

# The Role of Nonverbal IQ in Diagnosing Developmental Language Disorder in Multilingual Children

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

Department of Rehabilitation Sciences, TU Dortmund University, Germany

Tina Ghaemi

Department of Rehabilitation Sciences, TU Dortmund University, Germany

Anna-Lena Scherger

Department of Rehabilitation Sciences, TU Dortmund University, Germany

## Abstract

Previous research has been inconsistent in approaching the exclusion criterion of nonverbal IQ when investigating developmental language disorder (DLD) in monolingual and multilingual children. The present study investigates the influence of the controversial low nonverbal IQ range (between one and two standard deviations (SD) below the mean) on lexical and morpho-syntactical abilities. 91 multilingual children, aged 4-8;11, were tested on Crosslinguistic Lexical Task and Sentence Repetition Task in Germany. Data were analyzed using generalized linear mixed models, considering the factors nonverbal IQ, DLD status, age, gender, and length of exposure (LoE) to German. Results show that children with typical language development (TLD) outperformed those with DLD on the LITMUS tests, independent of their nonverbal IQ, supporting the validity of these tools. Language status (TLD/DLD) and LoE had the strongest impact on test performance, exceeding the effect of nonverbal IQ. Regardless of language status, nonverbal IQ affected only receptive vocabulary but not productive vocabulary or morpho-syntax. However, when applying the one SD threshold, its influence shifted from receptive vocabulary to morpho-syntactical abilities. No significant differences were found between average and low nonverbal IQ groups across most tests within the TLD and DLD groups.

**Keywords:** nonverbal IQ; developmental language disorder; multilingualism; vocabulary; grammar; morpho-syntax

## Background

There are inconsistencies in the research on the role of IQ in developmental language disorders (DLD) and in particular on the use of nonverbal IQ as an exclusion criterion in the diagnostic process of DLD. Compared to work on monolingual children in this field, bilingual children are even more underinvestigated with respect to IQ and its role in the diagnostic procedure of DLD. The present study aims to find out whether nonverbal IQ has an influence on performance in language tests in multilingual children with and without DLD.

## Definition of DLD and the Role of (Non)verbal IQ

For many years, research and clinical practice on children's language development and its disorders have been based on the definition of Specific Language Impairment (SLI) by Stark & Tallal (1981). Besides language-specific inclusion criteria they set exclusion criteria for underlying

neurological, emotional or sensory-related disorders. These also cover a diagnosis of an intellectual disability (at that time still called mental retardation), which is determined when a child scores two standard deviations (SD) below the normative mean on an overall IQ test ( $IQ < 70$ ). If there was no underlying overall IQ score, they also defined an exclusion criterion for only the nonverbal IQ: one SD below the mean, i.e.,  $< 85$ . This decision was based on the assumption that these children would score even lower in the verbal domain, so that an overall  $IQ < 70$  was expected. Accordingly, no child with a nonverbal  $IQ < 85$  was considered to have SLI.

In the following years, adaptations of the criteria were discussed controversially, especially with respect to the exclusion of nonverbal IQ as a criterion (DeThorne & Watkins, 2006; Plante, 1998; Rice, 2016; Thordadottir, 2015). Nevertheless, researchers investigating language development still consider IQ as a possible influencing variable and see the exclusion criterion as an important instrument to minimize the impact of cognitive factors on language performance (Thordadottir, 2015).

As part of the CATALISE consensus study on terminology, experts from the field of language acquisition agreed on a new definition of 'problems with language development' for English-speaking countries (Bishop et al., 2017). The term SLI has been replaced by the more comprehensive term DLD. According to that, DLD is diagnosed when language is affected but not associated with a known biomedical aetiology (e.g., Down syndrome or intellectual disability:  $IQ < 70$ ). The new definition also determined that DLD does not require a discrepancy between verbal and nonverbal abilities. This means that children with low nonverbal abilities who do not meet criteria for intellectual disability (Harris, 2013) can be included as cases of DLD (Bishop et al., 2017). Recent studies show that language problems are often associated with lower IQ scores. However, they suggest that many children with DLD may have a nonverbal IQ in the borderline range of one SD below the mean (about 89.7, Gallinat & Spaulding, 2014; see also Park et al., 2020).

For German-speaking countries, there is no agreement to date as to whether an IQ between 70 and 85 should be included in DLD terminus in the first place or whether it is a contributory condition for DLD as other biomedical

aetiologies (DLD associated with X, Lüke et al., 2023). Furthermore, the role of cognitive abilities in the definition of DLD in the ICD-11 allows also for some interpretation: “Developmental speech or language disorders [...] are outside the limits of normal variation expected for age and level of intellectual functioning” (WHO, 2022), whereby the latter is not concretized by a specific diagnostic marker.

Regarding language performance and IQ in monolingual children, literature reports mixed results: In line with expectations, vocabulary is correlated to verbal IQ (Carvajal et al., 1993; Smith et al., 2005). However, also weaker nonverbal cognitive levels seem to be associated with lower verbal skills (Lervag et al., 2019; Saar et al., 2022). On the other hand, Rice et al. (2004) and DeThorne & Watkins (2006) found no significant associations between semantic/morpho-syntactic abilities and nonverbal IQ as well as Bruinsma et al. (2022) found no correlation between nonverbal IQ and vocabulary growth which could be interpreted as processing and memorizing new words. Another more recent study by Rice et al. (2020) could show that children with DLD and a nonverbal IQ > 85 do not outperform children with DLD and IQ < 85 on vocabulary and grammar tests, which indicates that a low nonverbal IQ (< 85) is neither necessary nor sufficient for DLD.

### Multilingualism, DLD and Nonverbal IQ

Multilingualism as an additional influence on language development renders the diagnostic process of DLD even more complex. Multilingual children can have difficulties or show cross-linguistic influences in one or both languages, that are associated with their multilingualism. However, that does not necessarily mean they have DLD (Chilla et al., 2022). To meet these diagnostic challenges, so-called LITMUS tools have been developed (‘Language Impairment Testing in Multilingual Settings’) during and following COST Action IS0804 (Armon-Lotem et al., 2015). They have been proven to be meaningful tools for identifying DLD in multilingual children (Hamann & Ibrahim, 2017; Simonsen & Haman, 2017). Two of them are chosen for this study for investigating grammar and vocabulary: the Crosslinguistic Lexical Task (CLT) and Sentence Repetition tasks (SRT). CLT evaluates vocabulary skills in the modalities of comprehension and production (Haman et al., 2015), while SRT examines morpho-syntactic knowledge (Armon-Lotem & Marinis, 2015; Meir et al., 2015).

Up to date, the role of nonverbal IQ in the diagnosis of DLD in multilingual children is unclear because research concerning these LITMUS tools often focused on multilingualism-specific characteristics rather than cognitive influences. A literature review revealed 14 studies on LITMUS-CLT and LITMUS-SRT (2017 - 2024) that differentiated children with typical language development (TLD) from DLD children. As table 1 shows, research on LITMUS CLT and SRT has used changing exclusion criteria with regard to nonverbal IQ.

Table 1: Overview of exclusion criteria in CLT and SRT studies

No.	Exclusion	Studies
3	IQ < 85	Abed Ibrahim & Hamann (2024); Armon-Lotem & Meir (2016); Meir et al. (2016)
3	IQ < 80	Hamann & Abed Ibrahim (2017); Tuller et al. (2018); Zebib et al. (2020)
1	IQ < 75	Saliby et al. (2017)
1	IQ < 70	Hamdani et al. (2024)
2	Developmental/ health issues or cognitive deficits	Gagarina et al. (2019); Yılmaz Çiftci & Tunçer (2024)
4	No description of IQ-related exclusion criteria	Antonijevic-Elliott et al. (2020); Chondrogianni et al. (2018); San (2023)

The named studies included different IQ measures (e.g., Wechsler’s IQ scale or different versions of Raven’s Coloured Progressive Matrices (CPM) which further hamper the comparison with regard to investigated cognitive domains and tasks.

Previous research on multilingualism and cognitive influences on vocabulary or grammar has mainly been limited to studies on phonological working memory (Abed Ibrahim & Hamann, 2024; Weismer et al., 2017, Wimmer & Scherger, 2022), attention (Boerma et al., 2017), processing speed (Eber, 2021; Park et al., 2020) and executive functions (Bialystok et al., 2008; Boerma & Blom, 2020; Everaert et al., 2023; Giguere et al., 2022; Martin Vega, 2024; Yurtsever et al., 2023). Even though many studies use IQ as an exclusion criterion for their participants, nonverbal IQ in the absence of a biomedical condition has not yet been investigated as an influencing factor on language development disorders in multilingual children.

### Nonverbal IQ in Normative Samples

DLD is diagnosed when a child scores 1.5 SD below the mean in standardized language assessment (Bishop et al., 2017; Lüke et al., 2023). Language diagnostic tools therefore require a normative sample. It occurs that in existing diagnostic tests the samples differ largely with respect to the exclusion criteria. Loveall et al. (2022) analyzed 57 norm-referenced English language assessments for mono- and bilinguals and found out that 52 of them included individuals with a variety of disabilities and neurodevelopmental disorders to their normative samples. But ‘disability type’ was poorly defined and overall, no precise statements were made about nonverbal IQ.

The Bilingual English-Spanish Assessment, BESA (Peña et al., 2018) is an example for a diagnostic tool for multilingual children. Peña et al. (2018) do not report any exclusion criteria. As far as Germany is concerned, there are two assessment tools that take multilingual children into account in their evaluation: LiSe-DaZ (Schulz et al., 2011)

and PDSS (Kauschke et al., 2023). In the description of their normative samples, there are either no indications (Schulz et al., 2011) or the indication that cognitive impairments were excluded in advance by means of a parental questionnaire (Kauschke et al., 2022).

The relationship between DLD, multilingualism and IQ has not yet been sufficiently investigated. The question arises as to whether the exclusion criteria of a nonverbal IQ < 85 should apply to DLD in multilingual children, as in previous SLI studies.

## The Present Study

The current study aims at investigating whether nonverbal IQ has an influence on lexical and grammatical abilities for multilingual children with TLD and DLD. For this purpose, we want to examine whether IQ (operationalized as either low nonverbal IQ 70-85 [L-IQ] or average IQ > 85 [A-IQ]) has an effect on performance in grammar (LITMUS-SRT) and vocabulary (LITMUS-CLT) tests in multilingual children with TLD and DLD. Therefore, we differentiate four groups of children based on their language development and nonverbal IQ: TLD with A-IQ; TLD with L-IQ; DLD with A-IQ and DLD with L-IQ. A further aim is to see whether children with DLD underperform children with TLD on the LITMUS tools, when children with a nonverbal IQ 70 - 85 are explicitly included in the sample, as this has not been uniformly implemented in LITMUS research to date (see table 1).

The following research questions are of interest:

- I. How does nonverbal IQ influence CLT and SRT accuracy in general performance?
- II. TLD: Do children with A-IQ differ from those with L-IQ in CLT and SRT accuracy?
- III. DLD: Do children with A-IQ differ from those with L-IQ in CLT and SRT accuracy?
- IV. Do children with DLD underperform children with TLD in CLT and SRT?

In line with previous research (Bishop et al., 2017; DeThorne & Watkins, 2006; Gallinat & Spaulding, 2014; Park et al., 2020; Rice et al., 2020), a nonverbal IQ in the range between 70 and 85 should have no effect on language performance. Therefore, we would predict for research questions I, II and III:

**Hypothesis A:** IQ has no effect on children's performance in lexical-semantic and morpho-syntactic domains. Therefore, children with TLD and A-IQ and children with TLD and L-IQ perform equally high. Concerning DLD, children with A-IQ and those with L-IQ perform equally low.

**Hypothesis B:** With respect to research question IV, we expect all children with DLD to underperform TLD children.

## Methods

### Participants

This paper is part of a large-scale research project in Germany in which standard values for LITMUS tools for multilingual children are to be determined (Gagarina et al., in prep.) Ethical approval was granted by the ethics committee of the DGfS (German Linguistics Society). Although more children have been recruited in the project so far, only those children were included for whom complete background information, IQ test results and data from CLT and SRT were available. The current sample consists of 91 participants, after four children were excluded due to potential intellectual disabilities, as indicated by a nonverbal IQ more than two SD below the mean (IQ > 70). All participants were recruited in four major urban areas in Germany. The sample includes children from diverse linguistic backgrounds, with 27 different first languages spoken in addition to German. Table 2 provides an overview of the sample's demographic information, including age, length of exposure to German (LoE), the age of onset (AoO) i.e., start of contact to German and gender for the respective groups. There are no significant differences between compared groups with respect to age.

Children were classified as having DLD if a specialist (speech therapists, doctors, or special needs teachers) diagnosed them through a formal procedure. This group included those attending special language support schools or receiving speech and language therapy, as indicated in the parental questionnaire. Nonverbal IQ was assessed using Raven's Coloured Progressive Matrices (Raven's CPM, 2018). The average nonverbal IQ of the total sample is 90.07 (*SD* 11.37). The children with TLD have an average IQ of 90.57 (*SD* 10.79) and the children with DLD 89.48 (*SD* 9.61).

### Procedure

The data collection took place over two or three sessions of 30 minutes each conducted in a separate room at the participants' kindergarten or primary school. During the sessions, the nonverbal IQ was always assessed first. The LITMUS tools were displayed on a tablet, and children wore headphones to ensure audio clarity and minimize distractions. All standardized assessments were administered in strict accordance with the respective manual instructions to ensure reliability and validity of the data.

### Measures

**Background Information.** Background information about the multilingual children was obtained through a parental questionnaire, modeled on the PABIQ (Parents of Bilingual children Questionnaire, Tuller, 2015) and the Q-Bex (Quantifying Bilingual Experience, De Cat et al., 2022), which was administered to gather details about language-related background factors, including the child's exposure to and use of different languages.

Table 2: Overview of participants (mean and range in years; SD in months)

	IQ_score	Age	LoE	AoO	Gender
TLD & A-IQ <i>n</i> = 27	99.3 (8.71), 87-118	6;4 (13.56), 4;3- 7;11	4;3 (26.22), 0;8 – 7;9	2;1 (28.02), 0;0 - 7;2	m 38.5% f 61.5%
TLD & L-IQ <i>n</i> = 22	79.8 (4.75), 70-85	5;10 (17.05), 4;0 - 8;10	3;11 (15.88), 1;1 – 6;1	1;11 (18.13), 0;0 - 5;3	m 31.8% f 68.2%
DLD & A-IQ <i>n</i> = 27	95.5 (7.39), 86-110	7;3 (8.14), 6;0 - 8;7	4;11 (17.00), 2;2 – 7;5	2;3 (16.13), 0;0 - 5;6	m 70.4% f 29.6%
DLD & L-IQ <i>n</i> = 15	78.6 (5.74), 70-85	6;11 (13.39), 4;10 - 8;5	5;1 (24.43), 1;4 – 8;3	1;10 (15.73), 0;0 - 4;0	m 64.3% f 35.7%

**Raven’s 2 CPM.** CPM (Raven et al., 2018) was used to evaluate participants’ nonverbal IQ, including three test sets with a total of 36 items for the targeted age groups. This tool is designed to measure analytic reasoning skills through pattern recognition and problem-solving tasks. Children were required to select the best option to complete a pattern from a stimulus book.

**LITMUS-CLT.** Receptive and expressive vocabulary was assessed using the LITMUS-CLT, hereafter referred to as CLT. The German version (Rinker & Gagarina, 2014) contains 32 nouns and 32 verbs in the comprehension subtest and 32 nouns and 32 verbs in the production subtest. As receptive and productive vocabulary address different modalities and are assessed in different ways, we decided to treat CLT subtests for comprehension (CLT\_Rec) and for production (CLT\_Prod) as individual tests. Participants were presented with stimulus pictures on an app interface. For CLT\_Rec they were asked to select the correct item and for CLT\_Prod to name the items.

**LITMUS-SRT.** The German SRT (Hamann et al., 2013) was employed to evaluate participants’ grammatical skills. This task targets the ability to process and reproduce sentences of varying morpho-syntactic difficulty. It consists of 45 items containing three levels of difficulty concerning syntactic complexity. Participants listened to sentences presented via PowerPoint slides and were instructed to repeat them as accurately as possible. Children’s responses were rated with regard to the “Identical repetition” method.

### Statistical Analysis

Children’s test scores were treated as binary dependent variables with the levels 0 for incorrect and incomplete answers and 1 for correct answers. Subsequently, generalized linear mixed models (GLM) in R (version 4.3.3, R Core Team, 2024) have been conducted to analyze the data. The best-fitted models for each LITMUS test data were chosen when the Akaike Information Criterion (AIC) had a value < 2 compared to the most complex models. No separability was found, i.e., the parameters of the model are estimable. The variance inflation factors (VIFs) of all predictors are below 2, which indicates no problematic multicollinearity. All the final models included L-IQ/IQ-Score, DLD, age, gender and LoE as fixed factors on the dependent variable accuracy.

Items and participant were added as random factors. We calculated two regressions per LITMUS test in order to include IQ once as a categorical variable and once as a continuous variable. Lme4 package (version 1.1-27.1) and the lmerTest package (version 3.1.3) were used for GLMER implementation. After that emmeans package (version 1.10.5) was employed for each LITMUS test to conduct pairwise comparisons among the groups.

### Results

Figure 1 shows descriptive statistics for group performance on the three different tests (CLT\_Rec, CLT\_Prod and SRT): all groups perform highest on CLT\_Rec, followed by CLT\_Prod and lowest on SRT. The following sections report the results of separate models for each test in order to analyze the influence of DLD, nonverbal IQ and other factors on accuracy.

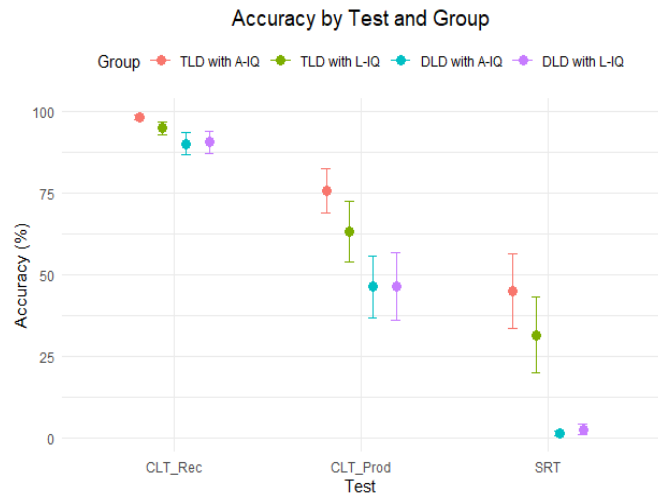


Figure 1: Accuracy by Test and Group (Mean and SD)

### Influence of Nonverbal IQ as a Categorical Variable

**CLT\_Rec.** The analysis revealed that DLD ( $\beta = -1.88, p < .001$ ) and LoE ( $\beta = 0.05, p < .001$ ) had a significant influence

on accuracy in CLT\_Rec. Whereas age, gender and low nonverbal IQ showed no such influence ( $p > .05$ ). The odds ratios (OR) showed that a higher LoE was associated with a slight increase in the probability of correct answers ( $OR = 1.05$ ), while DLD was associated with a lower probability ( $OR = 0.15$ ).

**CLT\_Prod.** As in CLT\_Rec, there is a significant influence of DLD ( $\beta = -1.28, p = .001$ ) and LoE ( $\beta = 0.05, p < .001$ ) for the accuracy in CLT\_Prod. Age, gender and low nonverbal IQ are no influencing variables in this test ( $p > .05$ ). DLD reduced the odds of correct answers ( $OR = 0.28$ ). Higher LoE values slightly increased the odds ( $OR = 1.05$ ).

**SRT.** In contrast to the CLT tests, it can be stated that in addition to DLD ( $\beta = -4.11, p < .001$ ) and LoE ( $\beta = 0.05, p < .001$ ), displaying a nonverbal IQ ( $\beta = -5.32, p = .046$ ) also has a significant influence on the accuracy of the SRT, but not age or gender. DLD significantly reduces the probability of correct repetitions ( $OR = 0.02$ ), while lower IQ further decreases this likelihood substantially ( $OR = 0.01$ ). In contrast, higher LoE values are associated with a slight increase in the odds of correct repetitions ( $OR = 1.05$ ).

### Influence of Nonverbal IQ as a Continuous Variable

**CLT\_Rec.** If IQ is analyzed as a continuous variable in a separate model, it shows a significant main effect on accuracy in CLT\_Rec ( $\beta = 0.74, p = .003$ ), with an increase in IQ score associated with a higher probability of correct answers ( $OR = 2.09$ ).

**CLT\_Prod.** For CLT\_Prod the effect is not significant ( $\beta = 0.43, p = .057$ ). A one-unit increase in the standardized IQ score is associated with an increase in the odds of correctness by a factor of 1.54.

**SRT.** The effect of IQ on accuracy in SRT is not statistically significant ( $\beta = 0.42, p = .279$ ). An increase in IQ score is associated with a higher probability of correct answers ( $OR = 1.50$ ), though this effect is not statistically significant.

### Comparison between A-IQ and L-IQ within the same language groups

We compared children with an A-IQ and children with a L-IQ within either the TLD or the DLD group on the three LITMUS tests. Table 4 shows this comparison.

**CLT\_Rec.** When contrasting the groups in terms of accuracy on CLT\_Rec, children with TLD and A-IQ differ significantly from those with TLD and L-IQ ( $p = .043$ ). Within the group of children with DLD, there is no significant difference in accuracy between those with A-IQ and L-IQ ( $p = .999$ ).

Table 4: Comparison of accuracy between groups

TLD: A-IQ vs. L-IQ			
	OR	z-value	p-value
CLT_Rec	3.22	2.63	.043*
CLT_Prod	1.80	0.89	.428
SRT	1.78	0.90	.808
DLD: A-IQ vs. L-IQ			
	OR	z-value	p-value
CLT_Rec	0.95	-0.11	.999
CLT_Prod	0.99	-0.01	1.000
SRT	0.51	-0.91	.800

**CLT\_Prod.** Children with TLD and A-IQ do not differ significantly from children with TLD and L-IQ ( $p = .428$ ) on CLT\_Prod. Similarly, children with A-IQ and L-IQ show no significant difference if they have DLD ( $p = 1.000$ ).

**SRT.** When the groups are compared in terms of accuracy in the SRT, results show that there are no significant differences between children with A-IQ and L-IQ within both groups: TLD ( $p = .808$ ) and DLD ( $p = .800$ ).

### Comparison between TLD and DLD

**CLT\_Rec.** When the DLD and TLD groups are compared in terms of accuracy CLT\_Rec and IQ is not taken into account, the DLD groups perform significantly worse than the TLD groups ( $\beta = -1.28, p < .001$ ). Participants with TLD had a higher probability of correct answers ( $OR = 3.61$ ).

**CLT\_Prod.** Children in the DLD group have a significantly lower probability of correct answers compared to the TLD group ( $\beta = -1.11, p < .001$ ). The odds of correct answers are reduced by a factor of 0.33 in the DLD group (compared to TLD).

**SRT.** Children with DLD also significantly underperform children with TLD in terms of accuracy on SRT ( $\beta = -3.74, p < .001, OR = 42$ ).

## Discussion

To answer **research question I**, we analyzed the influence of IQ on the three different tests, regardless of language status (TLD or DLD), treating IQ once as a continuous variable and once as a categorical variable. The analysis revealed that IQ seen as a continuous variable does not influence test performances except for receptive vocabulary. It may indicate that children with a lower IQ have difficulties in processing, building or recalling receptive vocabulary. This could be in line with the results of Lervag et al. (2019), who found an influence of nonverbal IQ on vocabulary growth, although the included nonverbal IQ ranges were not described in this study. However, when treating IQ as a categorical variable (L-IQ vs. A-IQ), this influence on

receptive vocabulary disappears in accordance with the results on expressive vocabulary, whereas an influence on morpho-syntactical abilities is shown. Nevertheless, this has no effect on the diagnosis of DLD, since the influence of low IQ disappears when categorizing into DLD / TLD. DLD and LoE were identified as the main influencing factors across all tests. Age and gender, on the other hand, showed no influence. The outcome that DLD has a strong influence on the accuracy in DLD diagnostic tools is unsurprising and support the validity of the tests. Furthermore, the finding that LoE is an influencing factor on the language performance confirms previous research (Ehl & Grosche, 2021; Hamann et al., 2020; Hansen et al., 2017).

To answer **research questions II and III**, the differences between average nonverbal IQ and low IQ were analyzed separately for the TLD and DLD group. Within the children with TLD, we found children with L-IQ to underperform children with A-IQ in receptive vocabulary. This finding appears to be neglectable as both show high accuracy with low variation (TLD with A-IQ:  $M = 98.40\%$  [ $SD = 0.6$ ] and TLD with L-IQ:  $M = 95.00.0\%$  [ $SD = 1.9$ ]). Since Brown (1973), in language acquisition research a linguistic structure is assumed as mastered when it is realized correctly  $>90\%$ . No differences between children with A-IQ and L-IQ could be identified in productive vocabulary and grammatical abilities in the TLD group. Within the DLD group, we found no differences between A-IQ and L-IQ in all language abilities. These results are consistent with hypothesis A and confirm that children with A-IQ and L-IQ perform equally within the TLD and DLD groups. (DeThorne & Watkins, 2006; Gallinat & Spaulding, 2014; Park et al., 2020; Rice et al., 2020).

Many studies have confirmed that children with DLD underperform children with TLD on CLT and SRT, but some excluded children with  $IQ < 85$  (see table 1). **Research question IV** therefore aims to determine whether the LITMUS tools can reliably identify children with DLD, if children with L-IQ are explicitly included in the sample. Our analysis shows that if we do not consider IQ in the comparison of children with TLD and DLD, children with DLD underperform their typically developing peers on all tests. Therefore, the findings indicate that children with DLD can be detected by the LITMUS tools regardless of having a low nonverbal IQ. This confirms our hypothesis B that we expect all children with DLD to underperform TLD children across the tests (Gagarina et al., 2019; Hamann & Abed Ibrahim, 2017).

The discrepancies in the results could be addressed in future studies by systematically varying methodological choices – such as testing different IQ categorization thresholds or applying nonlinear modeling approaches – to assess the robustness and stability of the findings. Additionally, incorporating interaction terms (e.g.,  $IQ \times DLD$ ,  $IQ \times LoE$ ) in future analyses could more accurately capture the complex interplay between cognitive abilities, language exposure, and developmental language status.

One potential methodological criticism involves our SRT rating method. Hamann and Abed Ibrahim (2017) describe two approaches: the 'identical repetition' method and the 'target structure' method. The 'identical repetition' method only ignores phonological errors, treating lexical substitutions as mistakes—which can confound vocabulary challenges in multilingual children. In contrast, the 'target structure' method evaluates mastery of complex syntactic constructions, offering deeper insights into morpho-syntactic abilities. Although the 'identical repetition' approach is quicker and widely used by speech-language pathologists for identifying children with DLD (Hamann & Abed Ibrahim, 2017), future studies should incorporate 'target structure' accuracy to better elucidate the relationship between nonverbal IQ and morpho-syntactic abilities.

Another limitation is the composition of the sample. The children show a large variation in heritage languages, which can have different crosslinguistic effects on linguistic abilities in German (Chilla et al., 2021; Lago et al., 2021). The age of the children is widely dispersed and even though age was not identified as a significant influencing variable in our analyses, linguistic abilities differ over the course of age. In addition, the group sizes differ from 15 to 27 participants, hampering comparisons. Finally, the assignment of the diagnosis of DLD must be treated with caution. As there are no diagnostic tools with reliable norm values for multilingual children in Germany, we relied on the diagnosis of specialists. Our results show that the children diagnosed with DLD by these professionals performed significantly worse on the three LITMUS tools developed to identify DLD. However, previous misdiagnoses that could have influenced the results cannot be ruled out.

## Conclusion

To summarize, children with TLD performed higher on the LITMUS tests than children with DLD, regardless of their nonverbal IQ. This supports the validity of the tests, also when children with a nonverbal IQ between 70-85 are included. The main results show that children with A-IQ vs. L-IQ do not differ across (almost) all tests within the TLD and DLD groups. Another important finding is that nonverbal IQ in the (continuous) range of 70-118 had no effect, neither on productive vocabulary nor on morpho-syntactic performance, but it did affect receptive vocabulary. When applying the limit of 1 SD below the mean in nonverbal IQ (categorical) this influence disappears in receptive vocabulary but arises in morpho-syntactical abilities. Therefore, further research might address this discrepancy. However, of all analyzed variables, language status (TLD/DLD) and LoE showed the greatest influence on LITMUS test performance, more than nonverbal IQ. Although the results of the present study cannot fully cover the complex interplay between IQ and multilingualism, the findings support the statement of the CATALISE study that children with low nonverbal ability who do not meet criteria for intellectual disability ( $IQ < 70$ ) can be included as cases of DLD (Bishop et al., 2017) – also in multilingual children.

## Acknowledgements

We would like to thank all participating caregivers, children, and our research assistants who have made this study possible. Special thanks to Tim Möller, Paul Bauten, Lilli Brokinkel, Jonna, Johanning, Anna Koban, Kristina Nagel and Marie Willenbrink for their invaluable help with data collection. We would also like to thank the project team from the ZAS in Berlin, the Goethe University in Frankfurt and the University of Eichstätt-Ingolstadt. We acknowledge financial support by the Federal Ministry of Education and Research (BMBF) for the project SPEAK (03VP11464). The main analyses were carried out with the support of the division of Statistical Consulting and Analysis Centre for Higher Education, TU Dortmund University.

## References

- Abed Ibrahim, L., & Hamann, C. (2024). German LITMUS Sentence-Repetition Task as a screening tool for the identification of SLI in bilingual children in Germany: The role of syntactic complexity and working memory. In H. A. A. AlThagafi & J. Ray (Eds.), *Proceedings of the 48th annual Boston University Conference on Language Development* (pp. 1–16). Somerville, MA: Cascadilla Press.
- Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M., & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, 34(4), 293–311. <https://doi.org/10.1080/02699206.2019.1637458>
- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. In *Communication Disorders Across Languages*. Multilingual Matters. <https://doi.org/10.21832/9781783093137>
- Armon-Lotem, S., & Marinis, T. (2015). Sentence repetition. In S. Armon-Lotem, J. De Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 95–122). Multilingual Matters. <https://doi.org/10.21832/9781783093137-007>
- Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51(6), 715–731. <https://doi.org/10.1111/1460-6984.12248>
- Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859–873. <https://doi.org/10.1037/0278-7393.34.4.859>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Boerma, T., & Blom, E. (2020). Effects of developmental language disorder and bilingualism on children's executive functioning: A longitudinal study. *Research in Developmental Disabilities*, 107, 103782. <https://doi.org/10.1016/j.ridd.2020.103782>
- Boerma, T., Leseman, P., Wijnen, F., & Blom, E. (2017). Language proficiency and sustained attention in monolingual and bilingual children with and without language impairment. *Frontiers in Psychology*, 8, 1241. <https://doi.org/10.3389/fpsyg.2017.01241>
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Bruinsma, G., Wijnen, F., & Gerrits, E. (2022). Language gains in 4–6-year-old children with developmental language disorder and the relation with language profile, severity, multilingualism, and non-verbal cognition. *International Journal of Language & Communication Disorders*, 58(5). <https://doi.org/10.1111/1460-6984.12821>
- Carvajal, H., Hayes, J. E., Miller, H. R., Wiebe, D. A., & Weaver, K. A. (1993). Comparisons of the vocabulary scores and IQs on the Wechsler Intelligence Scale for Children—III and the Peabody Picture Vocabulary Test—Revised. *Perceptual and Motor Skills*, 76(1), 28. <https://doi.org/10.2466/pms.1993.76.1.28>
- Chilla, S., Hamann, C., Prévost, P., Abed Ibrahim, L., Ferré, S., dos Santos, C., Zebib, R., & Tuller, L. (2021). The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German: A crosslinguistic approach. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Language impairment in multilingual settings* (pp. 203–231). John Benjamins. <https://doi.org/10.1075/tilar.29.08chi>
- Chilla, S., Rothweiler, M., & Babur, E. (2022). *Childhood multilingualism: Foundations – disorders – diagnostics* (3rd, fully revised ed.). Ernst Reinhardt Verlag.
- Chondrogianni, V., Butcher, M., & Garraffa, M. (2018). Supporting children with developmental language disorder in Gaelic-medium primary education. *UEA Eprints*. <https://ueaeprints.uea.ac.uk>
- De Cat, C., Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022). *Quantifying bilingual experience (Q-BEx): Questionnaire manual and documentation*. Charlottesville: Open Science Framework.
- DeThorne, L., & Watkins, R. (2006). Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics & Phonetics*, 20, 641–658. <https://doi.org/10.1080/02699200500074313>
- Ebert, K. D. (2021). Revisiting the influences of bilingualism and developmental language disorder on children's nonverbal processing speed. *Journal of Speech, Language,*

- and Hearing Research, 64(9), 3564–3570. [https://doi.org/10.1044/2021\\_JSLHR-21-00156](https://doi.org/10.1044/2021_JSLHR-21-00156)
- Ehl, B., & Grosche, M. (2020). Einbezug von Alter und Sprachkontaktdauer in die Wortschatzdiagnostik bei Mehrsprachigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 52(1–2), 41–53. <https://doi.org/10.1026/0049-8637/a000228>
- Everaert, E., Boerma, T., Selten, I., Gerrits, E., Houben, M., Vorstman, J., & Wijnen, F. (2023). Nonverbal executive functioning in relation to vocabulary and morphosyntax in preschool children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 66(10), 3954–3973. [https://doi.org/10.1044/2023\\_JSLHR-22-00732](https://doi.org/10.1044/2023_JSLHR-22-00732)
- Gagarina, N., Gey, S., & Sürmeli, N. (2019). Identifying early preschool bilinguals at risk of DLD: A composite profile of narrative and sentence repetition skills. *ZAS Papers in Linguistics*, 62, 168–189.
- Gagarina, N., Grimm, A., Rinker, T., & Scherger, A.-L. (in prep.). TEBIK 4-8: Test für bilinguale Kinder im Alter von 4 bis 8 Jahren (Test for bilingual children by age 4 to 8).
- Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 57(4), 1363–1382. [https://doi.org/10.1044/2014\\_JSLHR-L-12-0363](https://doi.org/10.1044/2014_JSLHR-L-12-0363)
- Giguere, D., Dickson, D. J., Tulloch, M. K., & Hoff, E. (2022). Majority language skill, not measures of bilingualism, predicts executive attention in bilingual children. *Journal of Experimental Child Psychology*, 213, 105256. <https://doi.org/10.1016/j.jecp.2021.105256>
- Haman E., Łuniewska M., Pomiechowska B. (2015). Designing Cross-linguistic Lexical Tasks (CLTs) for bilingual preschool children. In Armon-Lotem S., Jong J. d., Meir N. (eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment* (S. 196–240). Multilingual Matters.
- Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2, Article 16. <https://doi.org/10.3389/fcomm.2017.00016>
- Hamann, C., Chilla, S., Abed Ibrahim, L., & Fekete, I. (2020). Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Applied Psycholinguistics*, 41, 1375–1414. <https://doi.org/10.1017/S0142716420000399>
- Hamann, C., Chilla, S., Ruigendijk, E., and Abed Ibrahim, L. (2013). “A German sentence repetition task: testing bilingual Russian/German children,” in *Poster Presented at the COST Meeting in Krakow*, Oldenburg.
- Hamdani, S., Chan, A., Kan, R., Chiat, S., Gagarina, N., Haman, E., Łuniewska, M., Polišenská, K., & Armon-Lotem, S. (2024). Identifying developmental language disorder (DLD) in multilingual children: A case study tutorial. *International Journal of Speech-Language Pathology*, 1–15. <https://doi.org/10.1080/17549507.2024.2326095>
- Hansen, P., Simonsen, H. G., Łuniewska, M., & Haman, E. (2017). Validating the psycholinguistic aspects of LITMUS-CLT: Evidence from Polish and Norwegian. *Clinical Linguistics & Phonetics*, 31(11–12), 910–930. <https://doi.org/10.1080/02699206.2017.1307455>
- Harris, J. C. (2013). New terminology for mental retardation in DSM-5 and ICD-11. *Current Opinion in Psychiatry*, 26(3), 260–262.
- Kauschke, C., Dörfler, T., Sachse, S., & Siegmüller, J. (2023). *Pathologische Diagnostik bei Sprachentwicklungsstörungen* (3. Aufl.). Urban & Fischer.
- Lago, S., Mosca, M., & Stutter Garcia, A. (2021). The role of crosslinguistic influence in multilingual processing: Lexicon versus syntax. *Language Learning*, 71(S1), 163–192. <https://doi.org/10.1111/lang.12412>
- Lervåg, A., Dolean, D., Tincas, I., & Melby-Lervåg, M. (2019). Socioeconomic background, nonverbal IQ and school absence affects the development of vocabulary and reading comprehension in children living in severe poverty. *Developmental Science*, 22(5), e12858. <https://doi.org/10.1111/desc.12858>
- Loveall, S. J., Channell, M. M., Mattie, L. J., & Barkhimer, A. E. (2022). Inclusion of individuals with neurodevelopmental disorders in norm-referenced language assessments. *Frontiers in Psychology*, 13, 929433. <https://doi.org/10.3389/fpsyg.2022.929433>
- Lüke, C., Kauschke, C., Dohmen, A., Haid, A., Leitinger, C., Männel, C., Penz, T., Sachse, S., Scharff Rethfeldt, W., Spranger, J., Vogt, S., Niederberger, M., & Neumann, K. (2023). Definition and terminology of developmental language disorders—Interdisciplinary consensus across German-speaking countries. *PLOS ONE*, 18(11), e0293736. <https://doi.org/10.1371/journal.pone.0293736>
- Martin Vega, S. (2024). *Lexical processing and executive function in bilingual children with and without language disorder* (Doctoral thesis). University of Essex.
- Meir, N., Walters, J., & Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: The impact of L1 and L2 properties. *International Journal of Bilingualism*, 20(4), 421–452. <https://doi.org/10.1177/1367006915609240>
- Park, J. S., Miller, C. A., Sanjeevan, T., Van Hell, J. G., Weiss, D. J., & Mainela-Arnold, E. (2020). Bilingualism and processing speed in typically developing children and children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(5), 1479–1493. [https://doi.org/10.1044/2020\\_JSLHR-19-00403](https://doi.org/10.1044/2020_JSLHR-19-00403)
- Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English-Spanish Assessment™ (BESA™) Manual*. Brookes Publishing.
- Plante, E. (1998). Criteria for SLI: The Stark and Tallal Legacy and Beyond. *Journal of Speech, Language, and Hearing Research*, 41(4), 951–957. <https://doi.org/10.1044/jslhr.4104.951>

- Raven J., Rust J., Chan F., Zhou X. (2018). *Raven's 2 progressive matrices, clinical edition (Raven's 2)*. Pearson.
- Rice, M. L. (2016). Specific language impairment, nonverbal IQ, attention-deficit/hyperactivity disorder, autism spectrum disorder, cochlear implants, bilingualism, and dialectal variants: Defining the boundaries, clarifying clinical conditions, and sorting out causes. *Journal of Speech, Language, and Hearing Research*, 59(1), 122–132. [https://doi.org/10.1044/2015\\_JSLHR-L-15-0255](https://doi.org/10.1044/2015_JSLHR-L-15-0255)
- Rice, M. L., Taylor, C. L., Zubrick, S. R., Hoffman, L., & Earnest, K. K. (2020). Heritability of specific language impairment and nonspecific language impairment at ages 4 and 6 years across phenotypes of speech, language, and nonverbal cognition. *Journal of Speech, Language, and Hearing Research*, 63(3), 793–813. [https://doi.org/10.1044/2019\\_JSLHR-19-00012](https://doi.org/10.1044/2019_JSLHR-19-00012)
- Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research*, 47(4), 816–834. [https://doi.org/10.1044/1092-4388\(2004\)061](https://doi.org/10.1044/1092-4388(2004)061)
- Rinker, T. & Gagarina, N. (2014). *CLT – Crosslinguistic Lexical Task – German Version*. Universität Konstanz & ZAS Berlin.
- Saar, V., Komulainen, E., & Levänen, S. (2022). The significance of nonverbal performance in children with developmental language disorder. *Child Neuropsychology*, 29, 213–234. <https://doi.org/10.1080/09297049.2022.2077324>
- Saliby, C. K. A., dos Santos, C., Hreich, E. K., & Messarra, C. (2017). Assessing Lebanese bilingual children: The use of Cross-linguistic Lexical Tasks in Lebanese Arabic. *Clinical Linguistics & Phonetics*, 31(11–12), 874–892. <http://dx.doi.org/10.1080/02699206.2017.1308554>
- San, N. H. (2023). Subordination in Turkish heritage children with and without developmental language impairment. *Languages*, 8(4), 239. <https://doi.org/10.3390/languages8040239>
- Schulz, P., Tracy, R., Esser, G., Hasselhorn, M., & Schneider, W. (Hrsg.). (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache* (1. Aufl.). Hogrefe.
- Simonsen, H. G., & Haman, E. (2017). LITMUS-CLT: A new way to assess bilingual lexicons. *Clinical Linguistics & Phonetics*, 31(11–12), 811–817. <https://doi.org/10.1080/02699206.2017.1307454>
- Smith, B. L., Smith, T. D., Taylor, L., & Hobby, M. (2005). Relationship between intelligence and vocabulary. *Perceptual and Motor Skills*, 100(1), 101–108. <https://doi.org/10.2466/pms.100.1.101-108>
- Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, 46, 114–122.
- Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. De Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 331–358). Multilingual Matters. <https://doi.org/10.21832/9781783093137-014>
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Multilingual Matters. <https://doi.org/10.21832/9781783093137-013>
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, É., Prévost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53(4), 888–904. <https://doi.org/10.1111/1460-6984.12398>
- Weismer, S. E., Davidson, M. M., Gangopadhyay, I., Sindberg, H., Roebuck, H., & Kaushanskaya, M. (2017). The role of nonverbal working memory in morphosyntactic processing by children with specific language impairment and autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, 9(1), 28. <https://doi.org/10.1186/s11689-017-9209-6>
- Wimmer, E., & Scherger, A.-L. (2022). Working memory skills in DLD: Does bilingualism make a difference? *Languages*, 7(4), 287. <https://doi.org/10.3390/languages7040287>
- World Health Organization. (2022). *International classification of diseases for mortality and morbidity statistics* (11th ed.). <https://icd.who.int>
- Yılmaz Çiftçi, N., & Tunçer, A. M. (2024). Testing validity and reliability of Cross-linguistic Lexical Task—Turkish (CLT-TR) in preschoolers with and without risk of language disorder. *Arch Health Sci Res*, 11, 148–154. <https://doi.org/10.5152/ArcHealthSciRes.2024.23106>
- Yurtsever, A., Anderson, J. A. E., & Grundy, J. G. (2023). Bilingual children outperform monolingual children on executive function tasks far more often than chance: An updated quantitative analysis. *Developmental Review*, 69, 101084. <https://doi.org/10.1016/j.dr.2023.101084>
- Zebib, R., Tuller, L., Hamann, C., Abed Ibrahim, L., & Prévost, P. (2020). Syntactic complexity and verbal working memory in bilingual children with and without developmental language disorder. *First Language*, 40(4), 461–484. <https://doi.org/10.1177/0142723719888372>