Testing the continuum/spectrum model in Russian-speaking children with and without developmental language disorder

Militina Gomozova
Center for Language and Brain, HSE University, Moscow, Russia
ORCID: 0000-0003-4976-9868, email: militina.gomozova@ya.ru

Valeriia Lezzhova
Center for Language and Brain, HSE University, Moscow, Russia
ORCID: 0000-0003-1815-6540, email: val.lezzhova@gmail.com

Olga Dragoy
Center for Language and Brain, HSE University, Moscow, Russia
Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia
ORCID: 0000-0002-6777-5164, email: odragoy@hse.ru

Anastasiya Lopukhina
Royal Holloway, University of London, Egham, UK
Center for Language and Brain, HSE University, Moscow, Russia
ORCID: 0000-0002-2283-1644, email: nastya.lopukhina@gmail.com

Keywords: developmental language disorder, comprehensive language assessment, continuum model, spectrum model, Russian-speaking children

Abstract

Purpose: Previously, Lancaster and Camarata (2019) showed that the continuum/spectrum model of the developmental language disorder (DLD) best explained the high heterogeneity of symptoms in children with DLD. We hypothesize that the continuum/spectrum approach can include not only children with DLD, but also typically developing children with different timelines and patterns of language acquisition. This model can explain individual language profiles and deficits in children.

Method: We assessed language abilities in a group of Russian-speaking children with DLD aged 4-to-7 years (N=53) and their age- and gender-matched peers without speech and language diagnoses (N=53, TD). We evaluated the children's performance at four language levels in production and comprehension domains, using 11 subtests of the standardized language assessment for Russian – RuCLAB. Using the k-means cluster method and RuCLAB scores, we obtained two clusters of children and analyzed their language performance in individual subtests.

Results: The analysis revealed that the two clusters of children both included DLD and TD participants: Group 1, with higher test scores (TD=45, DLD=24 children) and Group 2, with lower scores (TD=8, DLD=29). Children from Group 1 mostly had lower scores at one of the language levels, whereas those from Group 2 struggled at several language levels. Furthermore, children with DLD from both Groups tended to be more sensitive to linguistic features like word length, noun case, sentence reversibility compared to TD children.

Conclusions: The presence of two mixed groups shows that children with diagnosed DLD could perform on par with TD children, whereas some younger TD children could perform similarly to children with DLD. Our findings support the continuum/spectrum model: linguistic skills in pre-school children is a continuum, varying from high to poor skills at all language levels in comprehension and production. To describe a child's language profile, the tasks assessing all language levels should be used.

Introduction

Epidemiological studies revealed that language impairments in children ranged from 3% to 7%, depending on definition and children's age (Norbury et al., 2016; Tomblin, Records et al., 1997; Weindrich et al., 2000). Children with impaired language development are often diagnosed as having developmental language disorder (DLD; e.g. Law et al., 2000; McLeod & McKinnon, 2007; Van Borsel et al., 2006). DLD is a well-studied impairment: its symptoms, diagnostic criteria, and treatment were widely described in previous literature (Bishop, 2014; Botting et al., 2001; Conti-Ramsden, 2003; Conti-Ramsden & Botting, 1999). However, since children have their individual pace in acquiring phonological, lexical, grammatical, and discourse skills, it may be problematic to distinguish between typical and atypical language development. There are many questions about patterns of language symptoms in DLD and relevant inclusion criteria that limit our understanding of DLD and its distinction from typical development.

DLD is characterized by difficulties in language acquisition, despite the normal intelligence level, hearing, and vision (Bishop et al., 2016, 2017; Leonard, 2008; Reilly et al., 2014). According to the Diagnostic and Statistical Manual of mental disorders (DSM-5), DLD should be presumed if a child has difficulties in language comprehension and production at the lexical, syntactic, and/or discourse levels. However, lexical, syntactic, and discourse impairments associated with DLD are described in the DSM-5 superficially: there is no sufficient information about possible patterns of impairment at these language levels.

Language impairments in children with DLD were described in more detail in experimental studies. Some results were in line with the DSM-5 criteria. For example, Hannus et al. (2013) showed that children with DLD had impairments at lexical, morphosyntactic (mostly with past tense marking and verb morphology), and discourse levels, and that with individual participants, each of the language levels might be impaired to greater or lesser degree. Other studies showed controversial results. For example, Marini et al. (2008) found that Italian-speaking children with DLD were more likely to have impairments at the morphosyntactic level, and less likely – at the lexical level. Several cross-linguistic studies confirmed that children with DLD struggle not only at lexical, morphosyntactic, and discourse levels (as noted in DSM-5), but also at the phonological level (e.g., Marshall & van der Lely, 2007 for English-speaking children; Roberts & Leonard, 1997 for German-speaking children; Bedore & Leonard, 2001 for Spanish-speaking children). Impairments at the phonological level are sometimes referred to as the core deficit that underlies impairments at the morphosyntactic level (i.e. phonological approach, Joanisse & Seidenberg, 1998). Thus, there is a discrepancy between the clinical diagnostic criteria for DLD in DSM-5 and the DLD symptoms that are described in experimental studies.

Due to the vagueness of diagnostic criteria and the variety of symptoms in children with DLD, there have been debates about DLD inclusion standards. For example, in the commentaries to Bishop et al. (2014) and Reilly et al. (2014) reviews on DLD, the debaters indicated that DSM-5, as a biomedically-derived diagnostic system, is widely ignored by clinicians (Hansson et al., 2014; Rice, 2014; Rutter, 2014; Snowling, 2014). Other studies proposed three main arguments against the validity of DSM-5. First, DSM-5 does not explain the possible diversity of language profiles in individuals with DLD (Bishop & Snowling, 2004; Leonard, 2014).

Second, it is unclear at what age, speech therapists should screen children to allocate individuals with language disorders, because the age of onset of language impairment symptoms can differ. For example, Ellis and Thal (2008) analyzed later language outcomes for children classified at 16 months into three groups: typically developing children; “late producers” (children with expressive delay only); and “late comprehenders” (children with receptive and expressive delays). The authors found that at the age of 6, some of the children from all 3 groups met the criteria for DLD. Of the 13 children diagnosed with DLD at the age of 6, 7 children came from typically developing group (1.5%), 3 – from late producers (3.7%), and 3 – from late comprehenders (8.5%). These results highlight the idea of variability in language development and the complexity of DLD identification.

Third, the validity of DSM-5 may be questioned because there is evidence that some children with symptoms of language impairment can catch up with their typically developing peers without special help at later stages of language development (e.g. Bishop et al., 2016 for the group of children with communicative deprivation). Also, as found in the study by Ellis and Thal (2008), some children may show DLD symptoms at the late stages of language development. In sum, there is a lot of confusion about how and when children should be diagnosed with DLD and what profiles of DLD should be identified.

The possible approach to embrace the variety of symptoms in the DLD population is a continuum/spectrum model of the impairment that was proposed by Lancaster and Camarata (2019). They used the results of the performance of 505 children with DLD in three tests: Test of Language Development - Primary 2nd Edition (Newcomer & Hammill, 2008), narrative task (Culatta et al., 1983) and Wechsler Preschool and Primary Scales of Intelligence - Revised (Wechsler, 1989). These results were then clustered to compare three possible models of DLD – the subgroup model, the individual differences model, and the continuum/spectrum model. The cluster analysis revealed 18 clusters of the participants. At this stage, the authors rejected the subgroup model. The further analyses of data randomness allowed the authors to reject the individual differences model. Thus, the continuum/spectrum model was the only model that was able to explain the high variability of performance in children with DLD. Lancaster and Camarata suggested that the continuum/spectrum approach should encourage clinicians to pay attention to the severity of symptoms, rather than to subgrouping.

We hypothesize that the continuum/spectrum approach to language performance can be extended not only to children with DLD, but also to children who have different timelines and patterns of language acquisition. This includes children at the lower border of normal development, e.g. late talkers, children with mild social deprivation, and children whose performance in language tests is below average. These children did not get much consideration in the experimental literature (Bishop et al., 2016). The advantage of expanding the continuum/spectrum model to a wider range of children is that the proposed approach can help us describe different patterns of language development. Furthermore, this model takes into account the severity of symptoms and may help the practitioners allocate children with individual weaknesses and needs for intervention that would not be diagnosed with DLD by the classical DSM-5 approach.

Our study aimed to test the continuum/spectrum model in Russian-speaking children with and without DLD diagnosis. We assessed language abilities in a group of 4-to-7-year-old children with DLD and their age- and gender-matched peers without diagnosed speech and language impairments. We evaluated children’s performance at all language levels in production and comprehension domains and we used clustering analysis that did not require previous subgrouping. Such analysis provided us with the opportunity to detect and describe language profiles in Russian-speaking young children.

The novelty of the present study is threefold: first, we tested the continuum/spectrum model simultaneously in children with DLD and in typically developing children. Second, we analyzed which linguistic parameters can affect children’s performance in language tasks. We compared typically developing children and children with DLD who had higher and lower language performance (as defined by the cluster analysis). This analysis allowed us to identify language levels and specific linguistic features that could be impaired in children with DLD. Third, we described patterns of language deficits in individual children. Such a description contributed to the discussion about the heterogeneity of symptoms in children with DLD and allowed us to explain the controversial results from previous studies. We suppose that the results of our study can be used as evidence in a theoretical discussion of language impairments in children with DLD and also can be useful for speech/language therapists.

Method

Participants

A total of 106 native Russian-speaking 4-7-year-old children participated in the study: 53 children with DLD and 53 age and gender matched typically developing (TD) children (M=5.5; SD=0.93; 70 boys, 36 girls). The information about the number of children in each age group is provided in Table 1. Children with DLD were recruited from speech therapy groups in public kindergartens in Moscow, Nizhniy Novgorod, Sochi, and Saratov. Typically developing children were recruited from the same kindergartens. The parents or primary caretakers of all participants signed an informed consent form before the start of the study. The study was approved by the HSE Committee on Interuniversity Surveys and Ethical Assessment of Empirical Research and conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013).

Table 1. Age groups of the participants (both DLD and TD)

Age group	Boys (N)	Girls (N)
4-year-old	16	8
5-year-old	26	24
6-year-old	18	2
7-year-old	10	2

To clarify the diagnostic procedure, DLD is not used as an official diagnosis in speech and language pathologists' practice in Russia. For our study, we selected children with a set of diagnoses which were used to identify children with delayed language and speech acquisition (Kornilov et al., 2016). These diagnoses are usually given to children starting from 3 years of age and include General Speech Underdevelopment (GSU; общее недоразвитие речи/ОНР); Phonetic-phonemic Underdevelopment (PPU; фонетико-фонематическое недоразвитие речи/ФФНР); Delayed Speech Development (DSD; задержка речевого развития/ЗРР). In our sample of the participants with DLD, two participants had PPU and DSD diagnosis, and 51 children were diagnosed as having GSU. The diagnoses were made by a committee that consists of a neurologist, a speech therapist, and a psychologist. The committee examined a child based on the diagnostic principles described in "Fundamentals of childhood speech pathology: clinical and psychological aspects" by Kornev (2006). Such complex observation provides an opportunity to qualitatively assess language abilities, but this diagnostic system does not include standardized language assessment.

In our study, exclusion criteria were any persistence of comorbid hearing impairment, uncorrected vision problems and a low level of non-verbal intelligence (NVI). All the participants were tested with Raven's Colored Progressive Matrices (Raven, 2000) to evaluate the NVI level. The results were within the normative range of NVI in both groups of the participants.

Materials and Procedure

All children were tested with a new standardized Russian Child Language Assessment Battery (RuCLAB) (Arutiunian et al., 2022) that consists of 11 subtests and assesses language development on phonological, lexical, morphosyntactic, and discourse levels in comprehension and production domains. The description of the subtests and its properties are provided in Table 2. All real words and visual stimuli were selected from the Verbs and Nouns Stimuli Database for Russian (Akinina et al., 2014, 2015, 2016) with name agreement for pictures above 85%. The remaining pictures of situations were drawn by the same artist. All audio stimuli were recorded by a female native speaker of Russian.

The assessment was performed using a Samsung Galaxy A (2018) SM-T595 tablet with a screen size 10.1", 1920 × 1200 px. The stimuli for all subtests were presented in the AutoRAT application (Ivanova et al., 2016).

The children were tested individually in a quiet room in a kindergarten. Each subtest was preceded by instruction and two or three training trials which were excluded from the analysis. The children had the opportunity to ask questions before each subtest and during the practice trials, but they did not receive any feedback during the testing. The order of subtests and items were the same for all participants. The assessment lasted for 40-60 minutes depending on the children's age and including time for breaks between the subtests.

All subtests that assess comprehension domains were scored as correct or not automatically in the application. For all production subtests, the vocal responses of the participants were recorded in the application, then analyzed and scored by the authors of the paper according to the instruction.

Table 2. The description of RuCLAB subtests

Language	Subtest	Task,	Controlled	Scoring

level / Domain		example, number of items	linguistic features	
Phonology / Comprehension	Phonological discrimination	Listen to pairs of nonwords and identify whether they are same or different. For example, *ga**sk**a-ga**ks**a* and *ponka-ponka*. *N* = 24.	Length in phonemes; Place of contrasting sound; Type of contrasting sound	Scored automatically: correct answer – 1; incorrect answer – 0
Phonology / Production	Nonword repetition	Listen to nonwords and repeat them. For example, *bunitsa*. *N* = 24.	Word length (1, 2, or 3 syllables); Number of articulatory switches	Accurate repetition – 1; repetition with phonological paraphasias when 50% of word was preserved – 0.5; repetition with phonological paraphasias with less than 50% of word preserved – 0
Lexicon / Comprehension	Noun comprehension	Listen to a noun and choose a	Frequency; Length; Age of	Scored automatically: correct answer

		corresponding picture out of four (e.g. *babochka* 'butterfly'). Three distractors are a phonologically similar word (e.g. *babushka* 'grandmother'), a semantically similar word (e.g. *muha* 'fly') and an irrelevant word (e.g. *dedushka* 'grandfather'). *N* = 24.	acquisition (matched with the other three subtests that assess lexicon)	– 1; incorrect answer – 0
Lexicon / Comprehension	Verb comprehension	Listen to a verb and choose a corresponding picture out of four (e.g. pet' 'to sing' – target; *pit'* 'to drink' – phonological distractor;	Frequency; Length; Age of acquisition	Scored automatically: correct answer – 1; incorrect answer – 0

		tancevat' 'to dance' – semantic distractor; *nalivat'* 'to pour' – irrelevant distractor). *N* = 24.		
Lexicon / Production	Object naming	Name a depicted object (e.g. *lodka* 'boat'). *N* = 24.	Frequency; Length; Age of acquisition	Response matches the target word or a word from the list of possible nominations – 1; incorrect response – 0. Types of errors: phonological, semantical, mixed, neologism, circumlocution, other
Lexicon / Production	Action naming	Name a depicted action (risovat' 'to draw'). *N* = 24.	Frequency; Length; Age of acquisition	Same as in Object naming

Morphosyntax / Comprehension	Sentence comprehension	Listen to a sentence and choose a corresponding picture out of two (e.g. "a cat is chasing a dog" – target; "a dog is chasing a cat" – distractor). *N* = 24.	Word order (SVO/OVS); Type of third argument (Prepositional or Instrumental); Number of verb arguments (1, 2, or 3); Type of the construction (simple constructions, relative clauses, reflexive constructions, constructions with prepositions).	Scored automatically: correct answer – 1; incorrect answer – 0
Morphosyntax / Production	Sentence production	Describe a picture using the syntactic construction of the priming sentence (priming sentence "A grandfather is cleaning a	Type of third argument (Prepositional or Instrumental); Number of verb arguments (1, 2, or 3); Semantic	Four parameters: priming, grammar, lexicon, other; 0.25 for each parameter

		carpet with a brush”; target sentence “A girl is writing a letter with a pencil”). *N* = 24.	reversibility (reversible or nonreversible).	
Morphosyntax / Production	Sentence repetition	Listen to a sentence and repeat it (e.g. “Pandas eat bamboo”). N = 12.	Sentence length (3 to 9 words); Word frequency (low or high).	Each word repeated correctly – 1. Incorrect word order was penalized with -1. The numbers were summarized and divided by the number of words in a sentence.
Discourse / Comprehension	Discourse comprehension	Listen to a story and answer true/false comprehension questions. *N* = 16. The questions are paired so that each pair is related to one fact mentioned	Line of story (main information or details); Type of information (explicit or implicit).	Scored automatically: correct answer in both paired questions – 1; incorrect answer in any or both paired questions – 0.

		in the story.		
Discourse / Production	Discourse production	Produce a story based on the presented picture with exposition, climax, and resolution. *N* = 1.	-	Four parameters: coherence, grammar, lexicon, and fluency; 0.25 for each parameter

Analysis

First, in Section 1, we ran a series of (G)LMMs and simple LMs to compare whether the performance of children with DLD differs from TD children (Baayen, Davidson, & Bates, 2008). In Section 2, we performed a cluster analysis to analyze the performance of the participants regardless of their group. Then, in Section 3, we ran a series of (G)LMMs and simple LMs to explore the influence of various linguistic features on childrens' performance. We compared the performance of TD children with the performance of the two groups of children with DLD, as revealed by the cluster analysis. Finally, in Section 4, we described patterns of errors in individual children.

In Sections 1 and 3, the analyses were conducted using the R statistical computing program (R Core Team, 2020). The models were estimated with the 'lme4' package (Bates et al., 2015) and the tables for models' outcomes were created with the 'sjPlot' package (Lüdecke, 2017). In the models, fixed effects included the participants' group (coded with treatment contrasts with 0 corresponding to TD children) and the linguistic parameters of the subtests, each nested within the participants' group (nested contrasts; Schad et al., 2020; in Section 3). The (G)LMM models included varying intercepts for participants and items. Significant effects were adjusted for multiple comparisons using the Bonferroni correction at an α-level of 0.005.

In Section 2, we applied the k-means clustering method with the participants' results on each subtest of the RuCLAB as grouping variables. We applied this analysis five

times: to all 11 tests, to the phonological level (2 subtests), to the lexical level (4 subtests), to the morphosyntactic level (3 subtests), and to the discourse level (2 subtests). The number of optimal clusters, *k,* for each analysis was defined using the average silhouette method in Python. This method uses a metric of how well each observation fits within a cluster; a higher value of the metric indicates a better clustering solution (Kaufman & Rousseeuw, 1990).
The data, analysis scripts, and tables for this study are freely available at the Open Science Framework project page: https://osf.io/6y23k/?view_only=ae460def79c9476fa8188b005341bd47.

Results

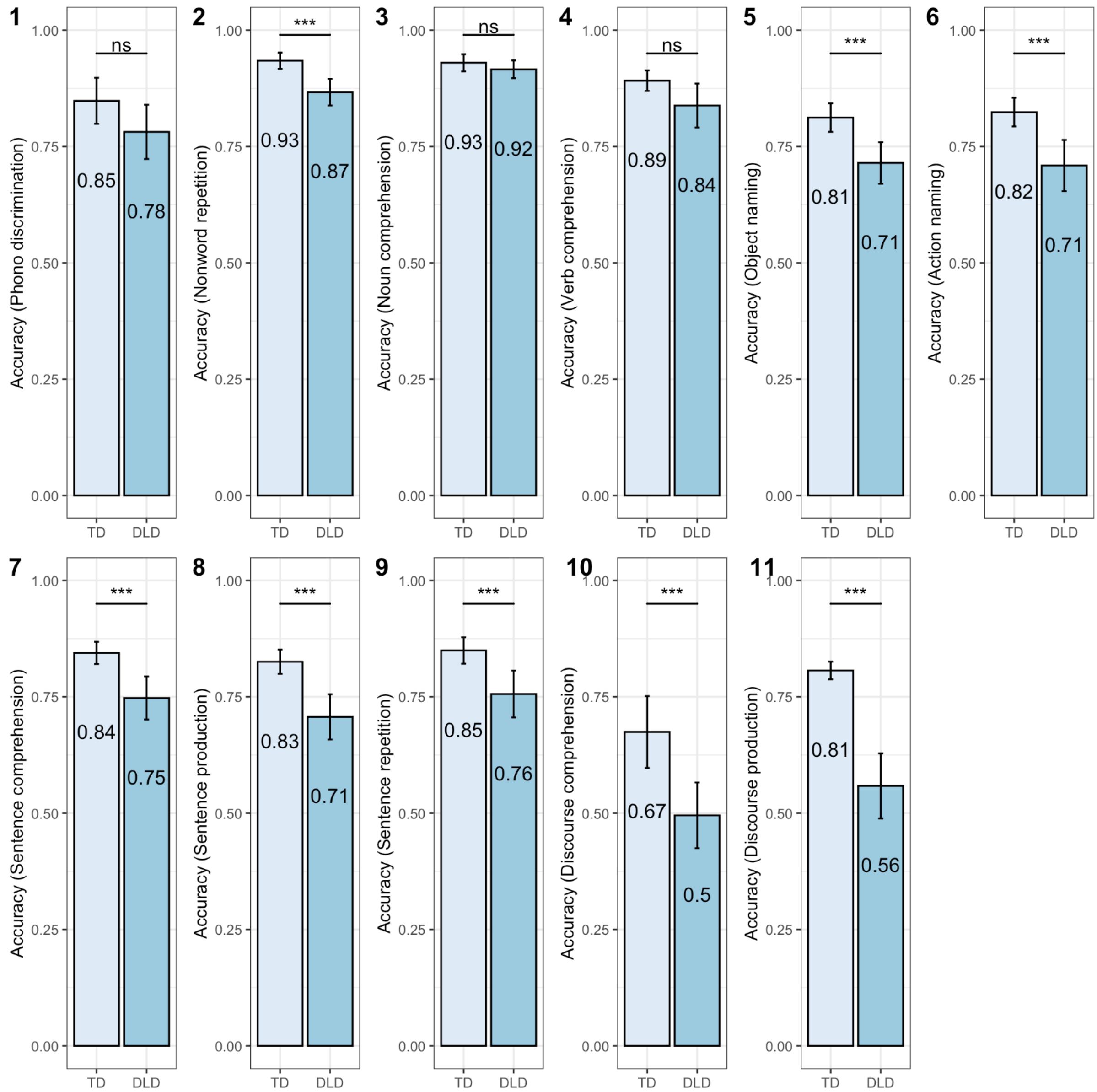
Comparing TD and DLD groups

All TD children completed all assessment subtests. However, four children with DLD were unable to complete at least one subtest, including Phonological discrimination, Sentence comprehension, Object naming, Sentence production, and Sentence repetition. Also, eight children with DLD did not complete the Discourse production subtest.

For the first round of analyses, we estimated the proportions of correct responses across the two groups of participants (TD vs. DLD). Figure 1 shows the corresponding descriptive statistics.

As expected, we found that children with DLD made more errors in Nonword repetition (β = -0.07, *SE* = 0.02, $p < 0.001$), Object naming (β = -0.77, *SE* = 0.21, $p < 0.001$), Action naming (β = -0.80, *SE* = 0.21, $p < 0.001$), Sentence comprehension (β = -0.71, *SE* = 0.19, $p < 0.001$), Sentence production (β = -0.12, *SE* = 0.03, $p < 0.001$), Sentence repetition (β = -0.09, *SE* = 0.03, $p < 0.001$), Discourse comprehension (β = -1.06, *SE* = 0.30, $p < 0.001$), and Discourse production (β = -0.25, *SE* = 0.04, $p < 0.001$) subtests compared to TD children. Notably, we did not observe significant differences in accuracy between TD and DLD children in Phonological discrimination (β = -0.67, *SE* = 0.35, p = 0.055), Noun comprehension (β = -0.24, *SE* = 0.20, p = 0.24), and Verb comprehension (β = -0.49, *SE* = 0.23, p = 0.03, non-significant after Bonferroni correction) subtests. The absence of the difference in word comprehension tasks is in line with the study by Marini et al. (2008). In the Phonological discrimination subtest, the absence of the difference between the two groups, and the high variability of scores within each group can be explained by the nature of the task: younger children struggled to keep in mind the instruction and decide whether the two sequences of sounds were the same or different. Overall, at the group level, children with DLD made significantly more errors in production subtests and in more effortful comprehension subtests that assessed their morphosyntactic and discourse skills.

Figure 1. Mean accuracy and confidence intervals for TD children and children with DLD across all subtests.



Cluster analysis

The silhouette scores method of cluster number selection indicated that the optimal number of clusters was two for the 11 subtests taken together, as well as for lexical and morphosyntactic language levels. The optimal number of clusters was three for phonological and discourse levels.

The two-cluster solution for all 11 subtests divided all children into two groups (Group 1, higher scores - 69 children) and (Group 2, lower scores - 37 children) of participants. The descriptive statistics of the two clusters are presented in Table 3, along with the information on the children's group (TD vs. DLD) and age. The data distribution for all 11 subtests is illustrated in Figure 2 below and for phonological, lexical, morphosyntactic, and discourse levels – in Figure S2 in Supplementary Materials. Welch t-tests for independent samples indicated significant differences in Group 1 and Group 2 average scores at all language levels: phonological (t = 6.32, p < 0.001), lexical (t = 8.17, p < 0.001), morphosyntactic (t = 9.57, p < 0.001), discourse (t = 6.32, p < 0.001).

Figure 2. Distribution of accuracy scores for all participants across all 11 subtests. The dimensions are reduced from 11 to 2 for plotting.

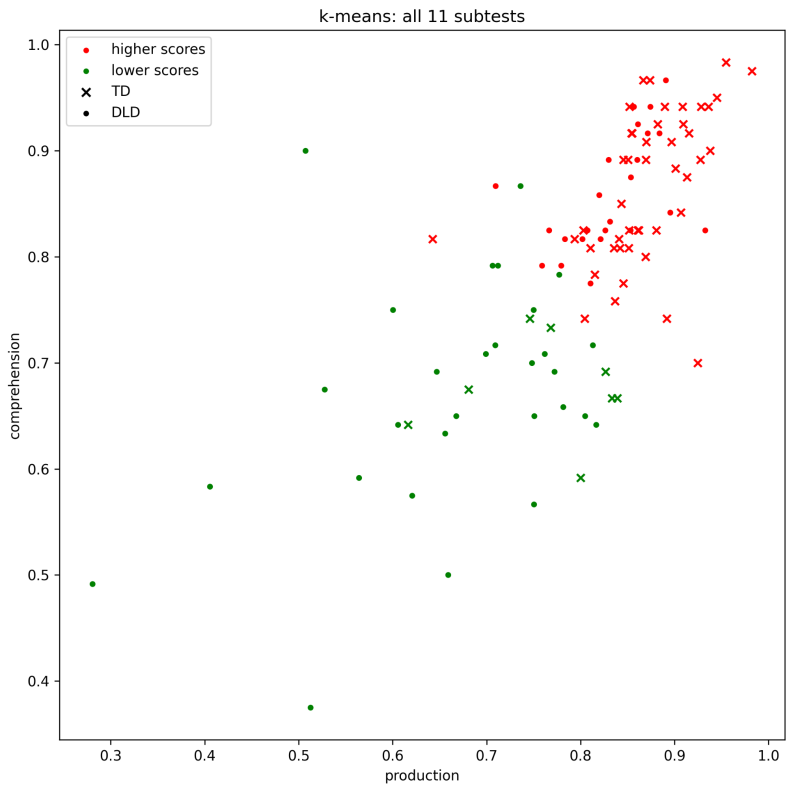


Table 3. The participants' performance is assorted by cluster (Group 1 and Group 2), clinical diagnosis (TD and DLD), and language level.

Cluster	Group	Distribution by age	Language level	Clustering within language level	Mean accuracy (SD)
Group 1 (*N*=69)	TD (*N*=45)	4 y.o.: 6 (50%) 5 y.o.: 23 (92%) 6 y.o.: 10 (100%) 7 y.o.: 6 (100%)	Phonology	**cluster 1: 80%** cluster 2: 18% cluster 3: 2%	0.91 (0.13)
			Lexicon	**cluster 1: 98%** cluster 2: 2%	0.88 (0.08)
			Morphosyntax	**cluster 1: 98%** cluster 2: 2%	0.85 (0.09)
			Discourse	**cluster 1: 78%** cluster 2: 22% cluster 3: 0	0.78 (0.17)
	DLD (*N*=24)	4 y.o.: 2 (17%) 5 y.o.: 12 (48%) 6 y.o.: 6 (60%) 7 y.o.: 4 (66%)	Phonology	**cluster 1: 75%** cluster 2: 25% cluster 3: 0	0.90 (0.09)
			Lexicon	**cluster 1: 96%** cluster 2: 4%	0.86 (0.12)

			Morphosyntax	**cluster 1: 100%** cluster 2: 0	0.84 (0.09)
			Discourse	**cluster 1: 87%** cluster 2: 13% cluster 3: 0	0.70 (0.12)
Group 2 (*N*=37)	TD (*N*=8)	4 y.o.: 6 (50%) 5 y.o.: 2 (8%)	Phonology	cluster 1: 12% **cluster 2: 88%** cluster 3: 0	0.78 (0.12)
			Lexicon	cluster 1: 50% cluster 2: 50%	0.75 (0.13)
			Morphosyntax	**cluster 1: 75%** cluster 2: 25%	0.77 (0.09)
			Discourse	cluster 1: 0 **cluster 2: 100%** cluster 3: 0	0.51 (0.28)
	DLD (*N*=29)	4 y.o.: 10 (83%) 5 y.o.: 13 (52%) 6 y.o.: 4 (40%) 7 y.o.: 2 (34%)	Phonology	cluster 1: 31% **cluster 2: 62%** cluster 3: 7%	0.76 (0.19)
			Lexicon	cluster 1: 48% cluster 2: 52%	0.74 (0.20)
			Morphosyntax	cluster 1: 41% **cluster 2: 59%**	0.65 (0.19)
			Discourse	cluster 1: 10% **cluster 2: 62%** cluster 3: 28%	0.38 (0.25)

Note. In this Table, we aggregated the results of all cluster analyses. Group 1 is the cluster with higher scores across all subtests, Group 2 is the cluster with lower overall scores. Cluster 1 within language level indicates clusters of participants with higher scores at a particular language level. Clusters 2 and 3 indicate clusters of participants with lower scores. The clusters with the majority of participants at the language level are in bold.

Most of the typically-developing children were in Group 1, whereas children with DLD were evenly distributed between the two Groups. Furthermore, the distribution between the Groups seems to be related to the participants' age: 67% of 4-year-old children were in Group 2 (i.e. 16 out of 24, TD and DLD together), and 5-to-7-year-old children were more likely to be in Group 1.

Within each of the four language levels, the children from Group 1 were assigned to clusters with the best performance. In contrast, the children from Group 2 were assigned to clusters with lower performance in phonology and discourse and were evenly distributed between ‘higher’ and ‘lower’ lexical clusters. Furthermore, in Group 2, TD children had higher scores at the morphosyntactic level compared to children with DLD (77% vs. 65% accuracy).

A deeper individual qualitative analysis of the performance of children in Group 1 showed that 15 participants with DLD were assigned to clusters with the best performance at all language levels (seven 5-year-old children, five 6-year-old children, and three 7-year-old children). Based on the cluster analysis, these participants' performance in language tests did not differ from the performance of TD children. In Group 2, ten children with DLD were assigned to clusters with worse performance at all language levels (four 4-year-old children, five 5-year-old children, and one 6-year-old child). Also, two 4-year-old TD children in Group 2 had lower scores at all language levels. For all the other children in Groups 1 and 2 (74% of the participants), the assignment to ‘higher’ or ‘lower’ clusters at each of the language levels varied without any special pattern.

Participants’ profiles across language levels

To describe language profiles in children with higher and lower performance in language subtests, we compared children with DLD from Group 1 (DLD-1, *N*=24) and children with DLD from Group 2 (DLD-2, *N*=29) with TD children from Group 1 (TD-1, N=45). We didn’t include TD children from Group 2 because of the small number of participants, *N*=8. Crucially, in this analysis, we explored the influence of linguistic features, e.g. word length, frequency, word class, word order, etc., in each language subtest on accuracy scores within each of the three groups of participants (nested effects). The key findings of the analysis are summarized in Table 4. Detailed summaries of model fits can be found in Supplementary Materials (Tables S1 - S5).

The comparison of DLD-1 to TD-1 children did not show any difference between the two groups after the Bonferroni correction to $p = .005$ (the only significant difference between TD-1 and DLD-1 was in the number of grammatical errors in the Discourse production subtest). The results before the Bonferroni correction showed that DLD-1 tended to be less accurate in Object and Action naming and in Discourse production subtests.

In contrast to this, the children from the DLD-2 group had lower accuracy compared to TD-1 children in Noun and Verb comprehension; Sentence comprehension; Discourse comprehension; Object and Action naming; Sentence production; and Discourse production subtests (all *p*'s < 0.003). They also tended to have lower scores in the Phonological discrimination subtest ($p = 0.02$). Interestingly, the children from the DLD-2 group did not differ significantly from TD-1 children in the Nonword repetition and Sentence repetition subtests (*p*'s = 0.07).

We also found that the children with DLD from both Groups tended to be more sensitive to linguistic features than TD-1 children. Namely, the children with DLD produced sentences with prepositional phrases less accurately than with instrumental phrases (DLD-1) and semantically reversible sentences less accurately than nonreversible ones (e.g., 'A boy is hugging a girl' vs. 'A woman is watering a flower'; DLD-2). Furthermore, the children with DLD struggled with questions in the Discourse comprehension subtest. Specifically, implicit questions were harder to respond to than explicit questions and questions about the details of the story were harder to respond to than those about the main theme of the story. Finally, the children with DLD had more grammatical errors than TD-1 children in the Sentence production and Discourse production subtests. They also had problems with coherence and fluency in the Discourse production subtest.

Table 4. The key findings in language subtests in the three groups of children (TD-1 children, children with DLD-1 from Group with higher test scores, and children with DLD-2 from Group 2 with lower test scores).

Subtest and linguistic features	TD	DLD-1	DLD-2

Phonological discrimination - length in phonemes	 n/s	= TD n/s	**< TD** n/s
Nonword repetition - length in syllables - number of articulatory switches	 n/s n/s	= TD n/s n/s	= TD **long < short** n/s
Noun & Verb comprehension - word class	 **verbs < nouns***	= TD **verbs < nouns**	**< TD*** **verbs < nouns***
Object & Action naming - word class	 n/s	**< TD** n/s	**< TD*** n/s
Sentence comprehension - word order - third argument	 n/s n/s	= TD n/s n/s	**< TD*** n/s n/s
Sentence production - reversibility - third argument - priming - grammar - lexicon - other	 n/s n/s n/s n/s n/s n/s	= TD n/s **prepositional < instrumental** = TD **< TD** = TD **< TD**	**< TD*** **reversible < irreversible*** n/s **< TD*** **< TD*** **< TD*** **< TD***
Sentence repetition - word frequency - sentence length	 **high < low** **long < short***	= TD **high < low** **long < short***	= TD **high < low*** **long < short***
Discourse comprehension - line of story - type of information	 n/s n/s	= TD n/s **implicit < explicit**	**< TD*** **detail < main** n/s
Discourse production		**< TD**	**< TD***

- coherence	n/s	**< TD**	**< TD***
- grammar	n/s	**< TD***	**< TD***
- lexicon	n/s	= TD	**< TD***
- fluency	n/s	**< TD**	**< TD***

Note. The ‘<’ means ‘less accurately’; ‘=’ means ‘with comparable accuracy’; n/s means ‘non-significant’. The level of significance was Bonferroni corrected to p = 0.005. Significant results before the Bonferroni correction are in bold, $p < 0.05$. Significant results after Bonferroni correction are also marked with an asterisk (*), $p < 0.005$.

Patterns of errors in individual children

Based on the cluster analyses at different language levels, we revealed the combinations of levels that triggered more errors in children with DLD. Here, we report the qualitative analysis of errors in different subtests for TD and DLD children.

Children from the DLD-1 group (with higher overall scores) showed worse performance only at 1-2 language levels: phonological and/or discourse. Notably, only 4-year-old children had low scores at both phonological and discourse levels, whereas 5-to-7-year-old children struggled only at one of the levels. Interestingly, these patterns of errors are similar to those of TD-1 children. Furthermore, all children who had lower scores at the phonological level had errors in comprehension (Phonological discrimination) and production (Nonword repetition).

Children from the DLD-2 group (with lower overall scores) showed worse performance at 2-4 language levels in different combinations. At the same time, six TD-2 children had lower performance at phonological and discourse levels and the other two TD participants also had lower scores at the lexical level. In contrast to children with DLD, none of the TD-2 children struggled at the morphosyntactic level. So, DLD-2 and TD-2 children showed a combination of weaknesses at several language levels. DLD-2 children are more impaired in Linguistic performance than TD-2 children.

Discussion

The present study aimed to test the predictions of the continuum/spectrum model of linguistic skills in a group of 4-to-7-year-old, Russian-speaking children, with and without DLD. We used 11 subtests of the Russian Child Language Assessment Battery (RuCLAB) to assess childrens' performance at all language levels in comprehension and production domains. With all the obtained scores we performed a cluster analysis and then we described the childrens' linguistic skills by comparing groups that emerged in the cluster analysis. We also analyzed the linguistic features of the stimuli that might influence children's performance in subtests. Finally, we described patterns of language deficits in children in different clusters.

We found that some children with diagnosed DLD could perform on par with TD children in most language subtests, whereas some younger TD children could have performance similar to children with DLD. Our results for children with DLD supported the results of the previous study that introduced the continuum/spectrum model and extended it to TD children (Lancaster & Camarata, 2019). We concluded that linguistic skills in pre-school children vary from high to low at all language levels in comprehension and production. Children who are in the middle of the continuum have individual patterns of high and low skills depending on the language level. We suggest that to properly describe a child's language profile, multiple tasks assessing all language levels should be used.

Cluster analysis and the continuum/spectrum model

Using cluster analysis, we divided all children into groups, based solely on their performance in the language assessment and regardless of the presence/absence of clinical diagnosis. The analysis revealed that all children can be divided into two clusters based on their cumulative score in RuCLAB. Both clusters included children with DLD from all four age groups and at least some TD children. Therefore, our a priori division of participants into TD and DLD groups, based on the diagnosis from speech-language pathologists, was not fully supported by the cluster analysis. Specifically, the fact that we obtained two mixed clusters, instead of two clusters with only TD or DLD participants in them, allowed us to conclude that subgrouping children into TD/DLD categories does not correlate with the actual linguistic skills of the participants.

Our results for children with DLD are in agreement with the study by Lancaster and Camarata (2019), in which they concluded that the continuum/spectrum model was the best model for embracing the variety of language profiles in the population of children with DLD. Crucially, the fact that some younger TD children were also a part of the Group with lower results can indicate that the continuum/spectrum approach initially proposed for children with DLD may be extended to TD children whose performance in language tests is below average (in our study, the average scores were calculated for the whole group of 4-to-7-year-old participants). Importantly, the hypothesis about the continuum/spectrum model in typical and atypical language development was discussed in the CATALISE panel (Bishop et al., 2016), but previously was not tested in the experimental studies. Thus, our study is the first to show that the continuum/spectrum model best describes the results of language assessment in pre-schoolers. Moreover, our findings indicate that 4-to-5-year-old participants tended to be in the Group with lower results, whereas 6-to-7-year-old children tended to be in the Group with higher results. This observation may highlight the necessity of using different criteria for intervention provision in 4-year-old children and children from 5 years old.

We hypothesize that the continuum/spectrum model may be applied to the entire range of emerging speakers, including late talkers, children with non-severe social deprivation, children who outgrew their impairment with the help of intervention, and TD children whose performance in language tests was below average (see similar debates about classifying children with Autism Spectrum Disorder (ASD) in (van Rentergem et al., 2021)). Future studies should test whether the model can explain patterns of language development in different groups of typically and atypically developing children. The advantage of the continuum/spectrum approach is that it takes into account the severity of symptoms and allocates children with needs for language pathologists' intervention that would not be diagnosed with DLD by the classical DSM-5 approach. In summary, we suppose that the continuum/spectrum model is the optimal approach for describing language development variability in children.

Language profiles in TD children and children with DLD

To describe the language profiles in children with DLD who were assigned to different Groups by cluster analysis, we compared the two groups of children with DLD to TD children. Below, we discuss the findings at each of the four language levels.

Phonological level. A few studies suggested that single phonological tasks could identify children with DLD (Coady & Evans, 2008; Thordardottir & Reid, 2022). However, in a recent metaanalysis, Schwob et al. (2021) found that the diagnostic accuracy of the Nonword repetition task was generally near the threshold of DLD identification, and to obtain more robust measures, the phonological task should be combined with other tasks. Similar conclusions were reported by Boerma & Blom (2017) for bilingual children with DLD: the authors suggested to combine the Nonword repetition with narrative tasks. In the present study, we found that both groups of children with DLD did not differ significantly from TD children in the Nonword repetition and Phonological discrimination tasks (except for the tendency for the DLD-2 group). So, for better assessment of DLD impairments, we propose combining phonological and other language level assessment and evaluate both comprehension and production domains.

Lexical level. The children from the DLD-2 group (with lower results) had lower accuracy compared to TD children in word comprehension and production, whereas the results of the DLD-1 group were comparable with TD children. Our results are partly in line with the previous studies that found differences between TD children and children with DLD in both word production and comprehension tasks (Vukovic et al., 2010; van Weerdenburg et al., 2006). We suggest that these tasks can help to identify children with severe language impairments and may not be discriminative for mild impairments. Interestingly, all children comprehended verbs less accurately than nouns, whereas there was no such difference in production. These results are in line with previous studies which revealed that verb learning is particularly challenging for children with DLD (Skipp et al., 2002) as well as for TD children (D'Odorico & Fasolo, 2007; Goldfield, 2000). The absence of the difference between nouns and verbs in word production subtests may be explained by the Sheng & McGregor (2010) study. They found that the gap between action- and object-naming depends on a child's cumulative vocabulary and the effect of word class on naming accuracy was more pronounced in younger children.

Morphosyntactic level. We found that children from the DLD-1 group (with higher results) did not differ significantly from TD children, whereas the DLD-2 group (with lower results) made significantly more errors in the Sentence comprehension and Sentence production subtests. Within the Sentence production subtest, we found that only DLD-2 children produced semantically reversible sentences less accurately than nonreversible ones. It means that the reversibility of the syntactic structure affected sentence production in children with more impaired linguistic skills. Our findings are in line with the results from Hinzen et. al (2022), who showed that TD children started to perceive reversible events at age of 4, while children with DLD lagged behind their chronological age controls. Interestingly, in the Sentence repetition subtest, we did not find differences between the three groups of participants, which is in contrast with previous research (Hesketh & Conti-Ramsden, 2013; Stokes et al., 2006). All three groups repeated long sentences less accurately than short sentences. However, only the DLD-2 group was sensitive to the word frequency effect: these children struggled to repeat sentences with low frequency words. For the other two groups, the effect of frequency was non-significant after Bonferroni correction. In summary, at the morphosyntactic level, we found that children with DLD, with more impaired linguistic skills, struggled with sentence comprehension and production and were sensitive to the linguistic features of the sentences. DLD-1 children, with better linguistic skills, did not differ significantly from TD children. We concluded that morphosyntactic tasks alone might not be discriminative for all children with language impairments.

Discourse level. In the Discourse comprehension subtest, children from the DLD-2 group made more errors compared to TD children. They also seemed to struggle more with answering questions about the details of the story. Children from the DLD-1 and TD groups responded to comprehension questions with comparable accuracy, and DLD-1 children tended to make more errors when answering implicit questions. These results are in line with previous findings that children with DLD may not have problems with surface narrative comprehension, but lose to interpret implicit information and to note the details (Filiatrault-Veilleux et al., 2015; Filiatrault-Veilleux et al., 2016).

Altogether, these results support the idea that discourse intervention models for children with DLD should emphasize implicit information because it is the main source of deficit in discourse comprehension (Swanson et al., 2005; Petersen et al., 2010). In the Discourse production subtest, children from the DLD-2 group made more errors of all types (in coherence, grammar, lexicon, and fluency) compared to TD children. Children from the DLD-1 group tended to make more errors in coherence and fluency, and made significantly more errors in grammar compared to TD children (in line with (Kornev & Balciuniene, 2021) for Russian-speaking children). Our results support previous studies which concluded that lower scores in lexicon, grammar, and fluency can differentiate children with DLD from their TD peers in narrative production tasks (Winters et al., 2022). Furthermore, our results support the view that problems in coherence are critical factors that distinguish children with DLD from TD children (Merritt & Liles, 1987; Van der Lely, 1997).

Finally, the analysis of the patterns of errors in individual children showed that many children with DLD had low scores at two or more language levels. It means that DLD is usually manifested at more than one language level. These findings are in line with Tomblin et al. (1996) EpiSLI study, in which the authors showed that the combination of two and more language levels' scores below 1.25 SD should be interpreted as DLD. Importantly, our results allow us to specify this recommendation: the application of Tomblin et al.'s principle would be more reliable in children from 5 years old. In our study, 4-year-old children who had lower scores in phonological and discourse subtests still belonged to the cluster with the overall good linguistic skills. Possibly, younger children should meet wider criteria for DLD diagnosis establishment (e.g., three or more language levels' scores below 1.25 SD).

Limitations

Our study has two limitations. The first limitation is a relatively small sample size (106 children in total), whereas the previous cluster analysis study for testing the continuum/spectrum model included data from 505 participants (Lancaster & Camarata, 2019). Although power analysis is generally not available for clustering methods, it is recommended to have as many participants as possible (Allen & Goldstein, 2013). The second limitation is different number of participants in each age group, which makes our samples unbalanced. Previous studies showed that at early ages linguistic skills changed dramatically each year (Ambridge & Lieven, 2011). So, selecting a larger and more balanced sample for testing the continuum/spectrum model may be a promising line of research.

Clinical implications

Several studies proposed that a single task can distinguish children with DLD from their typically developing peers. For example, some papers showed that the area of weaknesses for children with DLD is Sentence repetition (e.g. Pham & Ebert, 2020; Vang Christensen, 2019; Taha et al., 2021); or Nonword and Real-word repetition (e.g., Dispaldro et al., 2013; Girbau & Schwartz, 2008). However, our results do not support the single-task approach. Our findings highlight the necessity of comprehensive language assessment that should include evaluation of all language levels in children. For efficient diagnostics, it is essential to assess production and comprehension at all language levels. Our findings are in line with the principle of Tomblin et al. (1996) who suggested diagnosing children with DLD in the case of two or more language levels' scores below 1.25 SD. This principle should be widened for 4-year-old children.

There is little evidence of the application of the continuum/spectrum model in clinical routine. However, theoretical underpinnings of DLD and research dedicated to disorders such as ASD and attention deficit hyperactive disorder (ADHD) demonstrate the high potential of the continuum/spectrum approach for speech/language assessment and treatment planning (Agelink van Rentergem et al., 2021). The application of the continuum/spectrum model in DLD diagnostics will lead to an individualized approach, allowing the clinicians to focus on the specific symptoms and their severity.

The agreement on accepting the 1.25 SD below average on two or more language levels, as the cutoff to detect DLD may be not perfect. In the present study, we demonstrated that participants who did not reach that cutoff but performed below average on two or more language levels, may still need help from a speech therapist.

The complex and comprehensive assessment of all levels of language functioning during their development will be beneficial for understanding the nature of the deficit.

Acknowledgements

We want to thank the preschool department of Educational center №1 in Ivanteevka, Russia, the Center of children development (Kindergarten №188) in Saratov, Russia and all the children and their caretakers who participated in this study, as well as Nina Zdorova and Anastasiya Sycheva for their contribution to data collection. We are grateful to Konstantin Lopukhin for his help with cluster analysis.

Contribution (CRediT)

Militina Gomozova: Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing
Valeriia Lezzhova: Investigation, Data Curation, Writing - Review & Editing
Olga Dragoy: Resources, Writing - Review & Editing, Funding acquisition
Anastasiya Lopukhina: Conceptualization, Methodology, Formal analysis, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration

References

1. Agelink van Rentergem, J., Deserno, M., Geurts, H. (2021). Validation strategies for subtypes in psychiatry: A systematic review of research on autism spectrum disorder. Clinical Psychology Review, 87, 102033. https://doi.org/10.1016/j.cpr.2021.102033
2. Akinina, Y., Grabovskaya, M., Vechkaeva, A., Ignatyev, G., Isaev, D., & Khanova, A. (2016). Biblioteka psiholingvisticheskih stimulov: Novye dannye dlja russkogo i tatarskogo jazyka. In Y. Aleksandrov, & K. Anokhin (Eds.), The Seventh International Conference on Cognitive Science (pp. 93–95). Institute of Psychology of Russian Academy of Sciences.
3. Akinina, Y., Iskra, E., Ivanova, M., Grabovskaya, M., Isaev, D., Korkina, I., Malyutina, S., Sergeeva, N. (2014). Biblioteka stimulov «Sushestvitel'noe i ob'ekt»: Normirovanie psiholingvisticheskih parametrov. In B. Velichkovsky, V. Rubtsov, & D. Ushakov (Eds.), The Sixth International Conference on Cognitive Science (pp. 112–114). Institute of Psychology of Russian Academy of Sciences.

4. Akinina, Y., Malyutina, S., Ivanova, M., Iskra, E., Mannova, E., & Dragoy, O. (2015). Russian normative data for 375 action pictures and verbs. Behavior Research Methods, 47(3), 691–707. https://doi.org/10.3758/s13428-014- 0492-9
5. Allen, D. N., & Goldstein, G. (Eds.). (2013). *Cluster analysis in neuropsychological research: Recent applications.* Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-6744-1
6. Ambridge, B., & Lieven, E. (2011). Related debates and conclusions. In *Child Language Acquisition: Contrasting Theoretical Approaches* (pp. 359-376). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511975073.010
7. American Psychiatric Association, DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™*(5th ed.). American Psychiatric Publishing, Inc.. https://doi.org/10.1176/appi.books.9780890425596
8. Arutiunian, V., Lopukhina, A., Minnigulova, A., Shlyakhova, A., Davydova, E., Pereverzeva, D., Sorokin, A., Tyushkevich, S., Mamokhina, U., Danilina, K., & Dragoy, O. (2022). Language Abilities of Russian Primary-School-Aged Children with Autism Spectrum Disorder: Evidence from Comprehensive Assessment. *Journal of autism and developmental disorders*, *52*(2), 584–599. https://doi.org/10.1007/s10803-021-04967-0
9. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005
10. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01
11. Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of speech, language, and hearing research : JSLHR*, *44*(4), 905–924. https://doi.org/10.1044/1092-4388(2001/072)
12. Bishop, D.V.M. (2014). Ten questions about terminology for children with unexplained language problems. *International journal of language & communication disorders*, 49(4), 381-415. https://doi.org/10.1111/1460-6984.12101
13. Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin, 130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858
14. Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PloS one*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753
15. Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & 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
16. Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible?. *Journal of communication disorders*, *66*, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001
17. Botting, N., Faragher, B., Simkin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: what differentiates good and poor outcome?. *Journal of child psychology and psychiatry, and allied disciplines*, *42*(8), 1013–1020. https://doi.org/10.1111/1469-7610.00799
18. Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). International Journal ofLanguage & Communication Disorders, 43(1), 1–40. https://doi.org/10.1080/
19. Conti-Ramsden G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of speech, language, and hearing research : JSLHR*, *46*(5), 1029–1037. https://doi.org/10.1044/1092-4388(2003/082)
20. Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: longitudinal considerations. *Journal of speech, language, and hearing research : JSLHR*, *42*(5), 1195–1204. https://doi.org/10.1044/jslhr.4205.1195
21. Culatta, B., Page, J.L. & Ellis, J. (1983). Story telling as a communicative performance screening tool. Language, Speech, and Hearing Services in Schools, 14(2), 66–74. https://doi.org/10.1044/0161-1461.1402.66
22. Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: a study of diagnostic accuracy. *Journal of speech, language, and hearing research : JSLHR*, *56*(1), 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)
23. D'Odorico, L., & Fasolo, M. (2007). Nouns and verbs in the vocabulary acquisition of Italian children. *Journal of Child Language, 34*(4), 891–907. https://doi.org/10.1017/S0305000907008240
24. Ellis, E., Thal, D. (2008). Early Language Delay and Risk for Language Impairment. Perspectives on Language Learning and Education, 15 (3), 93-100. https://doi.org/10.1044/lle15.3.93
25. Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N. & Desmarais, C. (2015), Inferential comprehension of 3–6 year olds within the context of story grammar: a scoping review. *International Journal of Language & Communication Disorders*, 50(6), 737-749. https://doi.org/10.1111/1460-6984.12175
26. Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2016). Comprehension of Inferences in a Narrative in 3- to 6-Year-Old Children. *Journal of speech, language, and hearing research : JSLHR*, *59*(5), 1099–1110. https://doi.org/10.1044/2016_JSLHR-L-15-0252

27. Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish--English bilingual children with and without specific language impairment. *Journal of Communication Disorders, 41*(2), 124–145. https://doi.org/10.1016/j.jcomdis.2007.07.001
28. Goldfield, B. A. (2000). Nouns before verbs in comprehension vs. production: the view from pragmatics. *Journal of child language*, *27*(3), 501–520. https://doi.org/10.1017/s0305000900004244
29. Hannus, S., Kauppila, T., Pitkäniemi, J., & Launonen, K. (2013). Use of language tests when identifying specific language impairment in primary health care. *Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics (IALP)*, *65*(1), 40–46. https://doi.org/10.1159/000350318
30. Hansson, K., Sandgren, O., & Sahlén, B. (2014). Changing labels for a concept in change. Commentaries to: Bishop, D.V.M. Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders*, *49*(4), 407-408. https://doi.org/10.1111/1460-6984.12101
31. Hesketh, A., & Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *PloS one*, *8*(2), e56314. https://doi.org/10.1371/journal.pone.0056314
32. Hinzen, W., Peinado, E., Perry, S. J., Schroeder, K., & Lombardo, M. (2022). Language level predicts perceptual categorization of complex reversible events in children. *Heliyon*, *8*(7), e09933. https://doi.org/10.1016/j.heliyon.2022.e09933
33. Ivanova, M., Dragoy, O., Akinina, J., Soloukhina, O., Iskra, E., Khudyakova, M., & Akhutina, T. (2016). AutoRAT as your fingertips: Introducing the new Russian Aphasia Test on tablet. Frontiers in Psychology Conference Abstract: 54th Annual Academy of Aphasia Meeting. https://doi.org/10.3389/conf.fpsyg.2016.68.00116
34. Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: a deficit in grammar or processing?. *Trends in cognitive sciences*, *2*(7), 240–247. https://doi.org/10.1016/S1364-6613(98)01186-3
35. Kaufman, L. & Rousseeuw, P. (1990). Finding Groups in Data: An Introduction To Cluster Analysis. Wiley Series in Probability and Statistics. DOI:10.1002/9780470316801
36. Kornev, A. N. (2006). *Fundamentals of Childhood Speech Pathology: Clinical and Psychological Aspects*, Rech, St. Petersburg.
37. Kornev, A. N., & Balčiūnienė, I. (2021). Lexical and Grammatical Errors in Developmentally Language Disordered and Typically Developed Children: The Impact of Age and Discourse Genre. *Children (Basel, Switzerland)*, *8*(12), 1114. https://doi.org/10.3390/children8121114
38. Kornilov, S. A., Lebedeva, T. V., Zhukova, M. A., Prikhoda, N. A., Korotaeva, I. V., Koposov, R. A., Hart, L., Reich, J., & Grigorenko, E. L. (2016). Language development in rural and urban Russian-speaking children with and without

developmental language disorder. *Learning and individual differences*, *46*, 45–53. https://doi.org/10.1016/j.lindif.2015.07.001

39. Lancaster, H. S., & Camarata, S. (2019). Reconceptualizing developmental language disorder as a spectrum disorder: issues and evidence. *International journal of language & communication disorders*, *54*(1), 79–94. https://doi.org/10.1111/1460-6984.12433
40. Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *International journal of language & communication disorders*, *35*(2), 165–188. https://doi.org/10.1080/136828200247133
41. Leonard, L. B. (2014). Children with specific language impairment and their contribution to the study of language development. *Journal of child language*, *41 Suppl 1*(01), 38–47. https://doi.org/10.1017/S0305000914000130
42. Lüdecke, D. (2017). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.3.1. https://CRAN.R-project.org/package=sjPlot
43. Marini, A., Tavano, A., & Fabbro, F. (2008). Assessment of linguistic abilities in Italian children with Specific Language Impairment. Neuropsychologia, 46(11), 2816-2823. https://doi.org/10.1016/j.neuropsychologia.2008.05.013.
44. Marshall, C. R., & Van Der Lely, H. K. J. (2007). The impact of phonological complexity on past tense inflection in children with Grammatical-SLI. *Advances in Speech Language Pathology,* *9*(3), 191–203. https://doi.org/10.1080/14417040701261509
45. McLeod, S., & McKinnon, D. H. (2007). Prevalence of communication disorders compared with other learning needs in 14,500 primary and secondary school students. *International journal of language & communication disorders*, *42 Suppl 1*, 37–59. https://doi.org/10.1080/13682820601173262
46. Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech & Hearing Research, 30*(4), 539–552. https://doi.org/10.1044/jshr.3004.539
47. Newcomer P. L., Hammill D. D. (2008). *Test of language development–Primary* (4th ed.). Austin, TX: PRO-ED.
48. Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of child psychology and psychiatry, and allied disciplines*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573
49. Petersen, D. B., Gillam, S. L., Spencer, T., & Gillam, R. B. (2010). The effects of literate narrative intervention on children with neurologically based language impairments: an early stage study. *Journal of speech, language, and hearing research : JSLHR*, *53*(4), 961–981. https://doi.org/10.1044/1092-4388(2009/09-0001)

50. Pham, G., & Ebert, K. D. (2020). Diagnostic Accuracy of Sentence Repetition and Nonword Repetition for Developmental Language Disorder in Vietnamese. *Journal of speech, language, and hearing research : JSLHR*, *63*(5), 1521–1536. https://doi.org/10.1044/2020_JSLHR-19-00366
51. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
52. Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. Cognitive Psychology, 41(1), 1-48. https://doi.org/10.1006/cogp.1999.0735.
53. Reilly, S., Bishop, D. V., & Tomblin, B. (2014). Terminological debate over language impairment in children: forward movement and sticking points. *International journal of language & communication disorders*, *49*(4), 452–462. https://doi.org/10.1111/1460-6984.12111
54. Rice, M. (2014). Advocating for SLI. Commentary on Bishop, D.V.M., 2014, Ten questions about terminology for children with unexplained language problems. International journal of language & communication disorders, 49(4), 381.
55. Roberts, S. S., & Leonard, L. B. (1997). Grammatical deficits in German and English: a crosslinguistic study of children with specific language impairment. *First Language*, *17*(51), 131–150. https://doi.org/10.1177/014272379701705106
56. Rutter, M. (2014). Classificatory purposes and diagnostic concepts. Commentary on Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J.M. & Wake, M. (2014). Specific language impairment: a convenient label for whom? International journal of language & communication disorders, 49(4), 416-451.
57. Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of speech, language, and hearing research : JSLHR*, *53*(1), 146–159. https://doi.org/10.1044/1092-4388(2009/08-0160)
58. Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of speech, language, and hearing research : JSLHR*, *64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552
59. Skipp, A., Windfuhr, K. L., & Conti-Ramsden, G. (2002). Children's grammatical categories of verb and noun: a comparative look at children with specific language impairment (SLI) and normal language (NL). *International journal of language & communication disorders*, *37*(3), 253–271. https://doi.org/10.1080/13682820110119214
60. Snowling, M.J. (2014). SLI: Not just a semantic issue. Commentary on Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F.K., Morgan, A., Goldfeld S., Nicholson, J.M. & Wake, M. (2014). Specific language impairment: a

convenient label for whom? International journal of language & communication disorders, 49(4), 416–451.

61. Stokes, S. F., Wong, A. M., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *Journal of speech, language, and hearing research : JSLHR*, *49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)
62. Swanson, L. A., Fey, M. E., Mills, C. E., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American journal of speech-language pathology*, *14*(2), 131–143. https://doi.org/10.1044/1058-0360(2005/014)
63. Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Sentence Repetition as a Clinical Marker of Developmental Language Disorder: Evidence From Arabic. *Journal of speech, language, and hearing research : JSLHR*, *64*(12), 4876–4899. https://doi.org/10.1044/2021_JSLHR-21-00244
64. Thordardottir, E. & Reid, H. (2022). Nonword repetition to identify DLD in older school-age children. Dans Enfance, 1(1), 59-79. https://doi.org/10.3917/enf2.221.0059
65. Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research : JSLHR*, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245
66. Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of speech and hearing research*, *39*(6), 1284–1294. https://doi.org/10.1044/jshr.3906.1284
67. Van Borsel, J., Moeyaert, J., Mostaert, C., Rosseel, R., Van Loo, E., & Van Renterghem, T. (2006). Prevalence of Stuttering in Regular and Special School Populations in Belgium Based on Teacher Perceptions. *Folia Phoniatrica et Logopaedica*, 58 (4), 289–302. https://doi.org/10.1159/000093185
68. Van der Lely, H. K. (1997). Narrative discourse in Grammatical specific language impaired children: a modular language deficit?. *Journal of child language*, *24*(1), 221–256. https://doi.org/10.1017/s0305000996002966
69. Van Weerdenburg, M., Verhoeven, L., & van Balkom, H. (2006). Towards a typology of specific language impairment. *Journal of child psychology and psychiatry, and allied disciplines*, *47*(2), 176–189. https://doi.org/10.1111/j.1469-7610.2005.01454.x
70. Vang Christensen, R. (2019). Sentence Repetition: A Clinical Marker for Developmental Language Disorder in Danish. *Journal of speech, language, and hearing research : JSLHR*, *62*(12), 4450–4463. https://doi.org/10.1044/2019_JSLHR-L-18-0327
71. Vukovic, M., Vukovic, I., & Stojanovik, V. (2010). Investigation of language and motor skills in Serbian speaking children with specific language impairment and in typically developing children. *Research in developmental disabilities*, *31*(6), 1633–1644. https://doi.org/10.1016/j.ridd.2010.04.020

72. Wechsler, D. (1989). Wechsler Preschool and Primary Scale of Intelligence—Revised (San Antonio, TX: Psychological Corporation).
73. Weindrich, D., Jennen-Steinmetz, C., Laucht, M., Esser, G., & Schmidt, M. H. (2000). Epidemiology and prognosis of specific disorders of language and scholastic skills. *European child & adolescent psychiatry*, *9*(3), 186–194. https://doi.org/10.1007/s007870070042
74. Winters, K. L., Jasso, J., Pustejovsky, J. E., & Byrd, C. T. (2022). Investigating Narrative Performance in Children With Developmental Language Disorder: A Systematic Review and Meta-Analysis. *Journal of speech, language, and hearing research : JSLHR*, *65*(10), 3908–3929. https://doi.org/10.1044/2022_JSLHR-22-00017
75. World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053