	$U = 32.5$ $p = 0.006$ $r = 0.319$	$U = 58.0$ $p = 0.193$ $r = 0.227$

Notes

- ¹ The procedures for determining SN can vary depending on the federal state and are coordinated by the respective Ministry of Education. SN in the field of “speech and language education” is formulated in the KMK recommendation (1998) only in general and non-binding terms. Thus, it states that it can include information on, e.g., (1) impairments in language ability; (2) language acquisition, use of language, and speaking ability; (3) course of development and acquisition of language and speech; (4) measures and results of previous assessments; (5) individual living and upbringing circumstances; (6) social integration and school environment; or (7) hearing, auditory and visual perception, and motor skills. In addition to special education teachers, doctors from different disciplines such as ear, nose, and throat medicine, phoniatrics, neurology, orthodontics, pediatrics, psychologists, and representatives of medical-therapeutic and social services may also be involved.

- 2 The German school system is characterized by the existence of different school types across the different federal states, e.g., Hauptschule, Realschule, Gesamtschule, Sekundarschule, and the Gymnasium, among others (for an overview of the German education system, see Appendix A).
- 3 These children are also referred to as heritage child bilinguals, i.e., children of first-generation or second-generation immigrants who acquire their native heritage language in early childhood (i.e., before the age of 5;0; cf. Schwartz 2004; Unsworth 2005) either successively or simultaneously with the L2 (Montrul 2016).
- 4 The BiliSAT project was funded by DFG Grants to CH 1112/4-1S (to Chilla) and HA 2335/7-1C (to Hamann).
- 5 Two of the Refugee-BiTDs already had an age of 10;0 upon their arrival in Germany.
- 6 Passive was excluded from the qualitative analyses because there is only a single item eliciting this condition.

References

- Abed Ibrahim, Lina, and István Fekete. 2019. What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in Psychology* 9: 27–57. [CrossRef] [PubMed]
- Abed Ibrahim, Lina, Cornelia Hamann, and István Fekete. 2020. Language Assessment of Bilingual Arabic-German Heritage and Refugee Children: Comparing Performance on LITMUS Repetition Tasks. In *Proceedings of BUCLD 44*. Edited by Megan M. Brown and Alexandra Kohut. Somerville: Cascadia Press, pp. 1–17.
- Ahrenholz, Bernt, Fränze S. Wagner, Annkathrin Darsow, Anke Börsel, Brigitte Jostes, and Jennifer Paetsch. 2016. DaZ und Sprachbildung in der Berliner Lehrkräftebildung. *DDS—Die deutsche Schule. Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis. Beiheft* 13: 23–34.
- Al-Janaideh, Redab, Alexandra Gottardo, Sana Tibi, Johanne Paradis, and Xi Chen. 2020. The role of word reading and oral language skills in reading comprehension in Syrian refugee children. *Applied Psycholinguistics* 41: 1283–304. [CrossRef]
- Andreou, Georgia, and Lemoni Garyfallia. 2020. Narrative Skills of Monolingual and Bilingual Pre-School and Primary School Children with Developmental Language Disorder (DLD): A Systematic Review. *Open Journal of Modern Linguistics* 10: 429–58. [CrossRef]
- Armon-Lotem, Sharon, and Natalia Meir. 2016. Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language and Communication Disorder* 51: 715–31. [CrossRef] [PubMed]
- Armon-Lotem, Sharon, and Natalia Meir. 2019. The nature of exposure and input in early bilingualism. In *The Cambridge Handbook of Bilingualism*. Edited by Annick De Houwer and Lourdes Ortega. Cambridge: Cambridge University Press, pp. 193–211.
- Armon-Lotem, Sharon, Jan de Jong, and Natalia Meir. 2015. *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Armon-Lotem, Sharon, Joel Walters, and Natalia Gagarina. 2011. The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism* 1: 291–317. [CrossRef]
- Autorengruppe Bildungsberichterstattung. 2020. *Bildung in Deutschland 2020. Ein Indikatorengestützter Bericht mit einer Analyse zu Bildung in einer Digitalisierten Welt*. Bielefeld: wbv Media.
- Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, ed. 2019. *Deutschland kann Integration*. Berlin: Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration.
- Bedore, Lisa M., and Elizabeth D. Peña. 2008. Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism* 11: 1–29. [CrossRef]
- Bertelsmann Stiftung, ed. 2015. *Inklusion in Deutschland. Daten und Fakten*. Prof. Dr. phil. Klaus Klemm im Auftrag der Bertelsmann Stiftung. Gütersloh: Bertelsmann Stiftung.
- Birdsong, David. 2018. Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology* 9: 81. [CrossRef] [PubMed]
- Bishop, Dorothy V. M., Margaret J. Snowling, Paul A. Thompson, Trisha Greenhalgh, and the CATALISE-2 consortium. 2017. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry* 58: 1068–80. [CrossRef]
- BiSS-Trägerkonsortium. 2020. *Leitfaden für den Erstkontakt. Sprachliche Bildung für neu zugewanderte Kinder und Jugendliche*. Bielefeld: wbv Media.
- Botting, Nicola. 2020. Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International Journal of Language & Communication Disorders* 55: 255–65.
- Bundeszentrale für Politische Bildung (BPB). 2021. *Datenreport 2021. Ein Sozialbericht für die Bundesrepublik Deutschland*. Edited by Destasis, WZB and BiB. Bonn: bpb, Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKewjHILm2j5XzAhV0hf0HHVqRDToQFnoECBUQAQ&url=https%3A%2F%2Fwww.wzb.eu%2Fsystem%2Ffiles%2Fdocs%2Fsv%2Fk%2Fdr2021_buch_fuer_web_210212_gesamt.pdf&usg=AOvVaw2c_zt4V4ORmi0a7YWLDjbz (accessed on 20 September 2021).
- Carlisle, Joanne F. 2020. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing* 12: 169–90. [CrossRef]

- Catts, Hugh W., Marc E. Fey, Susan Ellis Weismer, and Sittner Bridges Mindy. 2014. The Relationship Between Language and Reading Abilities. In *Understanding Individual Differences in Language Development Across the School Years*, 1st ed. Edited by Bruce J. Tomblin and Marilyn A. Nippold. New York: Psychology Press, pp. 158–79.
- Chen, Alexandra, Catherine Panter-Brick, Kristin Hadfield, Rana Dajani, Amar Hamoudi, and Margaret Sheridan. 2019. Minds Under Siege: Cognitive Signatures of Poverty and Trauma in Refugee and Non-Refugee Adolescents. *Child Development* 90: 1856–65. [\[CrossRef\]](#)
- Chiat, Shula, and Kamila Polišenská. 2016. A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *JSLHR* 59: 1179–89. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chilla, Solveig, Cornelia Hamann, Phillipe Prévost, Lina Abed Ibrahim, Christophe Ferré Sandrine, dos Santos Zebib Racha, and Tuller Laurice. 2021. The influence of different first languages on L2 LITMUS-SRT in French and German: A crosslinguistic approach. In *LITMUS in Action: Cross-Comparison Studies across Europe. Trends in Language Acquisition Research*. Edited by Sharon Armon-Lotem and Kleanthes K. Grohmann. Amsterdam: Benjamins, pp. 228–62.
- Chilla, Solveig. 2008. *Erstsprache, Zweitsprache, Spezifische Sprachentwicklungsstörung? Eine Untersuchung des Erwerbs der deutschen Hauptsatzstruktur durch sukzessiv-bilinguale Kinder mit türkischer Erstsprache*. Hamburg: Dr. Kovač.
- Chondrogianni, Vasiliki. 2018. Child L2 acquisition. In *Bilingual Cognition and Language. The State of the Science across Its Subfields*. Edited by David Miller, Fatih Bayram, Jason Rothman and Ludovica Serratrice. Amsterdam: John Benjamins, pp. 103–26.
- Conti-Ramsden, Gina, Nicole Botting, and Brian Faragher. 2001. Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry* 42: 741–48. [\[CrossRef\]](#)
- Cummins, Jim. 2008. BICS and CALP: Empirical and theoretical status of the distinction. In *Encyclopedia of Language and Education*, 2nd ed. Edited by Brian Street and Nancy Hornberger. New York: Springer, pp. 487–99.
- Czinger, Christine, Katharina Korecky-Kröll, Kumru Uzunkaya-Sharma, and Wolfgang U. Dressler. 2015. Wie beeinflusst der sozioökonomische Status den Erwerb der Erst- und Zweitsprache? In *Deutsch als Zweitsprache in Schule und Unterricht*. Edited by Arne Ziegler and Klaus-Michael Köpcke. Berlin: De Gruyter, pp. 207–40. [\[CrossRef\]](#)
- Davies, Catherine, Clara Andrés-Roqueta, and Courtenay Frazier Norbury. 2016. Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology* 144: 98–113. [\[CrossRef\]](#)
- de Grüter, Theres, and Johanne Paradis. 2014. *Input and Experience in Bilingual Development*. Amsterdam: Benjamins.
- Deacon, S. Hélène, Jenna Benere, and Adrian Pasquarella. 2013. Reciprocal relationship: Children's morphological awareness and their reading accuracy across grades 2 to 3. *Developmental Psychology* 49: 1113. [\[CrossRef\]](#)
- Ditton, Hartmut. 2011. Familie und Schule—Eine Bestandsaufnahme der bildungssoziologischen Schuleffektforschung von James S. Coleman bis heute. In *Lehrbuch der Bildungssoziologie*, 2nd ed. Edited by Rolf Becker. Wiesbaden: Springer, pp. 245–64.
- dos Santos, Christophe, and Sandrine Ferré. 2018. A nonword repetition task to assess bilingual children's phonology. *Language Acquisition* 25: 58–71. [\[CrossRef\]](#)
- Duncan, Sorenson T., and Johanne Paradis. 2020. How does maternal education influence the linguistic environment supporting bilingual language development in child L2 learners of English? *International Journal of Bilingualism* 24: 46–61. [\[CrossRef\]](#)
- Durgunoğlu, Aydin Yücesan. 2002. Cross-linguistic transfer in literacy development and implications for language learners. *Annals of Dyslexia* 52: 189–204. [\[CrossRef\]](#)
- Edele, Aileen, Sebastian Kempert, and Petra Stanat. 2020. Mehrsprachigkeit und Bildungserfolg. In *Handbuch Mehrsprachigkeit und Bildung*. Edited by Ingrid Gogolin, Sarah McMonagle, Dominique Rauch and Antje Hansen. Wiesbaden: Springer, pp. 151–55.
- Elsner, Daniela. 2015. *Kompetenzorientiert Unterrichten in der Grundschule: Englisch 1–4*. München: Oldenbourg.
- Esser, Günter, Anne Wyschkon, and Katja Ballaschk. 2008. *BUEGA. Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter*. Göttingen: Hogrefe.
- Feilke, Helmuth. 2012. Bildungssprachliche Kompetenzen—Fördern und entwickeln. Basisartikel. *Praxis Deutsch* 233: 4–13.
- Fernald, Anne, Virginia A. Marchman, and Adriana Weisleder. 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science* 16: 234–48. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ferré, Sandrine, Laurie Tuller, Eva Sizaret, and Marie-Anne Barthez. 2012. Acquiring and avoiding phonological complexity in SLI vs. typical development of French: The case of consonant clusters. In *Consonant Clusters and Structural Complexity*. Edited by Philip Hoole, Lasse Bombine, Marianne Pouplier, Christine Mooshammer and Barbara Kühnert. Berlin: de Gruyter, pp. 285–308.
- Fox, Anette. 2009. *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner.
- Fox-Boyer, Anette. 2014. *Psycholinguistischer Analyse kindlicher Aussprachestörungen-II: PLAKSS II*. Frankfurt am Main: Pearson.
- Fredman, Marion. 2006. Recommendations for working with bilingual children – prepared by the multilingual affairs committee of IALP. *Folia Phoniatrica et Logopaedica* 58: 458–64. [\[CrossRef\]](#)
- Friedmann, Naama, and Rama Novogrodsky. 2008. Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI. In *Language Acquisition and Development*. Edited by Anna Gavarró and João M. Freitas. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 205–17.
- Gagarina, Natal'ja Vladimirovna, Klassert Annegret, and Topaj Nathalie. 2010. *Sprachstandtest Russisch für Mehrsprachige Kinder*. ZAS Papers in Linguistics 54. Berlin: Zentrum für Allgemeine Sprachwissenschaft.
- Genesee, Fred, Johanne Paradis, and Martha B. Crago. 2004. *Dual Language Development & Disorders: A Handbook on Bilingualism & Second Language Learning*. Baltimore: Paul H. Brookes Publishing, vol. 11.
- Glück, Christian. 2011. *WWT 6-10. Wortschatz- und Wortfindungstest für 6-10-jährige Kinder*, 2nd ed. Munich: Elsevier.

- Gogolin, Ingrid, and Joana Duarte. 2016. Bildungssprache. In *Handbuch Sprache in der Bildung*. Edited by Jörg Kilian, Birgit Brouer and Diana Lüttenberg. Berlin: De Gruyter, pp. 478–500.
- Gottardo, Alexandra, Norah Amin, Asma Amin, Redab Al-Janaideh, Xi Chen, and Johanne Paradis. 2020. Word reading in English and Arabic in children who are Syrian refugees. *Applied Psycholinguistics* 41: 1305–28. [CrossRef]
- Gough, Phillip, and William Tunmer. 1986. Decoding, reading, and reading disability. *Remedial and Special Education* 7: 6–10. [CrossRef]
- Graham, Rebecca A., Joy D. Osofsky, Howard J. Osofsky, and Tonya C. Hansel. 2017. School based post disaster mental health services: Decreased trauma symptoms in youth with multiple traumas. *Advances in School Mental Health Promotion* 10: 161–75. [CrossRef]
- Grimm, Angela, and Julia Hübner. in press. Nonword repetition by bilingual learners of German. The role of language-specific complexity. In *Bilingualism and Specific Language Impairment*. Edited by Christophe dos Santos and Laetitia de Almeida. Amsterdam: Benjamins.
- Grimm, Angela, Srine Ferré, Christophe dos Santos, and Shula Chiat. 2014. Can nonwords be language-independent? Cross-linguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese. In *Symposium Language Impairment Testing in Multilingual Setting (LITMUS): Disentangling Bilingualism and SLI*. Amsterdam: IASCL.
- Hadfield, Kristin, Aly Ostrowski, and Michael Ungar. 2017. What can we expect of the mental health and wellbeing of Syrian refugee children and adolescents in Canada? *Canadian Psychology/Psychologie Canadienne* 58: 194. [CrossRef]
- Hamann, Cornelia, and Lina Abed Ibrahim. 2017. Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication* 2: 16. [CrossRef]
- Hamann, Cornelia, Chilla Solveig, Abed Ibrahim Lina, and Fekete Istvan. 2020. Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Applied Psycholinguistics* 40: 1–40. [CrossRef]
- Hamann, Cornelia, Solveig Chilla, Esther Ruigendijk, and Lina Abed Ibrahim. 2013. A German Sentence Repetition Task: Testing Bilingual Russian/German Children. Paper presented at the COST Meeting, Krakow, Poland, 24 May 2013.
- Hamann, Cornelia. 2012. Bilingual Development and Language Assessment. In *Proceedings of the 36th Annual Boston Conference on Language Development (BUCLD)*. Edited by Alia K. Biller, Esther Y. Chung and Amelia E. Kimball. Somerville: Cascadia Press, pp. 1–28.
- Hammer, Carol Scheffner, Frank R. Lawrence, and Adele W. Miccio. 2007. Bilingual Children's Language Abilities and Early Reading Outcomes in Head Start and Kindergarten. *Language, Speech, and Hearing Services in Schools* 38: 237–48. [CrossRef]
- Hoff, Erika. 2003. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development* 74: 1368–78. [CrossRef]
- Hoff, Erika, Brett Laursen, and Twila Tardif. 2002. Socioeconomic status and parenting. In *Handbook of Parenting*, 2nd ed. Edited by Marc H. Bornstein. Mahwah: Erlbaum, pp. 231–52.
- Hußmann, Anke, Heike Wendt, Wilfried Bos, Kasper Bremerich-Vos, Daniel Albert, Eva-Maria Lankes, Nele McElvany, Tobias Stubbe, and Renate Valtin. 2017. *IGLU. Lesekompetenzen von Grundschulkindern in Deutschland im Internationalen Vergleich*. Münster: Waxmann.
- International Association of Logopedics and Phoniatrics [IALP]. 2011. Recommendations for Working with Bilingual Children. Available online: http://www.specchioriflesso.net/media/162083/linee_guida_bilingui_ialp_may_2011.pdf (accessed on 15 December 2021).
- IBM SPSS 27. 2020. *IBM SPSS Statistics for Windows, Version 27.0*. Armonk: IBM Corp.
- Jia, Gisela, and Akiko Fuse. 2007. Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research* 50: 1280–99. [CrossRef]
- Kaplan, Ida, Yvonne Stolk, Madeleine Valibhoy, Alan Tucker, and Judy Baker. 2016. Cognitive assessment of refugee children: Effects of trauma and new language acquisition. *Transcult Psychiatry* 53: 81–109. [CrossRef] [PubMed]
- Kempert, Sebastian, Aileen Edele, Dominique Rauch, Katrin M. Wolf, Jennifer Paetsch, Annkathrin Darsow, Jessica Maluch, and Petra Stanat. 2016. Die Rolle der Sprache für zuwanderungsbezogene Ungleichheiten im Bildungserfolg. In *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten*. Edited by Diehl Claudia, Hunkler Christian and Kristen Cornelia. Wiesbaden: Springer, pp. 157–241.
- Kirby, John R., Hélène S. Deacon, Peter N. Bowers, Leah Izenberg, Wade-Woolley Lesley, and Rauno Parrila. 2012. Children's morphological awareness and reading ability. *Reading and Writing* 25: 389–410. [CrossRef]
- Klieme, Eckhard, Cordula Artelt, Johannes Hartig, Nina Jude, Olaf Köller, Manfred Prenzel, Wolfgang Schneider, and Petra Stanat. 2010. *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- KMK. 2016. Erklärung der Kultusministerkonferenz zur Integration von jungen Geflüchteten durch Bildung. Beschluss der Kultusministerkonferenz vom 06.10.2016. Available online: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2016/2016_10_06-Erklaerung-Integration.pdf (accessed on 21 September 2021).
- Kohnert, Kathryn. 2013. *Language Disorders in Bilingual Children and Adults*, 2nd ed. San Diego, Oxford and Melbourne: Plural Publishing.
- Leonard, Laurence B. 2014a. *Children with Specific Language Impairment*, 2nd ed. Cambridge: MIT Press.
- Leonard, Laurence B. 2014b. Specific language impairment across languages. *Child Development Perspectives* 8: 1–5. [CrossRef] [PubMed]
- Marinis, Theodoros, and Sharon Armon-Lotem. 2015. Sentence repetition. In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Edited by Sharon Armon-Lotem, Jan de Jong and Natalia Meir. Bristol: Multilingual Matters, pp. 95–124.

- Marinis, Theodoros. 2011. On the nature and cause of specific language impairment: A view from sentence processing and infant research. *Lingua* 121: 463–75. [\[CrossRef\]](#)
- Meisel, Jürgen. 1990. *Two First Languages: Early Grammatical Development in Bilingual Children*. Dordrecht: Foris Publications.
- Meisel, Jürgen. 2009. Second Language Acquisition in Early Childhood. *Zeitschrift für Sprachwissenschaft* 28: 5–34. [\[CrossRef\]](#)
- Montanari, Elke G., and Roman Abel. 2017. Vocabulary Development in German-Turkish Language contact. In *The Rouen Meeting. Studies on Turkic Structures and Language Contacts*. Edited by Mehmet Ali Akıncı and Kutlay Yağmur. Heidelberg: Harrassowitz, pp. 253–66.
- Montrul, Silvina. 2016. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.
- Nagy, William, Virginia Berninger, and Robert D. Abbott. 2006. Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology* 98: 134. [\[CrossRef\]](#)
- Nation, Kate, and Courtenay F. Norbury. 2005. When reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders* 25: 21–32. [\[CrossRef\]](#)
- Norbury, Courtenay Frazier, Debbie Gogh, Charlotte Wray, Gillian Baird, Tony Charman, Emily Simonoff, George Vamvakas, and Andrew Pickles. 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: 1247–57. [\[CrossRef\]](#)
- Novogrodsky, Rama, and Varda Kreiser. 2015. What can errors tell us about specific language impairment deficits? Semantic and morphological cuing in a sentence completion task. *Clinical Linguistics & Phonetics* 29: 812–25.
- Olczyk, Melanie, Julian Seuring, Gisela Will, and Sabine Zinn. 2016. Migranten und ihre Nachkommen im deutschen Bildungssystem: Ein aktueller Überblick. In *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*. Edited by Claudia Diehl, Christian Hunkler and Cornelia Kristen. Wiesbaden: Springer, pp. 33–70.
- Palikara, Olympia, Julie E. Dockrell, and Geoff Lindsay. 2011. Patterns of Change in the Reading Decoding and Comprehension Performance of Adolescents with Specific Language Impairment (SLI). *Learning Disabilities: A Contemporary Journal* 9: 89–105.
- Paradis, Johanne, Adriana Soto-Corominas, Evangelika Daskalaki, Xi Chen, and Alexra Gottardo. 2021. Morphosyntactic Development in First Generation Arabic—English Children: The Effect of Cognitive, Age, and Input Factors over Time and across Languages. *Languages* 6: 51. [\[CrossRef\]](#)
- Paradis, Johanne, Brian Rusk, Tamara S. Duncan, and Krithika Govindarajan. 2017. Children's second language acquisition of English complex syntax: The role of age, input, and cognitive factors. *Annual Review of Applied Linguistics* 37: 148–67. [\[CrossRef\]](#)
- Paradis, Johanne, and Ruiting Jia. 2017. Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental Science* 20: e12433. [\[CrossRef\]](#) [\[PubMed\]](#)
- Paradis, Johanne. 2010. The interface between bilingual development and specific language impairment. *Applied Psycholinguistics* 31: 227–52. [\[CrossRef\]](#)
- Paradis, Johanne. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* 1: 213–37. [\[CrossRef\]](#)
- Pham, Giang, and Timothy Tipton. 2018. Internal and external factors that support children's minority first language and English. *Language, Speech, and Hearing Services in Schools* 49: 595–606. [\[CrossRef\]](#)
- Powell, Justin J.W., and Sandra J. Wagner. 2014. An der Schnittstelle Ethnie und Behinderung benachteiligt. In *Behinderung und Migration*. Edited by Gudrun Wansing and Manuela Westphal. Wiesbaden: Springer, pp. 177–99.
- Prevoo, Mariëlle, Maike Malda, Judi Mesman, and Rosanneke A.G. Emmen. 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: 1–22. [\[CrossRef\]](#)
- Quigley, Duana, Fíodhna Gardiner-Hyl, Deidre Murphy, and Ciara O'Toole. 2020. Best Practice Guidelines for Multilingual Children: A Cross-Disciplinary Comparison. *Learn* 41: 18–34.
- Rothman, Jason, Drew Long, Michael Iverson, Tiffany Judy, Tushar Chakravarty, and Anne Lingwall. 2016. Older age of onset in child L2 acquisition can be facilitative: Evidence from the acquisition of English passives by Spanish natives. *Journal of Child Language* 43: 662–86. [\[CrossRef\]](#)
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13: 155–63. [\[CrossRef\]](#)
- Rothweiler, Monika. 2006. Spezifische Sprachentwicklungsstörung und kindlicher Zweitspracherwerb. In *Sprache–Emotion–Bewusstheit. Beiträge zur Sprachtherapie in Schule, Praxis, Klinik*. Edited by Reiner Bahr and Claudia Iven. Idstein: Schulz-Kirchner Verlag, pp. 154–63.
- Schiff, Rachel, and Elinor Saiegh-Haddad. 2018. Development and relationships between phonological awareness, morphological awareness and word reading in spoken and standard Arabic. *Frontiers in Psychology* 9: 356. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schulz, Petra, and Angela Grimm. 2020. *Phonology and Sentential Semantics: Markers of SLI in Bilingual Children at Age 6*. LITMUS in Action. Amsterdam: Benjamins.
- Schulz, Petra, and Rosemarie Tracy. 2011. *LiSe-DaZ: Linguistische Sprachstandserhebung–Deutsch als Zweitsprache*. Göttingen: Hogrefe.
- Schwartz, Bonnie D. 2004. Why child L2 acquisition? In *Proceedings of Generative Approaches to Language Acquisition 2003*. LOT Occasional Series; Edited by Jacqueline Van Kampen and Sergio Baauw. Utrecht: LOT, pp. 47–66.

- Senatsverwaltung für Bildung, Jugend und Familie. 2017. *Leitfaden zur Feststellung Sonderpädagogischen Förderbedarfs an Berliner Schulen*. Berlin: Senatsverwaltung für Bildung, Jugend und Familie.
- Soto-Corominas, Adriana, Johanne Paradis, Redab AlJanaideh, Irene Vitoroulis, Xi Chen, Katholiki Georgiades Jenkins, and Alexra Gottardo. 2020. Socioemotional Wellbeing Influences Bilingual and Biliteracy Development: Evidence from Syrian Refugee Children. In *Proceedings of the 44th Boston University Conference on Language Development*. Edited by Megan M. Brown and Alexandra Kohut. Somerville: Cascadilla Press, pp. 620–33.
- Statistisches Bundesamt. 2021. Bevölkerung: Migration und Integration. Bevölkerung in Privathaushalten nach Migrationshintergrund im weiteren Sinn nach ausgewählten Geburtsstaaten. Available online: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Tabellen/migrationshintergrund-staatsangehoerigkeit-staaten.html> (accessed on 22 June 2021).
- Tabors, Patton O., Mariela Pérez, and Lisa M. López. 2003. Dual language abilities of bilingual four-year olds: Initial findings from the early childhood study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice* 1: 70–91.
- Thordardottir, Elin. 2015. Proposed Diagnostic Procedures for Use in Bilingual and Cross-Linguistic Context. In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters, pp. 331–58.
- Tomblin, Bruce J., Nancy L. Records, Paula Buckwalter, Xuyang Zhang, Elaine Smith, and Marlea O'Brien. 1997. Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research* 40: 1245–60. [CrossRef]
- Tsimpli, Ianthi, Eleni Peristeri, and Andreou Maria. 2016. Object Clitic production in monolingual and bilingual children with Specific Language Impairment: A comparison between elicited production and narratives. *Linguistic approaches to Bilingualism* 7: 394–430. [CrossRef]
- Tuller, Laurice, Cornelia Hamann, Solveig Chilla, Srine Ferré, Eléonore Morin, Philippe Prevost, Christophe dos Santos, Lina Abed Ibrahim, and Racha Zebib. 2018. Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders* 53: 888–904.
- Tuller, Laurie. 2015. Clinical use of parental questionnaires in multilingual contexts. In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Berlin: De Gruyter, pp. 229–328.
- Unsworth, Sharon. 2005. Child L2, Adult L2, Child L1 Differences and Similarities. A Study on the Acquisition of Object Scrambling in Dutch. Doctoral dissertation, Utrecht Institute of Linguistics OTS, LOT, Netherlands Graduate School of Linguistics, Amsterdam, The Netherlands.
- Unsworth, Sharon. 2016. Early child L2 acquisition. Age or input effects? Neither, or both? *Journal of Child Language* 43: 608–34. [CrossRef]
- Unsworth, Sharon. 2019. Quantifying experience in heritage language development. In *The Oxford Handbook of First Language Attrition*. Edited by Monika Schmid and Barbara Köpke. Oxford: Oxford University Press, pp. 74–93. [CrossRef]
- Vernon-Feagans, Lynne, Mary Bratsch-Hines, Elizabeth Reynolds, and Michael Willoughby. 2019. How early maternal language input varies by race and education and predicts later child language. *Child Development* 91: 1098–115. [CrossRef] [PubMed]
- Yew, Shaun Goh Kok, and Richard O'Kearney. 2013. Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry* 54: 516–24. [CrossRef] [PubMed]
- Zebib, Rasha, Guillemette Henri, Abdelhamid Khomsi, Camille Messara, and Edith Kouba Hreich. 2017. *Batterie d'Evaluation du Langage Oral chez l'enfant Libanais (ELO-L)*. Liban: LTE.