

Language assessment for multilingual children in Germany - developmental factors, environmental influences, and individual differences

Onur Özsoy (ooezsoy@uni-koeln.de)

University of Cologne, Cologne, Germany and Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Elna Bauer (bauer@leibniz-zas.de)

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Katharina Bleher (bleher@em.uni-frankfurt.de)

Goethe University Frankfurt, Frankfurt, Germany

Aylin Cil (aylin.cil@ku.de)

Catholic University of Eichstätt-Ingolstadt, Eichstätt-Ingolstadt, Germany

Tina Ghaemi (tina.ghaemi@tu-dortmund.de)

TU Dortmund, Dortmund, Germany

Annika Heitzmann (a.heitzmann@em.uni-frankfurt.de)

Goethe University Frankfurt, Frankfurt, Germany

Eugenia Rykova (eugenia.rykova@ku.de)

Catholic University of Eichstätt-Ingolstadt, Eichstätt-Ingolstadt, Germany

Jenny Thillmann (jenny.thillmann@tu-dortmund.de)

TU Dortmund, Dortmund, Germany

Angela Grimm (a.grimm@em.uni-frankfurt.de)

Goethe University Frankfurt, Frankfurt, Germany

Tanja Rinker (tanja.rinker@ku.de)

Catholic University of Eichstätt-Ingolstadt, Eichstätt-Ingolstadt, Germany

Anna-Lena Scherger (anna-lena.scherger@tu-dortmund.de)

TU Dortmund, Dortmund, Germany

Natalia Gagarina (gagarina@leibniz-zas.de)

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Abstract

Assessing language abilities in multilingual children is critical for identifying necessary support. The SPEAK project (German acronym for “Language assessment of multilingual children”) validates a comprehensive test battery for 4- to 8-year-olds in Germany. This study reports on data from 207 multilingual children with 50 first languages other than German. The battery includes German versions of internationally established tools: Nonword Repetition Task (phonological complexity), Cross-Linguistic Lexical Task (vocabulary comprehension and production), and Sentence Repetition Task (grammar), alongside a parental questionnaire. Results show that age strongly predicts task performance, with earlier exposure to German improving phonology, vocabulary and grammar. Parental education also consistently predicts outcomes. Suspected Developmental Language Disorder negatively affects receptive vocabulary and grammar. Findings highlight the complex interplay of factors in multilingual language development.

Keywords: Multilingual Language Assessment; Developmental Language Disorder; Nonword Repetition Task; Cross-Linguistic Lexical Task; Sentence Repetition Task

Introduction

Multilingualism, increasingly common worldwide, brings a range of cognitive, cultural, and social benefits (Xia et al., 2022). Cognitively, bilingual and multilingual individuals might demonstrate enhanced executive functions, such as improved attentional control, cognitive flexibility, and problem-solving skills, due to their frequent need to switch between languages (Bialystok, 2011). Culturally, multilingualism fosters a deeper appreciation of diverse perspectives, enabling individuals to engage more meaningfully with different traditions, literatures, and ways of thinking (Grosjean, 2010). Socially, speaking multiple languages facilitates communication across linguistic boundaries, as multilingual individuals may develop heightened empathetic abilities (Alqarni & Dewaele, 2020).

However, multilingualism also presents unique challenges in assessing and supporting language development in educa-

tional and clinical spheres. This applies particularly in contexts where diagnostic tools are designed for monolingual populations. In Germany, as in many other countries, language assessments often rely on monolingual norms, leading to systemic biases in identifying language difficulties in multilingual children/individuals. This results in overdiagnoses, where language differences are misinterpreted as deficits, and underdiagnoses, where genuine language disorders go unnoticed (Altman et al., 2022; Grimm & Schulz, 2014; Paradis et al., 2011). Diagnostic inaccuracies, with special attention to Developmental Language Disorder (DLD), have profound educational and social consequences for multilingual children.

DLD under- and overdiagnoses in multilingual children

Developmental Language Disorder is a neurodevelopmental condition that affects language acquisition and use. It is distinct from language disorders associated with known biomedical conditions such as brain injury or intellectual disability, though it may co-occur with other neurodevelopmental difficulties (Bishop, 2017). Children with Developmental Language Disorder exhibit difficulties in phonology, grammar, vocabulary acquisition, and complex sentence processing, which can impact their academic and social development (Thomas et al., 2019). Cognitive factors, such as working memory, are often implicated in Developmental Language Disorder, as these children struggle with linguistic encoding and retrieval (Gillam et al., 2019). Importantly, distinguishing Developmental Language Disorder from typical multilingual language development is crucial, as both groups may display delayed or non-standard language trajectories due to differing linguistic input and exposure.

The diagnostic challenges surrounding Developmental Language Disorder are particularly pronounced in multilingual populations. Overdiagnosis occurs when typical multilingual language patterns—such as cross-linguistic transfer or limited exposure effects—are mistaken for clinical disorders. Conversely, underdiagnosis arises when genuine language difficulties are dismissed as a consequence of multilingual exposure, leading to missed intervention opportunities (Paradis et al., 2011). Research has shown that standardized language assessments developed for monolinguals tend to misrepresent multilingual children’s linguistic abilities, reinforcing both false-positive and false-negative errors in diagnosis (Tuller, 2015). Addressing these issues necessitates assessment tools that account for multilingual language development and exposure-related variability (Schulz & Tracy, 2011).

Linguistic and socio-demographic factors shaping language acquisition

Typical and atypical language development are influenced by a complex interplay of linguistic and developmental as well as socio-demographic and environmental factors. Input quantity

and quality, socioeconomic status, parental language proficiency and parental education could all contribute to language acquisition outcomes (Ebert & Lee, 2024). More generally, studies confirm that children with sustained and high-quality exposure to both their home and societal languages generally achieve strong bilingual proficiency, whereas inconsistent input may lead to weaker linguistic skills in one or both languages. In the context of Developmental Language Disorder, these factors are particularly relevant, as they interact with inherent difficulties in language acquisition, either masking or exacerbating symptoms (Boerma & Blom, 2017).

The LITMUS test battery

To improve the assessment of multilingual children’s language abilities, the LITMUS (Language Impairment Testing in Multilingual Settings) battery was developed (Armon-Lotem et al., 2015). This test battery is designed to distinguish between Developmental Language Disorder and language difficulties arising from limited exposure. It includes tools such as the Nonword Repetition (NWR) task, Sentence Repetition Task (SRT), and Cross-Linguistic Lexical Task (CLT), which assess phonological complexity, syntactic competence, and lexical knowledge, respectively. These tasks are mostly language-independent and adapted to multilingual contexts to reduce biases stemming from monolingual norms. Henceforth, for the sake of brevity, the LITMUS prefix is omitted for all tests. Details about the LITMUS tasks are given in the Materials section.

Addressing the above-mentioned challenges requires linguistically and culturally sensitive assessment tools. The SPEAK (Language Diagnostics in Multilingual Children: Validation of a Test Battery) project’s goal is to respond to this need. Specifically, the project aims to validate an assessment battery for multilingual children, namely TEBIK 4–8 (Test for Bilingual Children aged 4 to 8), and to establish age-specific norms. The project primarily focuses on ensuring the applicability of the battery in German-speaking countries. However, the current study does not propose norms, and instead employs said test battery to investigate the interplay of linguistic, cognitive, and socio-demographic factors in multilingual children’s acquisition of German. By focusing on key domains of language development—phonology, vocabulary, and grammar—we aim to provide insights that bridge the gap between research and application in linguistically diverse populations.

To guide this investigation, we pose the following overarching research questions:

How do linguistic and socio-demographic factors as well as Developmental Language Disorder risk factors interact to shape language development in multilingual children?

Hypotheses

For each of the three assessment tools—NWR, CLT, and SRT—the study formulates the following hypotheses, each grounded in existing literature and empirical evidence.

1. Linguistic and socio-demographic variables

Linguistic and socio-demographic factors, such as parental education, have been shown to play a critical role in language development, with cumulative exposure and high-quality input often correlating with better linguistic outcomes.

Hypothesis: Chronological age, age of onset, length of exposure, sibling configuration, parental language proficiency, and parental education (highest school-leaving qualification) meaningfully influence multilingual children's language outcomes.

Prediction: Children who are older and who have been exposed to German from an earlier age and for a longer duration will demonstrate stronger phonology, vocabulary, and grammar skills. Additionally, children with older siblings and with parents with higher German levels and educational backgrounds are expected to exhibit better language outcomes (Gagarina et al., 2014; Schulz & Grimm, 2012).

2. Risk factors for Developmental Language Disorder

Developmental Language Disorder impacts language development across linguistic contexts, but multilingual assessment tools that account for variability can help differentiate between Developmental Language Disorder and underperformed language abilities due to limited L1 exposure. Additionally, there are critical risk factors for Developmental Language Disorder, such as familial history of Developmental Language Disorder or delayed occurrence of speech and language milestones.

Hypothesis: Multilingual children with suspected Developmental Language Disorder and associated risk factors (familial history of Developmental Language Disorder, delayed milestones such as first words and first sentences) will exhibit lower performance on language tasks compared to peers without these risk factors.

Prediction: Multilingual children with Developmental Language Disorder associated risk factors show deficits in NWR, CLT, and SRT, regardless of their multilingual language background, aligning with findings in monolingual contexts (Boerma & Blom, 2017; Tuller, 2015).

Methodology

Recruitment and participants

Participants, 4;00 to 8;11 year-old multilingual children living in Germany, were recruited in four urban areas in Germany. Recruitment was mostly conducted with the help of institutions such as nurseries and primary schools to ensure a wide distribution and participation of diverse populations. Children with at least six months of exposure to German, who were acquiring at least one additional language at home, were considered for the study.

The subsample of the present study includes a total of 207 children. The gender distribution was 49.5% female, 50%

male, and 0.5% diverse. Of these participants, 32% had been referred to speech and language therapy by their pediatrician.

The children's ages ranged from 4;00 to 8;11, with an average of 6;4 years. The age of onset of German ranged from 0;00 to 7;10, with an average of 2;0 years. The length of exposure to German varied between 0;8 and 8;6 years, with an average duration of 4;6 years.

The participants represented a linguistically diverse group, with 50 different first languages. Children were classified as suspected of having Developmental Language Disorder if they had been referred to a speech and language pathologists.

Materials

A range of tools specifically designed for multilingual language assessment were employed. These tools were selected based on their diagnostic validity and sensitivity to multilingual language acquisition contexts.

Nonword Repetition Task

The NWR is a phonological processing test that evaluates children's ability to repeat nonwords of varying phonological complexity (Grimm, 2022; Grimm et al., 2014). Nonword repetition is a robust indicator of phonological working memory and has been widely used to differentiate typical language development from developmental language disorder (Chiat, 2015). In this study, children were asked to repeat a series of nonwords presented via a standardized protocol to ensure consistency across participants.

Cross-Linguistic Lexical Task

The CLT assesses vocabulary knowledge in children. It involves picture-naming and picture-selection tasks, enabling the measurement of expressive and receptive vocabulary (Haman et al., 2015; Rinker & Gagarina, 2014). The task is particularly suited to multilingual contexts as it can account for cross-linguistic differences in lexical development.

Sentence-Repetition Task

The SRT is designed to evaluate syntactic and grammatical processing by requiring children to repeat sentences of varying complexity. This task targets the ability to process and reproduce grammatical structures, offering insights into the morphosyntactic competence and working memory of multilingual children (Hamann et al., 2013). The SRT has been validated across multiple languages and provides diagnostic sensitivity to Developmental Language Disorder.

Parental questionnaire

A parental questionnaire was administered to gather critical background information, including language exposure history, age of onset of each language, and potential risk factors for Developmental Language Disorder (Tuller, 2015). The questionnaire is available at <https://osf.io/hvadx>.

Procedure

Parents gave their written consent and filled in the parental questionnaire at home. Participation was entirely voluntary.

Appointments were scheduled at the participating children’s nursery or school. Typically, two to three sessions of 30-45 minutes each were required to complete all tests with a child. Children had the autonomy to decide whether to participate and were allowed to take as many breaks as needed. Sessions were generally supervised by educational personnel from the institution, ensuring a familiar and comfortable environment for the child. Examples were worked on together prior to each task to ensure that the child understood the instructions. If this was not the case, the testers moved on to the next task.

Ethical approval for the study was granted by the ethics committee of the German Linguistics Society (DGfS, Votum number 2022-16). Trained testers administered the assessments. All linguistic tasks were presented using a tablet to display materials. Additional cognitive tests were conducted using a tablet or pen-and-paper. Recordings were made using a tablet, a voice recorder, or a video camera to ensure high-quality data collection.

Data analysis

With the help of the recordings, children’s responses were later transcribed and manually coded by a different annotator. Double-coding was performed for a subset of the data to ensure coding consistency across annotators. For each subtest, one point was given per correct answer following standardized annotation guidelines. Following this, several binomial generalized linear mixed models were conducted using R (R Core Team, 2021) to analyze the results of each test separately. The models included random effects for participants and items to account for within-subject variability.

The children’s test score accuracies served as binary dependent variables, while independent variables included age, age of onset of German, length of exposure, previous referral to speech and language therapy for suspected Developmental Language Disorder, and other socio-demographic, linguistic, and cognitive background factors. These predictors allowed for a nuanced understanding of the factors influencing multilingual language acquisition and performance.

Data processing in R was conducted using the *tidyverse* package (Wickham et al., 2019), while the generalized linear mixed models were implemented using the *lme4* package (Bates et al., 2015). Continuous variables were centered around their mean to facilitate interpretation of model coefficients, while categorical variables with more than two levels were contrast-coded using sum-coding to ensure meaningful comparisons between factor levels. All code and data are openly accessible for transparency and reproducibility at <https://osf.io/tnzag/>.

Results

The results are presented in alignment with the research questions and organized according to the investigated domains: phonological development, vocabulary development, and grammatical development. Table 1 presents the overall results of four generalized linear indicating significant effects

with asterisks. In the following, we provide detailed descriptive and inferential results for the relevant variables in the sample.

Table 1: Overview of predictors in the four models

Predictor	NWR	CLT Rec.	CLT Prod.	SRT
Age (months)	***	***	***	***
Age onset German	*	***	***	**
Parental education	***	**	*	***
Sibling order	n.s.	n.s.	n.s.	*
Suspected DLD	n.s.	**	n.s.	***
Family DLD	n.s.	n.s.	n.s.	n.s.
Age first words	n.s.	n.s.	n.s.	n.s.
Parental German proficiency	n.s.	n.s.	n.s.	n.s.
Age × Age onset German	n.s.	n.s.	n.s.	n.s.

Note. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$, n.s. = not significant.

Phonological development

The descriptive analysis of the NWR scores in the sample revealed an average score of 62% (SE=0.06), with scores ranging from 13% to 100%. This indicates a diverse sample in terms of phonological proficiency and phonological working memory, with a wide spread of abilities within the group, reflecting substantial individual differences.

A generalized linear mixed model with a binomial link function was fitted to investigate the effects of various predictors on NWR accuracy. Age emerged as a significant predictor, with older children demonstrating higher accuracy ($\beta = 0.547, SE = 0.138, z = 3.949, p < .001$). Similarly, parental education level had a significant positive effect on NWR accuracy, suggesting that children from households with higher parental education levels performed better ($\beta = 0.488, SE = 0.136, z = 3.590, p < .001$). Age of onset of German showed a significant effect, with later onset being associated with slightly higher accuracy ($\beta = 0.272, SE = 0.139, z = 1.957, p < .05$). Family history of Developmental Language Disorder was not statistically significant at the conventional $\alpha = .05$ level. Suspected Developmental Language Disorder showed no significant effect ($p = .062$).

Other predictors, including sibling order, parental German proficiency, and age of onset of first words, did not reach statistical significance (all $p > .10$). The interaction between age and age of onset of German was also non-significant ($p = .704$), suggesting that the effect of age on NWR accuracy was not modulated by the timing of German language exposure.

Taken together, these results suggest that while age and parental education are strong predictors of NWR accuracy, the role of familial Developmental Language Disorder history and prior referral to speech and language therapy warrants further investigation.

Vocabulary development

Receptive vocabulary The descriptive analysis of the CLT receptive (CLT Rec.) scores revealed an average score of 82% ($SE = 0.03$), with scores ranging from 48% to 98%. The spread of scores demonstrates variability. These results highlight a relatively high overall level of lexical comprehension within the sample, alongside notable individual differences.

A generalized linear mixed model with a binomial link function was fitted to investigate the effects of various predictors on receptive accuracy. Age emerged as a strong predictor, with older children demonstrating significantly higher accuracy ($\beta = 0.772, SE = 0.103, z = 7.498, p < .001$). Conversely, age of onset of German had a significant negative effect, indicating that later exposure to German was associated with lower receptive performance ($\beta = -0.597, SE = 0.100, z = -5.990, p < .001$). Higher parental education level was associated with better receptive performance ($\beta = 0.298, SE = 0.096, z = 3.110, p = .002$).

Previously assumed Developmental Language Disorder showed a significant negative effect, where children who had potentially experienced Developmental Language Disorder tended to have lower receptive accuracy ($\beta = -0.529, SE = 0.209, z = -2.528, p = .011$), potentially reflecting underlying linguistic difficulties. In contrast, family history of Developmental Language Disorder, sibling order, parental German proficiency, and age of onset of first words did not significantly predict receptive performance (all $p > .10$).

Regarding interaction effects, the interaction between age and age of onset of German was not significant ($\beta = 0.002, SE = 0.097, z = 0.023, p = .981$), indicating that the effect of age on receptive accuracy was not modulated by the timing of German exposure.

Taken together, these findings indicate that age, early exposure to German, and parental education play key roles in shaping children's receptive lexical abilities, while suspected Developmental Language Disorder appears to be associated with lower performance.

Productive vocabulary The descriptive analysis of the CLT productive (CLT Prod.) scores revealed an average performance of 52% ($SE = 0.07$), with scores ranging from 11% to 94%. While the average indicates moderate proficiency, the range highlights variability, reflecting both low-performing and some high-performing participants. These findings suggest a diverse level of expressive lexical ability within the sample, with substantial individual differences.

A generalized linear mixed model with a binomial link function was fitted to investigate the effects of various predictors on productive vocabulary accuracy. Age emerged as a strong predictor, with older children demonstrating significantly higher accuracy ($\beta = 0.743, SE = 0.100, z = 7.432, p < .001$). Conversely, age of onset of German had a significant negative effect, indicating that later exposure to German was associated with lower productive performance ($\beta = -0.758, SE = 0.113, z = -6.722, p < .001$).

Parental education level was also significantly associ-

ated with better productive performance ($\beta = 0.334, SE = 0.121, z = 2.761, p = .006$).

In contrast, family history of Developmental Language Disorder, sibling order, parental German proficiency, and age of onset of first words did not significantly predict productive performance (all $p > .10$). Previously assumed Developmental Language Disorder also did not reach statistical significance ($\beta = -0.182, SE = 0.244, z = -0.743, p = .457$).

Regarding interaction effects, the interaction between age and age of onset of German was not significant ($\beta = 0.079, SE = 0.102, z = 0.768, p = .443$), indicating that the effect of age on productive accuracy was not modulated by the timing of German exposure.

Taken together, these findings indicate that age, early exposure to German, and parental education play key roles in shaping children's productive lexical abilities, while family history of Developmental Language Disorder and previous referral to speech and language therapy do not appear to have a significant impact.

Grammar development

The descriptive analysis of the SRT revealed an average performance of 32% ($SE = 0.04$), with scores ranging from 0% to 100%. While the average score suggests moderate performance overall, the range reflects the presence of both high-performing participants achieving the maximum score and others with minimal or no success. These results point to substantial individual differences in grammar and working memory development within the sample.

A generalized linear mixed model with a binomial link function was fitted to investigate the effects of various predictors on SRT accuracy. Age emerged as a strong predictor, with older children demonstrating significantly higher accuracy ($\beta = 0.659, SE = 0.162, z = 4.069, p < .001$). Conversely, age of onset of German had a significant negative effect, indicating that later exposure to German was associated with lower SRT scores ($\beta = -0.503, SE = 0.173, z = -2.909, p = .004$). Higher parental education level was associated with better performance in sentence repetition ($\beta = 1.075, SE = 0.191, z = 5.640, p < .001$).

Sibling order was also a significant predictor, with children who had a higher sibling order (i.e., more older siblings) showing lower SRT accuracy ($\beta = -0.536, SE = 0.234, z = -2.285, p = .022$). Furthermore, previously assumed Developmental Language Disorder showed a strong negative effect, where children who had been referred to speech and language therapy tended to have lower SRT accuracy ($\beta = -1.495, SE = 0.370, z = -4.040, p < .001$).

In contrast, family history of Developmental Language Disorder, parental German proficiency, and age of onset of first words did not significantly predict SRT performance (all $p > .10$). The interaction between age and age of onset of German was not significant ($\beta = -0.022, SE = 0.146, z = -0.149, p = .882$), suggesting that the effect of age on SRT accuracy was not influenced by the timing of German exposure.

Taken together, these findings indicate that age, early exposure to German, sibling order, and parental education play key roles in shaping children's morphosyntactic abilities, while family history of Developmental Language Disorder does not appear to have a significant impact.

Discussion

The present study examined the role of age, linguistic environment, and family background in shaping children's phonological, lexical, and grammatical abilities. It is one of the largest studies in Germany to assess the language abilities of multilingual children across different linguistic domains. Across all three linguistic domains—phonological complexity (NWR), receptive and productive vocabulary (CLT), and morphosyntactic knowledge (SRT)—age consistently emerged as a strong predictor of accuracy, highlighting the robust developmental trajectory of these skills. Additionally, parental education was a significant positive predictor across all models, reinforcing the well-documented link between educational background and children's linguistic outcomes.

The timing of German exposure played a crucial role, particularly in vocabulary and grammar development. A later onset of German was associated with lower accuracy in the receptive and productive vocabulary tasks as well as in sentence repetition, emphasizing the importance of early language exposure for lexical and morphosyntactic skills.

Although a family history of developmental language disorder was not a significant predictor, previously assumed Developmental Language Disorder—indicated by prior referral to speech and language therapy—was associated with lower scores in receptive vocabulary and sentence repetition, supporting its potential role in language difficulties. While a full and accurate diagnosis of Developmental Language Disorder would likely show a stronger effect, a key limitation of this study is the use of a proxy variable (previous referral) instead of a confirmed diagnosis. Sibling order negatively affected grammatical abilities, suggesting that younger children with more older siblings may receive less direct input in the target language, with their language exposure shaped more by interactions with their siblings than with adults.

Taken together, these findings highlight the interplay between developmental factors (age), environmental influences (early exposure and parental education), and individual differences (suspected DLD and sibling order) in shaping bilingual children's linguistic proficiency. The implications of these results are discussed in relation to bilingual language acquisition theories and educational interventions.

Individual developmental and environmental factors

The findings of this study align with previous research highlighting the significant influence of age and early language exposure on multilingual children's linguistic development (Klassert & Gagarina, 2010; Paradis et al., 2011). Consistent with earlier work, chronological age was a strong predictor across all linguistic domains, reinforcing the notion that

language proficiency develops cumulatively with increased exposure and cognitive maturation. Furthermore, the observed effects of parental education on children's lexical and grammatical skills align with studies emphasizing the role of home language environment in shaping bilingual language outcomes (Boerma & Blom, 2017; Gagarina et al., 2014). However, the lack of a strong effect for familial history of Developmental Language Disorder contrasts with previous studies suggesting that genetic predispositions contribute to language difficulties even in multilingual contexts (Tuller, 2015). A possible explanation for this discrepancy might be that some of the recently immigrated parents may not have received a diagnosis in their respective home country even though they showed signs of Developmental Language Disorder. Moreover, the negative association between sibling order and grammatical skills indicates that children with more older siblings may receive reduced direct linguistic input, particularly in multilingual language settings. These results underscore the need for nuanced, context-sensitive approaches to bilingual language assessment and intervention.

Implications for language assessment

The findings have significant implications for the assessment of multilingual children, particularly in distinguishing typical multilingual development from potential language disorders. Our findings underscore the need for comprehensive test batteries that assess multiple language domains (e.g., phonology, vocabulary, and grammar), as different variables can influence test scores differently across assessment levels. At the same time, we caution that norms should be interpreted and applied carefully to avoid overgeneralizations, given the complexity of the tests and variables involved. The significant effect of parental education also highlights the need for assessments that consider the broader context rather than relying on monolingual norms.

In this investigation, we deepen the empirical understanding of multilingual language acquisition in children and its underlying interaction of factors. The findings aim to bridge research and application, creating a better picture of linguistically diverse populations in Germany.

Acknowledgments

We thank the students assistant researchers in the SPEAK project for their valuable contributions in recording and coding the data. We are also thankful to the audience at the Bilingualism Matters Symposium 2025 for their feedback. The SPEAK project underlying this paper is funded by the German Federal Ministry of Education and Research.

References

- Alqarni, N., & Dewaele, J.-M. (2020). A bilingual emotional advantage? An investigation into the effects of psychological factors in emotion perception in Arabic and in English of Arabic-English bilinguals and Arabic/English monolinguals. *International Journal of Bilingualism*, 24(2), 141–158.
- Altman, C., Harel, E., Meir, N., Iluz-Cohen, P., Walters, J., & Armon-Lotem, S. (2022). Using a monolingual screening test for assessing bilingual children. *Clinical Linguistics & Phonetics*, 36(12), 1132–1152.
- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children. Disentangling bilingualism from language impairment*. Multilingual Matters.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(4), 229–235.
- Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, 52(6), 671–680.
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, 66, 65–76.
- Chiat, S. (2015). Nonword repetition. In *Assessing multilingual children. Disentangling bilingualism from language impairment* (pp. 125–150). Multilingual Matters.
- Ebert, K. D., & Lee, H. (2024). Individual Predictors of Language Treatment Response in Children With Developmental Language Disorder: A Systematic Review. *Journal of Speech, Language, and Hearing Research*, 67(8), 2708–2728.
- Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., & Walters, J. (2014). Age, Input Quantity and their Effect on Linguistic Performance in the Home and Societal Language among Russian-German and Russian-Hebrew Preschool Children. In R. Silbereisen, P. F. Titzmann, & J. Shavit (Eds.), *The Challenges of Diaspora Migration* (pp. 63–82). Routledge.
- Gillam, R. B., Montgomery, J. W., Evans, J. L., & Gillam, S. L. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for assessment and treatment. *International Journal of Speech-Language Pathology*, 21(3), 240–251.
- Grimm, A. (2022). The use of the LITMUS Quasi-Universal Nonword Repetition Task to identify DLD in monolingual and early second language learners aged 8 to 10. *Languages*, 7(3), 218.
- Grimm, A., Ferré, S., dos Santos, C., & Chiat, S. (2014). Can nonwords be language-independent? Cross-linguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese. *Symposium Language Impairment Testing in Multilingual Setting (LITMUS): Disentangling bilingualism and SLI*.
- Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, 7(4), 821–841.
- Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.
- Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 196–240). Multilingual Matters.
- Hamann, C., Chilla, S., Ruigendijk, E., & Abed Ibrahim, L. (2013). A German sentence repetition task: Testing bilingual Russian/German children. *Poster Presented at the COST Meeting in Krakow*.
- Klassert, A., & Gagarina, N. (2010). Der einfluss des elterlichen inputs auf die sprachentwicklung bilingualer kinder: Evidenz aus russischsprachigen migrantenfamilien in berlin. *Diskurs Kindheits-und Jugendforschung/Discourse. Journal of Childhood and Adolescence Research*, 5(4), 11–12.
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning*. Paul H. Brookes.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rinker, T., & Gagarina, N. (2014). Clt – crosslinguistic lexical task – german version.
- Schulz, P., & Grimm, A. (2012). Spracherwerb. In H. Drügh, S. Komfort-Hein, A. Kraß, C. Meier, G. Rohowski, R. Seidel, & H. Weiß (Eds.), *Germanistik. Sprachwissenschaft – Literaturwissenschaft – Schlüsselkompetenzen* (pp. 155–172). J. B. Metzler.
- Schulz, P., & Tracy, R. (2011). *LiSe-DaZ: Linguistische Sprachstandserhebung-Deutsch als Zweitsprache*. Hogrefe.
- Thomas, S., Schulz, J., & Ryder, N. (2019). Assessment and diagnosis of Developmental Language Disorder: The experiences of speech and language therapists. *Autism & Developmental Language Impairments*, 4, 2396941519842812.
- Tuller, L. (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. In S. Armon-Lotem, J. de Jong, & N. Mair (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 301–330). Multilingual Matters.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P.,

- Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Xia, T., An, Y., & Guo, J. (2022). Bilingualism and creativity: Benefits from cognitive inhibition and cognitive flexibility. *Frontiers in Psychology*, 13, 1016777.