

methods

Assessing the Validity of the Standardized Assessment of Reading Skills in Russian and Verifying the Relevance of Available Normative Data

Svetlana V. Dorofeeva

National Research University Higher School of Economics, Moscow, Russia

Victoria Reshetnikova

National Research University Higher School of Economics, Moscow, Russia

Margarita Serebryakova

Center for Speech Pathology and Neurorehabilitation, Moscow, Russia

Daria Goranskaya

National Research University Higher School of Economics, Moscow, Russia

Tatiana V. Akhutina

Lomonosov Moscow State University, Moscow, Russia

Olga Dragoy

National Research University Higher School of Economics, Moscow, Russia;
Federal Center for Cerebrovascular Pathology and Stroke, Moscow, Russia

Abstract. Standardized tests with normative data have become the gold standard for assessing reading skills and for diagnosing a specific reading disorder (developmental dyslexia) in many languages. For Russian, there is one such reading assessment test called the Standardized Assessment of Reading Skills, developed by A.N. Kornev and first published in 1997. However, the most recent available normative data on this methodology were collected more than a decade ago, and researches did not control for several important variables. Furthermore, no details have been published about the diagnostic validity of this test. We used the test to assess reading skills in 90 typically developing Russian primary school children in 2018. In this article, we present the results of testing typically developing children, including updated values of reading fluency and, for the first time, metrics of reading comprehension and weighted error scores. Additionally, we tested 50 children with clinically diagnosed developmental dyslexia and provide information about the sensitivity and specificity of the Standardized Assessment of Reading Skills.

Correspondence: Svetlana V. Dorofeeva, sdorofeeva@hse.ru, Center for Language and Brain, National Research University Higher School of Economics, 21/4 Staraya Basmannaya str., Office 510, 105066 Moscow, Russian Federation; Victoria Reshetnikova, victoriaregina98@gmail.com; Margarita Serebryakova, margoha1986@yandex.ru; Daria Goranskaya, daria.goranskaya@gmail.com; Tatyana V. Akhutina, akhutina@mail.ru; Olga Dragoy, odragoy@hse.ru

Keywords: reading fluency, reading comprehension, Russian, developmental dyslexia, SARS, Standardized Assessment of Reading Skills

Copyright © 2019. Svetlana V. Dorofeeva, Victoria Reshetnikova, Margarita Serebryakova, Daria Goranskaya, Tatyana V. Akhutina, Olga Dragoy. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author is credited and that the original publication in this journal is cited, in accordance with accepted academic practice.

Acknowledgements. We gratefully acknowledge the help provided by L.A. Pushkareva and E.I. Larionova (a psychologist and the Director of School 556 of the Southern District of Moscow, Russia), A.V. Kornilov (the Head of the Department 968 of School 1236 named after S.V. Milashenkov, Moscow, Russia), N.S. Volovik and M.A. Sukiasyan (the Assistant Director and the Director of General School “Pokolenie”, Volgograd, Russia), and R.A. Cheremin (the Chief Doctor of the Center for Speech Pathology and Neurorehabilitation, Moscow, Russia). We are thankful to all the children who took part in this research and to their parents. This study was financially supported by the Russian Foundation for Basic Research, Project No. 17-29-09122.

Received January 15, 2019, accepted March 20, 2019.

Introduction

Standardized tests are necessary for reliable and valid quantification of reading abilities, and for comparison between groups and individuals in order to diagnose reading disorders. Reading tests with uniform administration protocol and normative data from a large sample of individuals have become a methodological gold standard in last decades. Complex reading and spelling tests were developed for the German language (Moll & Landerl, 2010; Wimmer & Mayringer, 2014), standardized word-reading test (Brus & Voeten, 1973) and standardized pseudoword-reading test (van den Bos, Spelberg, Scheepstra, & DeVries, 1994) were developed for Dutch. For English, several variants of standardized reading assessments are available, including specialized reading test batteries (Torgesen, Wagner, & Rashotte, 1999; Wechsler, 1990; Woodcock, 1999) and reading subtests of more comprehensive cognitive assessment tools, such as Wide Range Achievement Test (Wilkinson & Robertson, 2006). A standardized reading test is also available for Russian.

The Standardized Assessment of Reading Skills (SARS) for Russian-speaking children was developed in 1982, first published in 1997 (Kornev, 1997), and republished in 2003 (Kornev, 2003). The second round of normative data collection was carried out in 2007–2008 (Kornev & Ishimova, 2010) to capture changes in the mean reading performance of children which happen over time. This test, according to the author, is intended for the “diagnosis of reading disorders and differential diagnosis of developmental dyslexia and non-specific reading disorders” (Kornev & Ishimova, 2010). Other existing reading assessment tools for Russian are parts of comprehensive neuropsychological assessment protocols (Akhutina et al., 2016; Semenovich, 2002), but their diagnostic application is limited due to the lack of published normative data.

The SARS has a number of obvious strengths. Firstly, during restandardization the number of children tested for normative data was substantially increased. First edition (Kornev, 1997) was published with normative data of 150 schoolchildren. The current edition (Kornev & Ishimova, 2010) was published with normative data of 700 schoolchildren. Secondly, this reading test provides the opportunity to measure not only reading fluency, but also reading comprehension using content questions with straightforward guidelines for scoring the responses as correct or incorrect. Thirdly, it includes texts of varying

difficulty (simple Text I and more difficult Text II), which allows to compare the performance of a reader as a function of increasing text difficulty. Kornev and Ishimova (2010) report that 10 and more points difference between reading performance of the two texts indicates greater severity of reading disorder. Finally, the SARS includes additional texts intended for retesting if necessary, with levels of difficulty comparable to the first two texts, and also with normative data of 400 Moscow schoolchildren tested in 2009–2010.

However, the SARS in its present form has limitations. Randomly chosen children from public schools took part in testing for normative data collection (Kornev & Ishimova, 2010), while it is preferable to include children without diagnosed neurological disorder or intellectual disabilities to determine appropriate cutoff scores for normal performance (Ivanova & Hollowell, 2013). In addition, despite the provided content questions for assessing reading comprehension and the well-developed criteria for scoring responses as correct or incorrect, the authors have not published normative data for reading comprehension. They suggested considering the level of performance with at least 7 out of 10 correct responses as “functionally sufficient”. At the same time, they provided information on the distribution of reading comprehension levels in the tested Moscow schoolchildren. From these data of the SARS’ authors, it can be calculated that 53% and 69% of second-graders correctly answered fewer than 7 questions for Text I and Text II, respectively (Kornev & Ishimova, 2010). Thus, the suggested level-based recommendation cannot serve as a guide for diagnosing reading comprehension disorder. Normative scores for reading comprehension are required, similarly to reading fluency.

Even more importantly for diagnostic purposes, in the last edition of the SARS (Kornev & Ishimova, 2010), direct normative data, e.g., the mean reading fluency and the standard deviation, were not provided. Instead, the authors suggested a transformation with a special formula resulting in coefficients of reading technic (see Method section). Those coefficients were not given for every possible number of correctly read words, but with gaps of 3 points. As a result, when the performance of a particular child is at the borderline (e.g., between “at risks” and “dyslexia”), there is no definite answer to which group the performance should be assigned to. Finally, Kornev (2003, p. 213) claimed that “the test showed sufficient validity in the clinical population”, but in fact no psychometric properties of the SARS were provided (e.g., sample size, sensitivity, and

specificity of the test). As a result, the diagnosis of developmental dyslexia in Russian-speaking children based on this test remains a challenge to clinical practice.

We conducted a study aiming to overcome some current limitations of the SARS and to improve its clinical application. First, to verify the relevance of the existing normative data we collected new data on the SARS performance by typically developing Russian school-children with measured intact non-verbal intelligence and normal hearing, normal or corrected to normal vision, and no diagnosed neurological disorders. Unlike the authors of the SARS, who provided norms for grades 2 to 6 of the Russian school system, we also tested children in the second half of the grade 1, to make it possible to identify children with or at risk of developing reading impairment as early as possible. As a result, we report new direct data (means and standard deviations) for reading fluency and reading comprehension in typically developing Russian school-children for each of the primary school grades.

Additionally, we assessed the validity of the SARS for the diagnosis of developmental dyslexia. For that, we tested 50 children who were clinically diagnosed with developmental dyslexia. We examined the correspondence between the clinical diagnosis and the performance of these children on the SARS using original norms provided in the last published manual for the test (Kornev & Ishimova, 2010), and our new data.

Method

Participants

Typically developing participants were recruited at three Russian public schools, two in Moscow ($n=58$) and one in Volgograd ($n=47$). All children (total $N=105$) were native Russian speakers, 7 to 11 years of age, and were elementary school students in the first through fourth grades. The children had no history of diagnosed neurological disorders, no diagnosed problems with reading acquisition, and all had normal or corrected-to-normal vision (as reported in the informed consents given by their parents or legal representatives). Screening for primary auditory impairments (using the Audiogramm program version 4.6.1.3, Professional Audiometric System; Sennheiser HDA 280 audiometry headphones) resulted in exclusion of three participants. Screening for non-verbal intelligence with the Raven's Colored Progressive Matrices (Raven, 2004; Raven, Raven, & Kort, 2012) resulted in the exclusion of 12 participants who scored below norms. Ninety children (48 girls; 50 children from Moscow; 7 left-handers; $M_{age}=8.7$, $SD=1.13$) remained in the analysis. General information about typically developing participants included in the analysis is presented in Table 1.

Before using the test results of children from different cities in the consolidated analysis, we analyzed the effect of the city on reading performance, taking into account the grade and gender of participants. For this, we used linear regression models that were built with the *lme4* package for the statistical data processing program *R* (Bates et al., 2015). Factors such as city, grade, and gender were included in the model as predictors for reading fluency. No statis-

Table 1. Overview of Typically Developing Participants

N=90	First graders		Second graders		Third graders		Fourth graders	
	girls	boys	girls	boys	girls	boys	girls	boys
Total (M/V)	12 (12/0)	6 (6/0)	15 (11/4)	12 (7/5)	13 (7/6)	13 (5/8)	8 (1/7)	11 (1/10)

Note. (M/V) — the number of participants tested in Moscow / in Volgograd.

tically significant effect of the *city* factor on reading speed was found ($Est=6.681$, $SE=3.94$, $t=1.69$, $Pr(>|t|)=.094$). In a similar model, factors such as city, grade and gender were also included as predictors for the number of correctly answered questions. Again, we found no statistically significant effect of the *city* factor on the level of reading comprehension ($Est=-0.873$, $SE=0.575$, $t=-1.52$, $Pr(>|t|)=.133$).

All participants with developmental dyslexia ($N=50$, girls = 17; 1 left-hander; $M_{age}=8.9$, $SD=1.2$) were native Russian speakers. They were elementary school students, first through fourth graders, and, based on clinical assessment, had oral language skills typical for their age. The inclusion criterion was the confirmation of dyslexia by clinical specialists of the Center for Speech Pathology and Neurorehabilitation (Moscow) right before the study. This involved a certified speech therapist and neuropsychologist of the children's department assessing the children's speech development and other higher mental functions. It is worth noting that only children whose parents applied to the Center of Speech Pathology and Neurorehabilitation on their own initiative were invited to participate in the study. These parents reported persistent difficulties in their children with acquiring reading, which they tried to resolve either on their own or with the help of school or private speech therapists (this parameter varied), but did not achieve a persistent sufficient effect.

According to the neuropsychological assessment (Akhutina et al., 2016), the children with developmental dyslexia were heterogeneous in terms of the types of deficits observed. Some of the children had difficulties with processing visual and visual-spatial information, some children showed decreased skills in phonological processing, verbal memory, or the perception of rhythms, and some of the children displayed rapid exhaustion from activities related to the perception of written text. In some cases, a combination of two or more deficits with varying severity was observed. The quality of the children's writing ranged from nearly normal (when rewriting the text from the sample) to almost completely absent. Additional inclusion criteria were normal or corrected-to-normal vision, normal hearing, normal non-verbal intelligence (measured with the same instruments as in the typically developing children), and the absence of the history of neurological disorders.

Written informed consent forms were signed by parents or legal representatives of the children; children also orally agreed to participate. The study was approved by the Committee on Interuniversity Surveys and Ethical Assessment of Empirical Research, National Research

University Higher School of Economics, Russia (protocol No. 47 at 08.05.2018).

Materials and Procedure

Children's reading fluency and reading comprehension were assessed with the SARS (Kornev & Ishimova, 2010). For the initial testing, the authors recommended using the first halves of Texts I and II (three paragraphs from each text). "In secondary assessments it is possible to offer the child to read the second halves of texts I or II or to use the additional texts given in Appendix 2" (Kornev & Ishimova, 2010). Each half of the text has its own 10 questions for assessing reading comprehension. We used the first halves of Texts I and II ("How I caught crayfish" and "Ungrateful spruce") and their questions.

The first text (3 paragraphs, 97 words) was simpler; the second text (3 paragraphs, 127 words) was more difficult according to the authors of the methodology (Kornev & Ishimova, 2010). To quantify the level of difficulty of Text I and Text II, we analyzed the words included in terms of frequency (in instances per million) and length (in syllables and in letters). We used information on the frequency of words from the linguistic database for Russian words "StimulStat Project" (Alexeeva, Slioussar, & Chernova, 2017). When counting, we took into account all full-meaning words (by lemmas). Words that occur in the text several times were counted by the number of occurrences. The results of the analysis confirmed that Text I contains higher frequency words, while Text II uses longer words (see Table 2).

Table 2. Properties of Text I and Text II

	Text I	Text II
Average frequency of full-meaning lexical words (ipm)	2249.5	1102.0
Average length of full-meaning words in syllables	1.94	2.15
Average length of full-meaning words in letters	4.90	5.88

All children were informed in advance that after reading each text, they would have to answer questions to the text. Among our tested cohort of typically developing children there were no first graders who could not read the proposed texts. Considering that the SARS is the only standardized method for assessing reading skills in Russian, and that it is widely used by speech therapists, we cannot strictly claim that all children were seeing these texts for the first time. However, none of the children involved in the testing reported after reading that she or he had already read the text. These texts were not used to test reading skills when selecting children with dyslexia for our study.

According to the original procedure, children had to read the texts aloud and to answer ten content questions immediately after reading each of the texts. The vast majority of children could not read more than three paragraphs in one minute. If the child read slowly, we did not stop him until he had finished reading the third paragraph — that is, a fragment sufficient to answer all the questions. If the

child read quickly, we did not interrupt him at the end of reading the third paragraph, and waited until one minute had passed from the moment when he finished reading the title of the text, and only then stopped, to be able to count the number of words correctly read in one minute.

According to original guidelines, the examiner should listen to a child's reading and control the reading parameters using the form with the number of words printed at the end of each line. However, for some children it was distractive, and we audio-recorded children while they were reading. These records were further analyzed off-line.

Reading fluency was measured as the number of words read accurately in the first minute (the total number of words read minus the number of incorrectly read words (Kornev & Ishimova, 2010)). Reading comprehension was measured as the number of correct responses to the questions. When a participant could not read the second text at all (10 children with dyslexia), we considered reading fluency equal to 0, the level of reading comprehension equal to 0, and the number of incorrectly read words was marked as NA (no data).

For diagnostic purposes, the authors (Kornev, 1997, 2003; Kornev & Ishimova, 2010) suggested to transform raw reading fluency scores to a coefficient of reading technique (CoRT) by the following formula:

$$\text{CoRT} = 100 + ((M - m) / m) * 50,$$

where M is the number of words read correctly by a child in the first minute, m is the mean number of words read correctly in one minute by children of the same grade. The correspondence between reading fluency and CoRTs for particular grade was presented in the CoRTs table (Kornev & Ishimova, 2010). As an argument for converting direct data on children's reading fluency into CoRTs, the authors of the SARS mentioned the possibility of bringing the data to a standard scale that coincides in dimension with the IQ scale used in ABM-WISC, the adapted version of Wexler's test (Panasyuk, 1973). This was to provide the convenience of comparing the results of children tested with SARS and ABM-WISC. However, at present, Wexler's test for Russian-speaking children has limitations as a method of psychometry (Bazylchik, 2016). The comparison of reading test results with intelligence test results is extremely important for identifying children with dyslexia, but it is necessary to use methods with actual normative data. Considering the fact that normative data can be outdated, and that the restandardization of tests can take place at different rates, in our opinion, it is expedient to provide normative data on reading that indicate sample means and standard deviations, so that researchers and practitioners have the opportunity to choose the best intelligence tests available at a particular time. In our study we used the test of J. Raven, because it has more relevant norms for Russian-speaking children, which are given in the official guidelines (Raven et al., 2012) and which were collected with the participation of Russian-speaking primary schoolchildren (Davydov & Chmyhova, 2016).

The authors of the SARS recommended considering the performance within one standard deviation lower from the mean reading fluency for a particular grade as non-pathological forms of reading delay, which can be

presumably corrected by pedagogical efforts. In turn, reading fluency lower than one and a half standard deviations from the mean reading fluency was supposed to be considered as belonging to dyslexia spectrum. The values in between one and one and a half standard deviations would signal the risk of dyslexia. We used the above-mentioned recommendations to assess the validity of the SARS (Kornev & Ishimova, 2010).

Due to the lack of other standardized tests for reading in Russian, it was not possible to assess that type of *criterion validity*, which is calculated as the correlation coefficient between the scores on the test under study and an external measure — results of similar tests acknowledged as the “gold standard” (Ivanova & Hallowell, 2013). Therefore, we used a type of criterion validity known as *concurrent validity*, which is calculated by comparing it with another criterion — the presence or absence of a clinical diagnosis of developmental dyslexia in the child participant. We analyzed the data of 90 typically developing children and 50 children clinically diagnosed as having developmental dyslexia. We assessed such psychometric properties of the SARS as sensitivity (the percentage of children with clinically diagnosed developmental dyslexia who perform below a cutoff score for normal performance and who would be diagnosed as having dyslexia while using SARS), and specificity (the proportion of participants without developmental dyslexia who obtain results above the cutoff for normal reading abilities and who would be considered as typically developing based on the SARS; Ivanova & Hallowell, 2013). Additionally, we examined the correspondence between the clinical diagnosis and the performance of these children on the SARS using our new data for typically developing children.

Results

Reading Fluency

Table 3 shows the mean reading fluency for the typically developing children from grades 1 to 4. For comparison, we present the data published by the author of the SARS in 1997 (the data re-published in 2003 was the same as in 1997). The manual published in 2010 did not contain means and standard deviations for reading fluency, but reported the CoRT formula transformations of raw norms (see Method section). Therefore, we took the formula of CoRT from Kornev (2003), and based on the CoRT table in Kornev and Ishimova (2010), we calculated the average reading fluency for children from grades 2 to 4, which should have been used to get such CoRTs.

The authors of the SARS manual (Kornev & Ishimova, 2010) suggested that non-pathological forms of reading delay should be considered within one standard deviation lower from the mean reading fluency of the correspondent grade, and as belonging to dyslexia spectrum should be considered results lower than one and a half standard deviations from the mean (see Method section). We used the same cut-offs for our new data. It was not possible to reliably reconstruct standard deviations for the Kornev and Ishimova sample (2010), since the authors did not provide raw individual data in open sources. Therefore, we present

Table 3. Mean Reading Fluency

Grade	Mean Reading Fluency (Kornev, 1997, 2003)		Calculated Means from the CoRTs (Kornev & Ishimova, 2010)		Mean Reading Fluency from Our New Data	
	Text I	Text II	Text I	Text II	Text I	Text II
1	NA	NA	NA	NA	43.3	31.2
2	52.2	41.7	62.5	52.0	70.3	55.1
3	101.4	89.5	84.7	72.5	90.4	75
4	96.1	98.5	106.4	92.6	94.5	95.4

Note. NA — data were not provided.

the comparison of mean reading fluency and standard deviations between our normative data and the data published by Kornev (1997, the same in 2003) in Table 4. In Table 5, we present the reference levels for assessing reading fluency based on our data: normative, at risk, and dyslexia values.

Table 4. Comparison of Mean Reading Fluency and Standard Deviations of Reading Fluency Provided in Kornev (1997, 2003) and from Our Data

Grade	Kornev Data				Our New Data			
	Text I		Text II		Text I		Text II	
	Mean Reading Fluency	SD	Mean Reading Fluency	SD	Mean Reading Fluency	SD	Mean Reading Fluency	SD
1	NA	NA	NA	NA	43.3	13.5	31.08	10.1
2	52.2	11.2	41.7	14.9	70.3	16.6	55.1	15.1
3	101.4	29.3	89.5	23.7	90.4	15.8	75.0	17.1
4	96.1	27.7	98.5	27.8	94.5	7.8	95.4	15.2

Note. SD — standard deviation.

Reading Comprehension

The second important criterion for assessing reading skills is the level of reading comprehension. For this purpose, the author of the SARS offered 10 content questions to each of the texts. It was recommended to evaluate the results at 4 levels: A — 10 out of 10 correct answers, B — 7 to 9 correct answers, C — 4 to 6 correct answers, and D — 3 or less correct answers. Levels A and B should be considered

Table 5. Reference Levels for Assessing Reading Fluency (Based on Our New Data)

Typically Developing Children, N=90						
Grade	Text I			Text II		
	Mean (typical reading)	Mean -1 SD (at risk)	Mean -1.5 SD (dyslexia)	Mean (typical reading)	Mean -1 SD (at risk)	Mean -1.5 SD (dyslexia)
1	43.3	29.8	23.2	31.8	21.7	16.6
2	70.3	53.7	45.4	55.1	40.0	32.4
3	90.4	74.6	66.6	75.0	57.9	49.2
4	94.5	86.7	82.8	95.4	80.2	72.5

Note. *SD* — standard deviation.

as “functionally sufficient”, that is allowing to use the full information contained in a text. However, mean data for the number of correctly answered questions to the texts in the typically developing group were never published by Kornev and Ishimova, thus we cannot compare the results of reading comprehension based on our data with their findings.

In Table 6, we present reference levels for assessing reading comprehension based on the performance of 90 TD children in our tested cohort: normative, at risk, and dyslexia values. These data show that even in the fourth grade, typical children with normal non-verbal intelligence give an average of about 5 correct responses to 10 questions for Text II (the younger ones give even fewer correct responses). Therefore, the original criterion recommended by the SARS manual (Kornev & Ishimova, 2010), that is, 7 to 10 correct answers, can hardly serve as a basis for diagnosing reading comprehension impairment.

Table 6. Reference Levels for Assessing Reading Comprehension (Based on Our New Data)

Typically Developing Children, N=90						
Grade	Text I			Text II		
	Mean (typical reading)	Mean -1 SD (at risk)	Mean -1.5 SD (dyslexia)	Mean (typical reading)	Mean -1 SD (at risk)	Mean -1.5 SD (dyslexia)
1	6.17	3.9	2.7	3.6	1.3	0.2
2	7.0	5.5	4.8	4.3	2.2	1.1
3	7.8	6.3	5.6	4.9	2.7	1.6
4	7.3	5.9	5.3	5.3	3.4	2.5

Note. *SD* — standard deviation.

Validity of the SARS

As described in the Method section, we examined the validity of the SARS (Kornev & Ishimova, 2010) for the diagnosis of developmental dyslexia, calculating psychometric properties of the test such as sensitivity and specificity using original norms provided in the last published manual for the test (Kornev & Ishimova, 2010), and our new data for typically developing children.

Specificity

To assess the specificity of the test, we analyzed what percentage of typically developing children would fall into the typically developing group based on the SARS (Kornev & Ishimova, 2010). Since the normative data on the SARS were published only for the second and subsequent grades, we could not evaluate the results of 18 first-graders from our sample. Among 72 children in our group of typically developing schoolchildren in grades 2 to 4, there was not a single child whose CoRT would fall into the dyslexia group (that is, below 1.5 *SDs*), according to the CoRT table published in Kornev and Ishimova (2010). In 69 children, the CoRT values were within 1 *SD* (typical reading), and in three children (one fourth grader and two second graders) — between 1 and 1.5 *SDs* (risk group). According to these results, no child would be erroneously diagnosed as having dyslexia, and specificity of the SARS can be assessed as 100 %.

It is worth noting that in the absence of screening for non-verbal intelligence, vision and hearing, non-specific reading disorders can be erroneously attributed to dyslexia, while the problem may be caused by another comorbid impairment (Snowling, Nash, Gooch, Hayiou-Thomas, & Hulme, 2019). Our initial cohort contained 15 children (further excluded from the analysis) with no dyslexia diagnosis, but with registered hearing impairments or performing below norms on the non-verbal intelligence test (Raven, 2004). In case we did not screen them for these exclusion criteria, four children (two children for both texts and two other children for Text II) could be erroneously attributed to the dyslexia group based on their CoRT values, which would reduce the specificity of the test.

We checked how the specificity of the test would change when using diagnostic criteria based on our new data. Out of 90 students of grades 1 to 4 from our group of typically developing schoolchildren, the reading performance of five children (two second graders, one child from the third grade and two children from the fourth grade) were below the level of 1.5 standard deviations from the average. That is, these children would fall into the dyslexia group, which would reduce the specificity of the test from 100 % to 94.4 %.

Sensitivity

To assess the sensitivity of the test, we analyzed what percentage of children with clinically diagnosed developmental dyslexia fell into the dyslexia group based on the SARS (Kornev & Ishimova, 2010). Nine children of the first grade could not be assessed, since the normative data for the SARS were only published starting from the second grade. Table 7 shows the percentage distribution of the remaining 41 children of grades 2 to 4 among typically reading, at risk and dyslexia groups, based on the CoRT for Text I and Text II.

Table 7. Distribution of Children with Clinically Diagnosed Dyslexia into Groups Based on the SARS and Norms from Kornev and Ishimova (2010) (Grades 2 to 4)

N=41	Typical Reading Fluency, %	At Risk, %	Dyslexia, %
Text I	51.2	12.2	36.6
Text II	43.9	17.1	39.0

Thus, according to the reading performance in Text I, 15 out of 41 children (from grades 2 to 4) with clinically diagnosed dyslexia would be classified as having dyslexia using normative data published in 2010. This gives a test sensitivity of 36.6%. According to the reading performance in Text II, 16 children would be classified as having dyslexia, which corresponds to a test sensitivity of 39.0%.

We investigated how the situation would have changed if we used the cutoff levels according to our new data obtained in 2018 (summarized in Table 5). Table 8 shows the percentage distribution of all tested children with clinically diagnosed dyslexia into groups based on the SARS performance in reading test and according to the new data for Text I and Text II.

Table 8. Distribution of Children with Clinically Diagnosed Dyslexia into Groups Based on the SARS and Our New Data for Reading Fluency (Grades 1 to 4)

N=50	Typical reading fluency, %	At risk, %	Dyslexia, %
Text I	28.0	0.0	72.0
Text II	26.0	14.0	60.0

With the new cutoff values for reading Text I, 34 out of 50 children with clinically diagnosed dyslexia would be classified as having dyslexia using our normative data. This gives a test sensitivity of 72.0%. According to the reading fluency of Text II, 30 out of 50 children with a diagnosis of developmental dyslexia would be classified as dyslexic, which corresponds to a sensitivity of 60.0%.

We also examined if the sensitivity value of the SARS would have changed if, in addition to the reading performance, reading comprehension would have been assessed. We used the same diagnostic cutoff values as were suggested by original guidelines for reading fluency: one standard deviation from the mean for the at-risk performance, and one and a half standard deviations for dyslexic performance. We used only our new data (see Table 6), because the authors of the SARS have not published normative data for reading comprehension. In our tested cohort, all participants with impaired reading comprehension had impaired reading fluency, but there were 13 children who had typical comprehension combined with impaired reading fluency. Table 9 shows the percentage distribution of all tested children with clinically diagnosed dyslexia among groups based on the SARS performance in

reading comprehension and according to our data for Text I and Text II.

Table 9. Distribution of Children with Clinically Diagnosed Dyslexia Among Groups Based on the SARS and Our New Data for Reading Comprehension (Grades 1 to 4)

N=50	Typical reading comprehension, %	At Risk, %	Dyslexia, %
Text I	78.0	2.0	20.0
Text II	62.0	0.0	38.0

Among 50 children with clinically diagnosed dyslexia, 10 performed below 1.5 SD answering the questions to Text I. These children would be diagnosed as dyslexic if only reading comprehension performance were used as diagnostic criterion. This gives a test sensitivity of 20%. 19 children performed below that cutoff value answering the questions to Text II, which corresponds to a test sensitivity of 38%.

We performed an additional analysis in search of an explanation for why 14 children (28% of all dyslexic children tested), who were within the normative range in terms of reading fluency and reading comprehension performance based on the SARS, were clinically diagnosed as dyslexic. For that, we listened again to the audio recordings of children's reading and weighted types of errors using a system of penalty points. Kornev and Ishimova (2010) claimed that

Table 10. System of Penalty Points for Different Types of Errors

Types of Errors	Penalty Points	
	Without Self-Correction	With Self-Correction
Wrong stress in a word	1	0.5
One sound skipped, changed or added	1	0.5
2–3 sounds swapped	2	1
2–3 sounds skipped, changed or added	2	1
One word skipped	2	1
One word repeated	2	1
One functional word added	2	1
One content word added	3	1.5
One word changed (4 or more sounds changed in a word)	3	1.5
Lost line with disorientation	3	1.5

“the errors that made children with dyslexia do not qualitatively differ from the errors that are normally observed in beginner readers”. In our tested cohort children made errors at different levels: from minor, such as repeating one sound or reducing sounds in a weak position, to major, such as saying another word instead of the one presented, or losing a line and disorienting. Table 10 presents the types of errors that occurred in the typically developing and dyslexic groups, along with the system of penalty points for these errors.

For each child, we calculated the total weighted score for errors by summing the penalty points for each error made. The results of 90 typically developing children are shown in Table 11. Then, we calculated the penalty points for each of those 14 children with a clinical diagnosis of dyslexia but no reading deficit identified by the main scores of the SARS. We found that for all of these children, the penalty points were more than 1.5 standard deviation higher than the means for the corresponding grade for at least one of the texts. In other words, these children read fluently but with a lot of errors of greater weight. Thus, they have an impairment of reading quality, not speed, which remains unnoticed when using the original scoring criteria for the SARS (Kornev & Ishimova, 2010).

Table 11. Reference Levels for Evaluation the Weighted Error Scores (Based on Our New Data)

Typically Developing Children, <i>N</i> =90						
Text I			Text II			
Grade	Mean Error Score	Mean + 1 SD	Mean + 1.5 SD	Mean Error score	Mean + 1 SD	Mean + 1.5 SD
1	2.75	4.61	5.54	2.72	4.70	5.69
2	3.77	7.15	8.84	4.44	7.69	9.32
3	3.40	5.92	7.18	5.54	9.56	11.57
4	2.95	5.90	7.38	5.76	9.52	11.4

Note. *SD* — standard deviation.

We explored whether the sensitivity of the SARS would change if, in addition to reading fluency, reading quality was assessed taking into account a new criterion — weighted penalty points for errors. We used the same diagnostic cutoff levels that were proposed in the original manual for reading speed: one standard deviation from the mean for the results considered as a risk group, and one and a half standard deviations from the mean for the results related to dyslexia. We used our data for 50 children with dyslexia for Text I, but for 40 children with dyslexia for Text II, since 10 children with dyslexia (20% of participants with dyslexia) could not read Text II at all, and there are no data about errors in the second text for them. Table 12 shows the percentage distribution of all tested children with clinically diagnosed dyslexia into groups based on weighted error scores in the SARS in accordance with our data for Texts I and II.

Table 12. Distribution of Children with Clinically Diagnosed Dyslexia into Groups Based on Weighted Error Scores in SARS (in Accordance with Our Data, Grades 1 to 4)

	Weighted Error Scores Within Normative Range, %	At Risk, %	Dyslexia, %
Text I (<i>N</i> =50)	30.0	6.0	64.0
Text II (<i>N</i> =40)	7.5	15.0	77.5

Among 50 children with clinically diagnosed dyslexia, 32 showed results below the cutoff level of 1.5 standard deviations when reading Text I. These children would be diagnosed as having developmental dyslexia if only weighted errors scores were used as a diagnostic criterion. This gives a test sensitivity of 64%. Among the 40 children with dyslexia who could read the second text, 31 children showed results below the cutoff level, which corresponds to a test sensitivity of 77.5%.

Discussion

Standardized tests should be used with actual normative data for diagnostic purposes, because normative performance can change over time (Raven, 2000). The fact that normative data need to be updated concerns not only this particular methodology but is characteristic of most standardized tests. In the methodical manual published in 2010 (Kornev & Ishimova, 2010), the authors also noted the reasonability of normative data updating. Importantly, for appropriate cutoff scores for normal performance it is preferable to test not randomly chosen children, but those who match such inclusion criteria as normal intellectual abilities, vision and hearing, and the absence of diagnosed neurological disorder (Ivanova & Hallowell, 2013).

The data we collected for the SARS in 2018 provide an update of reference levels for assessing reading performance and are controlled for the mentioned inclusion criteria for the tested normative sample. Although the number of typically developing children participating in our study (*N*=90) is not enough for complete restandardization of the test, the fact that the norms calculated on a sample of 90 children were diagnostically more productive than the norms from the author’s manual (Kornev & Ishimova, 2010) to identify a clinical phenomenon, suggests that it is necessary to review the normative levels.

As for the diagnostic power of the SARS, the specificity of the test is perfect. That means that any normative data set of SARS allows the identification of typical reading performance and the absence of dyslexia in a new cohort with high accuracy: from 100% (when using normative data obtained in 2007–2008) to 94.4% (when using reference levels, calculated on the data of 2018). However, sensitivity of the SARS substantially depends on the normative data used for cutoff values. When normative data obtained in 2007–2008 were used, sensitivity of the test lied between 36.6% and 39.0% (depending on the text), while using our new normative data obtained in 2018, the sensitivity of the test reached 60.0–72.0%. A test sensitivity of even 72.0% means that 28.0% of children with impaired reading skills

may be erroneously not diagnosed as such. As it was shown in many studies, the consequence of the non-recognition of dyslexia is the lack of adequate intervention, leading to anxiety and depressive behaviors (Mugnaini, Lassi, La Malfa, & Albertini, 2009; Willcutt & Pennington, 2000; Nelson & Harwood, 2011; Törő, Miklósi, Horanyi, Kovács, & Balázs, 2018), suicidal ideation, school failure, and drop out (Barbiero et al., 2019; Daniel et al., 2006; Wilson, Armstrong, Furrie, & Walcot, 2009).

The relatively low sensitivity of the SARS is a result of using only one (albeit the most important) diagnostic criterion — the speed of decoding. However, the tool developed by Kornev and Ishimova in fact allows us to use two additional diagnostic criteria: the level of reading comprehension, and weighted errors scores. To further improve the clinical validity of the SARS, we provide reference levels for assessing reading comprehension and weighted error scores based on the results of our tested cohort of typically developing children.

Our study showed that the insertion of a weighted error assessment as an additional diagnostic criterion can improve the diagnostic validity of the SARS. We were able to calculate the penalty points for errors due to the changes in the testing procedure (we audio recorded children's voice while reading). Considering that this innovation solves the problem of children's distracted attention and allows for more accurate assessment of reading parameters, we can recommend the inclusion of an audio recording of the children's voices while reading in the diagnostic procedure.

As for the assessment of reading comprehension (which is carried out in the SARS by means of content questions which children answer after reading the texts), the results of these tests are rather variable, but on the whole are quite low even in the typically developing group. This leads to the fact that only a very low level of reading comprehension is diagnosed as impairment (see Table 6, as well as comments on page 16 in the author's manual of Kornev and Ishimova, 2010). Actual data show that the application of a reading comprehension criterion does not lead to an increase in the overall sensitivity of the test. This suggests either that this test can be useful only in specific cases, when a child has a very low level of reading comprehension combined with normal reading fluency and accuracy (in our sample we had no such children), or that the diagnostic power of this task is doubtful. However, since reading comprehension is extremely important for schooling in general, future research might investigate variants of more effective reading comprehension tests.

It is important that three aforementioned deficits (slow reading, lack of understanding of what is read, and numerous major errors while reading) measured with the SARS' metrics (reading fluency, reading comprehension and weighted error scores, correspondingly) may occur in children with dyslexia both in isolation and in combination. Our study showed that children with only one type of deficit — those with high speed of reading and normal reading comprehension, but with a large number of errors — were still clinically diagnosed as having developmental dyslexia.

We propose the use of not a single criterion but two or better yet three criteria for diagnostic purposes — reading speed, reading comprehension, and weighted error scores —

since none of the criteria separately provides test sensitivity close to 100 % (see the Results section). Reference levels are presented in Table 5, Table 6 and Table 11, correspondingly. If a child's performance is one and a half standard deviation lower than the mean for the same grade, developmental dyslexia should be diagnosed. Additionally, screening for non-verbal intellectual abilities, hearing and vision should necessarily accompany reading abilities testing, since these measures interact with the assessment of reading. It is worth noting that western researchers also use tests for reading words (see for example Brus & Voeten, 1973) and for reading pseudowords (see for example van den Bos, Spelberg, Scheepstra, & DeVries, 1994). The development of such standardized instruments for the Russian language and investigation of their effectiveness in the diagnosis of dyslexia is one of the possible directions for future studies.

In conclusion, we express our great respect to A.N. Kornev and O.A. Ishimova for developing the only available standardized test for the assessment of reading skills in Russian-speaking children. We hope that our new data, their clinical evaluation, and suggested amendments for scoring the results (concerning a more precise evaluation of reading comprehension and the introduction of weighted error scores) will expand the efforts of the authors of the methodology and will contribute to an even wider use of the SARS.

References

- Akhutina, T. V., Korneev, A. A., Matveeva E. Yu., Romanova, A. A., Agris, A. R., Polonskaya, N. N., Pylaeva, N. M., Voronova, M. N., Maksimenko, M. Y., Yablokova, L. V., Melikyan, Z. A., Kuzeva, O. V. (2016). *Metody neyropsikhologicheskogo obsledovaniya detey 6–9 let [Methods of neuropsychological examination of children from 6 to 9 years old]*. Moscow: Sekachev. (In Russian).
- Alexeeva, S., Slioussar, N., & Chernova, D. (2018). StimulStat: A lexical database for Russian. *Behavior Research Methods*, 50(6), 2305–2315. doi:10.3758/s13428-017-0994-3
- Barbiero, C., Montico, M., Lonciari, I., Monasta, L., Penge, R., Vio, C., ..., Ronfani, L. on behalf of the EpiDit (Epidemiology of Dyslexia in Italy) working group (2019). The lost children: The underdiagnosis of dyslexia in Italy. A cross-sectional national study. *PLoS One*, 14(1), e0210448. doi:10.1371/journal.pone.0210448
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bazyltchik, S. V. (2016). Prigodnost rusifitsirovannykh versiy detskogo testa Vekslera (WISC) dlya diagnostiki umstvennoy otstalosti [Applicability of Russian version of the children's test Wechsler (WISC) for the diagnosis of mental retardation]. *Psychiatry, Psychotherapy and Clinical Psychology*, 2016(2), 12–20. (In Russian). Retrieved from <https://elibrary.ru/item.asp?id=26140287>.
- van den Bos, K. P., Spelberg, H. C. L., Scheepstra, A. J. M., & DeVries, J. R. (1994). *De Klepel. Vorm A en B. Eentestvoordeleesvaardigheid van pseudoworden. Verantwoording, handleiding, diagnostiekenbehandeling [The Clapper. Form A and B. A test for reading the pseudowords. Accountability, manual, diagnostics treatment]*. Nijmegen, the Netherlands: Berkhout. (In Dutch).
- Brus, B. T., & Voeten, M. J. (1973). *Eén-minuut-test vorm A en B. Verantwoording en handleiding [One-minute word reading test version A and B. Justification and manual]*. Nijmegen, the Netherlands: Berkhout. (In Dutch).
- Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout,

- and reading problems among adolescents. *Journal of Learning Disabilities*, 39(6), 507–514. doi:10.1177/00222194060390060301
- Davydov, D. G., & Chmykhova, E. V. (2016). Primenenie testa Standartnye progressivnye matritsy Ravena v rezhime ogranicheniya vremeni [Administration of the Raven's Standard Progressive Matrices with a time limit]. *Voprosy Psikhologii*, 2016(4), 129–139. (In Russian).
- Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, 27(8), 891–920. doi:10.1080/02687038.2013.805728
- Kornev, A. N. (1997). *Narusheniya chteniya i pisma u detey. Uchebno-metodicheskoe posobie [Reading and writing impairments in children]*. Saint Petersburg: MiM. (In Russian).
- Kornev, A. N. (2003). *Narusheniya chteniya i pisma u detey. Uchebno-metodicheskoe posobie [Reading and writing impairments in children]*. Saint Petersburg: Rech. (In Russian).
- Kornev, A. N., & Ishimova, O. A. (2010). *Metodika diagnostiki disleksii u detey. Metodicheskoe posobie [Methods of diagnosis of dyslexia in children. Methodical manual]*. Saint Petersburg: Publishing House of the Polytechnic University. (In Russian).
- Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest [Reading and Spelling Test]*. Bern: Hans Huber. (In German).
- Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics*, 5(4), 255–264. doi:10.1007/s12519-009-0049-7
- Nelson, J. M., & Harwood, H. (2011). Learning disabilities and anxiety: A meta-analysis. *Journal of Learning Disabilities*, 44(1), 3–17. doi:10.1177/0022219409359939
- Panasyuk, A. Y. (1973). *Adaptirovanny variant metodiki D. Wekslera WISC [Adapted version of D. Wechsler's WISC]*. Moscow: Meditsina. (In Russian).
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. doi:10.1006/cogp.1999.0735
- Raven, J. (2004). *Tsvetnye progressivnye matritsy serii A, Ab, B [Raven's progressive matrices]*. Moscow: Cogito-Center. (In Russian).
- Raven, J., Raven, J. K., & Kort, J. (2012). *Rukovodstvo k Progressivnym Matritsam Ravena i Slovarnym Shkalam. Razdel 3. Standartnye Progressivnye Matritsy (vkluyuchaya Paralelnye i Plyus versii) [Guidelines to Raven's Progressive Matrices and Vocabulary Scales. Section 3. Standard Progressive Matrices (including Parallel and Plus versions)]*. Moscow: Cogito-Center. (In Russian).
- Semenovich, A. V. (2002). *Neyropsikhologicheskaya diagnostika i korrektsiya v detskom vozraste [Neuropsychological assessment and intervention in childhood]*. Moscow: Academia. (In Russian).
- Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., & Wellcome Language and Reading Project Team (2019). Developmental outcomes for children at high risk of dyslexia and children with developmental language disorder. *Child Development*, 90(5), e548–e564. doi:10.1111/cdev.13216
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency (TOWRE)*. Austin, TX: PRO-ED.
- Törő, K. T., Miklósi, M., Horanyi, E., Kovács, G. P., & Balázs, J. (2018). Reading disability spectrum: Early and late recognition, subthreshold, and full comorbidity. *Journal of Learning Disabilities*, 51(2), 158–167. doi:10.1177/0022219417704169
- Wechsler, D. (1990). *Wechsler objective reading dimensions*. London: The Psychological Corporation.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.
- Willcutt, E. G., & Pennington, B. F. (2000). Psychiatric Comorbidity in Children and Adolescents with Reading Disability. *Journal of Child Psychology and Psychiatry*, 41(8), 1039–1048. doi:10.1017/S0021963099006368
- Wilson, A. M., Deri Armstrong, C., Furrie, A., & Walcot, E. (2009). The mental health of Canadians with self-reported learning disabilities. *Journal of Learning Disabilities*, 42(1), 24–40. doi:10.1177/0022219408326216
- Wimmer, H., & Mayringer, H. (2014). *SLS 2–9: Salzburger Lese-Screening für die Schulstufen 2–9 [Salzburg Reading and Spelling Test]*. Bern: Hans Huber. (In German).
- Woodcock, R. (1999). *Woodcock Reading Mastery Test–Revised*. Circle Pines, MN: American Guidance Services.

МЕТОДЫ

Оценка диагностической валидности Стандартизированной методики исследования навыков чтения на русском языке и проверка актуальности имеющихся нормативных данных

Светлана Валентиновна Дорофеева

Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Виктория Алексеевна Решетникова

Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Маргарита Никитична Серебрякова

Центр патологии речи и нейрореабилитации, Москва, Россия

Дария Николаевна Горанская

Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Татьяна Васильевна Ахутина

Московский государственный университет имени М. В. Ломоносова, Москва, Россия

Ольга Викторовна Драгой

Национальный исследовательский университет «Высшая школа экономики», Москва, Россия;
Федеральный центр цереброваскулярной патологии и инсульта, Москва, Россия

Аннотация. Стандартизированные тесты с нормативными данными стали золотым стандартом для оценки навыков чтения и для диагностики специфического расстройства чтения (дислексии) для многих языков. Для русского языка имеется такой тест для исследования чтения — Стандартизированная методика исследования навыков чтения, разработанная А. Н. Корневым и впервые опубликованная в 1997-м году. Однако наиболее свежие доступные нормативные данные по этой методике были собраны более десяти лет назад, и они не были проконтролированы по нескольким важным параметрам. Также не было опубликовано никаких деталей по поводу диагностической валидности теста. Мы использовали этот тест для оценки навыков чтения у девятилетнего типично развивающегося российского ученика начальной школы в 2018-м году. В данной статье мы представляем современные результаты тестирования типично развивающихся детей, включая обновленные уровни по скорости чтения и, впервые, метрики для оценки уровня понимания прочитанного и взвешенной оценки ошибок. Дополнительно мы протестировали пятьдесят детей с клинически диагностированной дислексией и сообщаем информацию о чувствительности и специфичности Стандартизированной методики исследования навыков чтения.

Контактная информация: Светлана Валентиновна Дорофеева, sdorofeeva@hse.ru, 105066, Москва, ул. Старая Басманная, 21/4, каб. 510, Национальный исследовательский университет «Высшая школа экономики»; Виктория Алексеевна Решетникова, victoriaregina98@gmail.com; Маргарита Никитична Серебрякова, margoha1986@yandex.ru; Дария Николаевна Горанская, daria.goranskaya@gmail.com; Татьяна Васильевна Ахутина, akhutina@mail.ru; Ольга Викторовна Драгой, odragoy@hse.ru.

Ключевые слова: скорость чтения, понимание прочитанного, русский язык, дислексия, СМИНЧ, Стандартизированная методика исследования навыков чтения

Благодарности. Мы благодарим за помощь, оказанную Л. А. Пушкаревой и Е. И. Ларионовой (психолог и директор школы № 556 Южного округа г. Москвы, Россия), А. В. Корниловым (руководитель подразделения 968 школы № 1236 имени С. В. Милашенкова, Москва, Россия), Н. С. Воловик и М. А. Сукиасяном (помощник директора и директор общеобразовательной школы «Поколение», Волгоград, Россия), а также Р. А. Череминим (главный врач Центра патологии речи и нейрореабилитации, Москва, Россия). Мы благодарны всем детям, которые приняли участие в этом исследовании, и их родителям. Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проект № 17-29-09122.

© 2019 Светлана Валентиновна Дорофеева, Виктория Алексеевна Решетникова, Маргарита Никитична Серебрякова, Дария Николаевна Горанская, Татьяна Васильевна Ахутина, Ольга Викторовна Драгой. Данная статья доступна по лицензии [Creative Commons "Attribution" \(«Атрибуция»\) 4.0. всемирная](https://creativecommons.org/licenses/by/4.0/), согласно которой возможно неограниченное распространение и воспроизведение этой статьи на любых носителях при условии указания автора и ссылки на исходную публикацию статьи в данном журнале в соответствии с канонами научного цитирования.

Статья поступила в редакцию 15 января 2019 г. Принята в печать 20 марта 2019 г.

Введение

Стандартизированные тесты необходимы для надежной и валидной количественной оценки способностей к чтению, а также для сравнения групп и отдельных лиц с целью диагностики нарушений чтения. Тесты на чтение с единым протоколом проведения и нормативными данными, полученными при тестировании большой выборки людей, стали методическим золотым стандартом в последние десятилетия. Комплексные тесты на чтение и письмо были разработаны для немецкого языка (Moll, Landerl, 2010; Wimmer, Mayringer, 2014), стандартизированный тест на чтение слов (Brus, Voeten, 1973) и стандартизированный тест на чтение псевдослов (van den Bos et al., 1994) были разработаны для голландского языка. Для английского языка доступно несколько вариантов стандартизированных инструментов для оценки чтения, включая специализированные батареи тестов на чтение (Torgesen et al., 1999; Wechsler, 1990; Woodcock, 1999) и субтесты на чтение в составе комплексных инструментов когнитивной оценки, таких как Wide Range Achievement Test (Wilkinson, Robertson, 2006). Стандартизированный тест на чтение доступен и для русского языка.

Стандартизированная методика исследования навыков чтения (СМИНЧ) для русскоязычных детей была разработана в 1982 году, впервые опубликована в 1997 году (Корнев, 1997) и переиздана в 2003 году (Корнев, 2003). Второй раунд сбора нормативных данных был проведен в 2007–2008 годах (Корнев, Ишимова, 2010), чтобы отразить изменения в средней производительности чтения у детей, которые происходят с течением времени. Эта методика, согласно утверждению автора, предназначена для «диагностики нарушений чтения и дифференциальной диагностики

ки дислексии и неспецифических нарушений чтения» (Корнев, Ишимова, 2010, с. 7). Другие существующие для русского языка инструменты оценки чтения являются составной частью протоколов комплексного нейропсихологического обследования (Ахутина и др., 2016; Семенович, 2002), но их диагностическое применение ограничено из-за отсутствия опубликованных нормативных данных.

СМИНЧ имеет ряд очевидных преимуществ. Во-первых, во время рестандартизации количество детей, протестированных для сбора нормативных данных, было значительно увеличено. Первое издание (Корнев, 1997) было опубликовано с нормативными данными, полученными в результате тестирования 150 школьников. Текущее издание (Корнев, Ишимова, 2010) было опубликовано с нормативными данными, полученными в результате тестирования 700 школьников. Во-вторых, этот тест на чтение дает возможность измерить не только скорость чтения, но и уровень понимания прочитанного, для этого используются вопросы о содержании прочитанного с четкими рекомендациями для оценки правильности ответов. В-третьих, он включает в себя тексты различной сложности (простой текст I и более сложный текст II), что позволяет проводить сравнительный анализ результатов тестируемого в зависимости от сложности текста. Корнев и Ишимова (2010, с. 13) сообщают, что разница в 10 и более баллов между показателями чтения двух текстов указывает на большую выраженность нарушений чтения. Наконец, СМИНЧ включает дополнительные тексты, предназначенные для повторного тестирования при необходимости, с уровнями сложности, сопоставимыми с первыми двумя текстами, а также с нормативными данными 400 московских школьников, протестированных в 2009–2010 гг.

Тем не менее СМНЧ в ее нынешнем виде имеет ограничения. Случайно выбранные дети из государственных школ принимали участие в тестировании для сбора нормативных данных (Корнев, Ишимова, 2010), хотя для определения надлежащих уровней нормативной производительности предпочтительно включать в исследование только детей без диагностированных неврологических расстройств или умственной отсталости (Ivanova, Hallowell, 2013). Кроме того, несмотря на наличие вопросов о содержании прочитанного для оценки уровня понимания и хорошо разработанные критерии для подсчета ответов как правильных или неправильных, авторы не опубликовали нормативные данные для понимания прочитанного. Они предложили рассматривать как «функционально достаточный» такой уровень производительности, при котором дети дают как минимум 7 верных ответов на 10 вопросов. В то же время они предоставили информацию о распределении протестированных в 2007–2008 годах 700 московских школьников по уровням понимания прочитанного. Из этих данных авторов методики можно рассчитать, что 53% и 69% второклассников правильно ответили менее чем на 7 вопросов к тексту I и тексту II соответственно (Корнев, Ишимова, 2010, с. 49). Таким образом, предложенная рекомендация не может служить руководством для диагностики расстройства понимания прочитанного. Для оценки понимания прочитанного требуются нормативные баллы, аналогичные таковым для оценки скорости чтения.

Что еще более важно для диагностических целей, в последнем издании СМНЧ (Корнев, Ишимова, 2010) прямые нормативные данные, например средняя скорость чтения и стандартное отклонение, не были представлены. Вместо этого авторы предложили преобразование по специальной формуле, в результате чего были получены коэффициенты техники чтения (см. раздел «Метод»). Эти коэффициенты были представлены не для каждого возможного числа правильно прочитанных слов, но с пропусками в три пункта. В результате, когда производительность конкретного ребенка находится на границе (например, между «группой риска» и «дислексией»), нет однозначного ответа, к какой группе следует отнести результат. Наконец, Корнев (2003, с. 213) утверждал, что «апробация на клиническом материале показала достаточную валидность методики», но при этом психометрические свойства СМНЧ не были представлены (например, размер выборки, чувствительность и специфичность теста). В результате диагностика дислексии у русскоязычных детей на основе этого теста остается сложной задачей для клинической практики.

Мы провели исследование, направленное на преодоление некоторых текущих ограничений СМНЧ и улучшение возможностей ее клинического применения. Во-первых, для проверки актуальности имеющихся нормативных данных мы собрали новые данные по СМНЧ, протестировав типично развивающихся российских школьников, у которых не было диагностированных неврологических нарушений, с измеренным ненарушенным невербальным интеллектом и нормальным слухом, а также с нормальным или скоррек-

Таблица 1. Общая информация о типично развивающихся участниках исследования

	Первоклассники		Второклассники		Третьеклассники		Четвероклассники	
	Девочки	Мальчики	Девочки	Мальчики	Девочки	Мальчики	Девочки	Мальчики
<i>N</i> = 90								
Всего (М/В)	12 (12/0)	6 (6/0)	15 (11/4)	12 (7/5)	13 (7/6)	13 (5/8)	8 (1/7)	11 (1/10)

Примечание. (М/В) — количество участников, протестированных в Москве / в Волгограде.

тированным до нормального зрением. В отличие от авторов СМНЧ, которые предусмотрели нормы для 2–6-го классов российской школьной системы, мы также протестировали детей 1-го класса (во второй половине учебного года), чтобы можно было идентифицировать детей с нарушениями чтения или с риском развития нарушений чтения как можно раньше. В результате мы сообщаем новые прямые данные (средние значения и стандартные отклонения) по скорости чтения и пониманию прочитанного у типично развивающихся российских школьников для каждого из классов начальной школы.

Кроме того, мы оценили валидность СМНЧ для диагностики дислексии. Для этого мы протестировали 50 детей, у которых была клинически диагностирована дислексия, и проанализировали соответствие между клиническим диагнозом и показателями этих детей по СМНЧ, используя оригинальные нормы, приведенные в последнем опубликованном руководстве к тесту (Корнев, Ишимова, 2010), и наши новые данные.

Метод

Участники

Типично развивающиеся участники были набраны в трех российских государственных школах, две из которых находятся в Москве ($n = 58$) и одна в Волгограде ($n = 47$). Все дети (всего $N = 105$) были носителями русского языка в возрасте от 7 до 11 лет и были учениками начальной школы с первого по четвертый классы. У детей не было истории диагностированных неврологических расстройств, не было диагностированных проблем с освоением чтения, все имели нормальное или скорректированное до нормального зрение (как сообщается в информированных согласиях, данных их родителями или законными представителями). Скрининг на первичные нарушения слуха (с использованием программы Audiogramm версии 4.6.1.3, профессиональной аудиометрической системы; аудиометрических наушников Sennheiser HDA 280) привел к исключению трех участников. Скрининг на уровень невербального интеллекта с помощью цветных прогрессивных матриц Равена (Равен, 2004; Равен и др.,

2012) привел к исключению 12 участников, набравших баллы ниже нормы. Девяносто детей (48 девочек; 50 детей из Москвы; 7 левшей; $M_{age} = 8.7$, $SD = 1.13$) остались в анализе. Общая информация о типично развивающихся участниках, включенных в анализ, представлена в таблице 1.

Прежде чем использовать результаты тестирования детей из разных городов в общем анализе, мы проанализировали влияние эффекта города на показатели чтения с учетом класса и пола участников. Для этого мы использовали линейные регрессионные модели, которые были построены при помощи пакета *lme4* для программы статистической обработки данных *R* (Bates et al., 2015). В модель в качестве предикторов для скорости чтения были включены такие факторы, как город, класс и пол протестированных детей. Статистически значимого эффекта влияния фактора *город* на скорость чтения обнаружено не было ($Est = 6.681$, $SE = 3.94$, $t = 1.69$, $Pr(>|t|) = .094$). В аналогичную модель в качестве предикторов для количества правильных ответов на вопросы были также включены такие факторы, как город, класс и пол протестированных детей. Статистически значимого эффекта влияния фактора *город* на уровень понимания прочитанного также не было обнаружено ($Est = -0.873$, $SE = 0.575$, $t = -1.52$, $Pr(>|t|) = .133$).

Все участники с дислексией ($N = 50$, девочек 17; 1 левша; $M_{age} = 8.9$, $SD = 1.2$) были носителями русского языка. Они были учащимися начальной школы, с первого по четвертый классы, и, основываясь на клинической оценке, имели устные языковые навыки, типичные для их возраста. Критерием включения было подтверждение наличия дислексии клиническими специалистами Центра патологии речи и нейрореабилитации (Москва) непосредственно перед исследованием. Для этого сертифицированным логопедом и нейропсихологом детского отделения проводилось исследование речи и других высших психических функций. Стоит отметить, что для отбора в исследование приглашались дети, чьи родители по собственной инициативе обращались в Центр патологии речи и нейрореабилитации. Эти родители сообщали о стойких трудностях у их детей с освоением чтения, с которыми они пытались справиться как самостоятельно, так и с помощью школьных или частных логопедов (этот параметр варьирует), но стойкого достаточного эффекта не получали.

Согласно данным проведенного нейропсихологического обследования (Ахутина и др., 2016), дети были разнородны по типам имеющихся дефицитов при дислексии. У части протестированных детей отмечались трудности с обработкой зрительной и зрительно-пространственной информации, у части детей наблюдалось снижение навыков фонематической обработки, слухоречевой памяти, восприятия ритмов, у некоторых из детей была отмечена быстрая утомляемость от деятельности, связанной с восприятием письменного текста. В ряде случаев можно было отметить комбинацию двух или более дефицитов с разной степенью выраженности. Качество письма у детей варьировало от близкого к норме (при переписывании текста с образца) до почти полностью отсутствующего письма. Также критериями включения были нормальное или скорректи-

Таблица 2. Оценка уровня сложности текста I и текста II

	Текст I	Текст II
Средняя частотность входящих полнозначных слов (вхождений на миллион)	2249.5	1102.0
Средняя длина входящих полнозначных слов в слогах	1.94	2.15
Средняя длина входящих полнозначных слов в буквах	4.90	5.88

рованное до нормального зрение, нормальный слух, нормальный невербальный интеллект (измеренные теми же инструментами, что и в группе типично развивающихся детей) и отсутствие истории неврологических нарушений.

Письменные формы информированного согласия были подписаны родителями или законными представителями детей; дети также устно согласились участвовать. Исследование было одобрено Комиссией по внутриуниверситетским опросам и этической оценке эмпирических исследовательских проектов Национального исследовательского университета «Высшая школа экономики» (Россия) (протокол № 47 от 08.05.2018).

Материалы и процедура

Скорость чтения и понимание прочитанного у детей были оценены с помощью СМИНЧ (Корнев, Ишимова, 2010). Для первичного тестирования авторами методики рекомендовано использовать первые половины текстов I и II (по три абзаца из каждого текста). «При повторных исследованиях можно предлагать ребенку чтение второй половины текстов I или II или использовать дополнительные тексты, приведенные в приложении 2» (Корнев, Ишимова, 2010, с. 11). Для каждой половины текста предусмотрены свои 10 вопросов на понимание прочитанного. Мы использовали первые половины текстов I и II («Как я ловил раков» и «Неблагодарная ель») и вопросы к ним.

Первый текст (3 абзаца, 97 слов) был проще; второй текст (3 абзаца, 127 слов) был более сложным, по утверждению авторов методики (Корнев, Ишимова, 2010, с. 13). Для количественной оценки уровня сложности текста I и текста II мы проанализировали слова, входящие в их состав, с точки зрения частотности (в количествах вхождений на миллион, *ipm*) и длины (в слогах и в буквах). Мы использовали информацию о частотности слов из Лингвистической базы данных для слов русского языка StimulStat Project (Alexeeva et al., 2017). При подсчете мы учитывали все полнозначные слова (по леммам). Слова, встречающиеся в тексте несколько раз, учитывались по количеству вхождений. Результаты анализа подтвердили, что текст I содержит более высокочастотные слова, а в тексте II использованы слова большей длины (см. таблицу 2).

Все дети до начала чтения были предупреждены, что после прочтения каждого текста им нужно будет отвечать на вопросы к тексту. Среди принявших уча-

стие в нашем исследовании типично развивающихся детей не было ни одного первоклассника, который не смог бы прочитать предложенные тексты. Учитывая, что СМНЧ — единственная стандартизированная методика для оценки навыков чтения на русском языке и что она широко используется логопедами, мы не можем строго утверждать, что все дети видели данные тексты впервые. Однако никто из детей, участвующих в тестировании, не сообщил после чтения о том, что этот текст он уже читал. Данные тексты не использовались для проверки навыков чтения при отборе детей с дислексией для нашего исследования.

В соответствии с оригинальной процедурой дети должны были читать тексты вслух и сразу после прочтения каждого из текстов ответить на десять вопросов к содержанию прочитанного. Подавляющее большинство детей не успевали прочитать больше трех абзацев за одну минуту. Если ребенок читал медленно, мы не останавливали его, пока он не закончит читать третий абзац — то есть фрагмент, достаточный для ответов на все вопросы. Если ребенок читал быстро, мы не прерывали его в момент окончания чтения третьего абзаца, а ждали, пока пройдет одна минута с того момента, когда он закончил читать название текста, и только потом останавливали, чтобы иметь возможность посчитать количество правильно прочитанных за одну минуту слов.

Согласно оригинальным рекомендациям, тестирующий должен слушать, как ребенок читает, и контролировать параметры чтения, используя форму с текстом и количеством слов, напечатанных в конце каждой строки. Однако некоторых детей это отвлекало, и мы фиксировали аудиозаписи, пока дети читали. Эти записи были в дальнейшем проанализированы в автономном режиме.

Скорость чтения измерялась как количество слов, прочитанных верно за первую минуту (общее количество прочитанных слов минус количество ошибочно прочитанных слов (Корнев, Ишимова, 2010, с. 12)). Понимание прочитанного оценивалось как количество правильных ответов на вопросы. Когда участник не мог прочитать второй текст совсем (10 детей с дислексией), мы считали скорость чтения равной 0, уровень понимания прочитанного равным 0, а количество неправильно прочитанных слов было помечено как NA (нет данных).

Для диагностических целей авторы (Корнев, 2003; Корнев, Ишимова, 2010) предложили преобразовать исходные данные по скорости чтения в коэффициент техники чтения (КТЧ) по следующей формуле:

$$\text{КТЧ} = 100 + ((M - m) / m) * 50,$$

где M — количество слов, правильно прочитанных ребенком за первую минуту, m — среднее количество слов, правильно прочитываемых за одну минуту детьми того же класса. Соответствие между скоростью чтения и КТЧ для конкретного класса было представлено в таблице КТЧ (Корнев, Ишимова, 2010, с. 46–47). В качестве аргумента в пользу преобразования прямых данных о скорости чтения детей в коэффициенты КТЧ авторы СМНЧ отмечали возможность приведения данных к стандартной шкале, совпадающей по

размерности со шкалой IQ при использовании адаптированного варианта методики Д. Векслера АВМ-WISC (Панасюк, 1973). Это должно было обеспечить удобство сравнения результатов детей, протестированных с помощью СМНЧ и АВМ-WISC. Однако в настоящее время тест Д. Векслера для русскоязычных детей имеет ограничения для применения в качестве метода психометрии (Базыльчик, 2016). Сравнение результатов теста на чтение с результатами тестов на интеллект чрезвычайно важно для выявления детей с дислексией, но необходимо использовать методики с актуальными нормативными данными. С учетом того, что нормативные данные могут устаревать, а рестандартизация тестов может проходить с разной скоростью, на наш взгляд, целесообразно приводить нормативные данные по чтению с указанием выборочных средних и стандартных отклонений, чтобы у исследователей и практиков была возможность выбирать наилучший из доступных в конкретный момент времени тестов на интеллект. Мы в нашем исследовании использовали тест Дж. Равена, поскольку к нему есть более актуальные нормы для русскоязычных детей, которые приведены в официальном руководстве (Равен и др., 2012) и которые собраны с участием русскоговорящих младших школьников (Давыдов, Чмыхова, 2016).

Авторами СМНЧ было рекомендовано рассматривать показатели в пределах от одного стандартного отклонения до полутора стандартных отклонений ниже среднего уровня скорости чтения для определенного класса как непатологические формы задержки чтения, которые предположительно могут быть преодолены педагогическими средствами. В свою очередь, скорость чтения более чем на полтора стандартных отклонения ниже средней должна рассматриваться как принадлежащая к спектру дислексии. Значения от одного до полутора стандартных отклонений будут сигнализировать о риске дислексии. Мы использовали вышеупомянутые рекомендации для оценки валидности СМНЧ (Корнев, Ишимова, 2010).

Поскольку для русского языка отсутствуют другие стандартизированные тесты на чтение, было невозможно оценить тот вид *критериальной валидности*, которая рассчитывается как коэффициент корреляции между баллами по изучаемому тесту и внешней мерой — результатами аналогичных тестов, признанных «золотым стандартом» (Ivanova, Hallowell, 2013). Поэтому мы использовали такой вид *критериальной валидности*, как *валидность по одновременности* (*конкурентная валидность*), которая рассчитывается посредством сравнения с другим критерием — наличием или отсутствием клинического диагноза дислексии у ребенка, принимающего участие в исследовании. Мы проанализировали данные 90 типично развивающихся детей и 50 детей с клиническим диагнозом «дислексия». Мы оценили такие психометрические свойства СМНЧ, как чувствительность (процент детей с клинически диагностированной дислексией, которые показали результаты ниже контрольного уровня и которым был бы поставлен диагноз «дислексия» при использовании СМНЧ), а также специфичность (процент участников без дислексии, которые показали результаты в пределах нормы и которые были бы диа-

гностированы как типично развивающиеся — на основе СМИНЧ; Ivanova & Hallowell, 2013). Кроме того, мы изучили соответствие между клиническим диагнозом и показателями этих детей по СМИНЧ, используя наши новые данные о типично развивающихся детях.

Результаты

Скорость чтения

В таблице 3 представлена средняя скорость чтения типично развивающихся детей от 1-го до 4-го класса. Для сравнения мы представляем данные, опубликованные автором СМИНЧ в 1997 году (данные, переизданные в 2003 году, были такими же, как и в 1997 году). Руководство, опубликованное в 2010 году, не содержало информации о средних и стандартных отклонениях для скорости чтения, но представляло результаты преобразования по формуле КТЧ исходных норм (см. раздел «Метод»). Поэтому мы взяли формулу КТЧ Корнева (2003) и на основе таблицы КТЧ из публикации Корнева и Ишимовой (2010, с. 46–47) вычислили средние значения скорости чтения для учеников 2–4-го классов, которые должны были быть использованы для получения таких КТЧ.

Таблица 3. Средняя скорость чтения

Класс	Средняя скорость чтения (Корнев, 1997, 2003)		Рассчитанная средняя скорость чтения из КТЧ (Корнев, Ишимова, 2010)		Средняя скорость чтения, по нашим новым данным	
	Текст I	Текст II	Текст I	Текст II	Текст I	Текст II
1	NA	NA	NA	NA	43.3	31.2
2	52.2	41.7	62.5	52.0	70.3	55.1
3	101.4	89.5	84.7	72.5	90.4	75
4	96.1	98.5	106.4	92.6	94.5	95.4

Примечание. NA — данные отсутствовали.

Авторы методических рекомендаций к СМИНЧ (Корнев, Ишимова, 2010) предложили рассматривать как непатологические формы задержки чтения результаты по скорости чтения на одно стандартное отклонение ниже среднего уровня для соответствующего класса, а как принадлежащие к спектру дислексии рассматривать результаты ниже чем на полтора стандартных отклонения по сравнению со средним значением для соответствующего класса (см. раздел «Метод»). Мы использовали те же уровни отсечки для наших новых данных. Было невозможно надежно реконструировать стандартные отклонения для выборки Корнева и Ишимовой (2010 г.), поскольку авторы не предоставили

исходные индивидуальные данные в открытых источниках. Поэтому мы представляем сравнение средней скорости чтения и стандартных отклонений между нашими нормативными данными и данными, опубликованными Корневым (1997, те же данные в 2003) в таблице 4. В таблице 5 мы представляем контрольные уровни для оценки скорости чтения на основе наших нормативных данных: значения для нормы, группы риска и дислексии.

Понимание прочитанного

Вторым важным критерием оценки навыков чтения является уровень понимания прочитанного. С этой целью автор СМИНЧ предложил 10 вопросов о содержании к каждому из текстов. Было рекомендова-

Таблица 4. Сравнение данных о средней скорости чтения и стандартных отклонениях в скорости чтения, представленных Корневым (1997, 2003) и полученных на основе наших данных

Класс	Данные Корнева (1997, с. 183; 2003, с. 209)		Наши новые данные	
	Текст I	Текст II	Текст I	Текст II
	Средняя скорость чтения	SD	Средняя скорость чтения	SD
1	NA	NA	43.3	13.5
2	52.2	11.2	41.7	14.9
3	101.4	29.3	89.5	23.7
4	96.1	27.7	98.5	27.8

Примечание. SD — стандартное отклонение.

Таблица 5. Контрольные уровни для оценки скорости чтения (основанные на наших данных)

Класс	Типично развивающиеся дети, N=90					
	Текст I			Текст II		
	Среднее (типичное чтение)	Среднее -1 SD (группа риска)	Среднее -1.5 SD (дислексия)	Среднее (типичное чтение)	Среднее -1 SD (группа риска)	Среднее -1.5 SD (дислексия)
1	43.3	29.8	23.2	31.8	21.7	16.6
2	70.3	53.7	45.4	55.1	40.0	32.4
3	90.4	74.6	66.6	75.0	57.9	49.2
4	94.5	86.7	82.8	95.4	80.2	72.5

Примечание. SD — стандартное отклонение.

Таблица 6. Контрольные уровни для оценки понимания прочитанного (основанные на наших данных)

Типично развивающиеся дети, N=90						
Класс	Текст I			Текст II		
	Среднее количество верных ответов на вопросы (норма)	Среднее -1 SD (группа риска)	Среднее -1.5 SD (дислексия)	Среднее количество верных ответов на вопросы (норма)	Среднее -1 SD (группа риска)	Среднее -1.5 SD (дислексия)
1	6.17	3.9	2.7	3.6	1.3	0.2
2	7.0	5.5	4.8	4.3	2.2	1.1
3	7.8	6.3	5.6	4.9	2.7	1.6
4	7.3	5.9	5.3	5.3	3.4	2.5

Примечание. SD — стандартное отклонение.

но оценивать результаты по четырем уровням: А — 10 правильных ответов на 10 вопросов, В — 7–9 правильных ответов, С — 4–6 правильных ответов и D — 3 или менее правильных ответов. Уровни А и В рекомендуется рассматривать как «функционально достаточные», то есть позволяющие использовать полную информацию, содержащуюся в тексте. Тем не менее средние данные о количестве правильных ответов на вопросы к текстам в группе типично развивающихся детей никогда не публиковались Корневым и Ишимовой, поэтому мы не можем сравнивать результаты понимания прочитанного на основе наших данных с их результатами.

В таблице 6 мы представляем контрольные уровни для оценки понимания прочитанного на основе показателей 90 типично развивающихся детей протестированной нами группы: значения для группы нормы, группы риска и дислексии. Эти данные показывают, что даже в четвертом классе типично развивающиеся дети с нормальным уровнем развития невербального интеллекта дают в среднем около 5 правильных ответов на 10 вопросов для текста II (младшие дети дают еще меньше правильных ответов). Таким образом, первоначальный критерий, рекомендованный в методическом пособии к СМИНЧ (Корнев, Ишимова, 2010), то есть 7–10 правильных ответов, вряд ли может служить основой для диагностики нарушений понимания прочитанного.

Валидность СМИНЧ

Как описано в разделе «Метод», мы исследовали валидность СМИНЧ (Корнев, Ишимова, 2010) для диагностики дислексии, рассчитав такие психометрические свойства теста, как чувствительность и специфичность, используя оригинальные нормы, приведенные в последнем опубликованном методическом пособии

к тесту (Корнев, Ишимова, 2010), и наши новые данные для группы типично развивающихся детей.

Специфичность

Чтобы оценить специфичность теста, мы проанализировали, какой процент типично развивающихся детей попал бы в группу нормы на основе СМИНЧ (Корнев, Ишимова, 2010). Поскольку нормативные данные по СМИНЧ опубликованы только для второго и последующих классов, мы не имели возможности оценить результаты 18 первоклассников из нашей выборки. Среди 72 детей, учащихся 2–4-го классов, из нашей группы типично развивающихся школьников не было ни одного ребенка, у которого КТЧ попадал бы в группу дислексии (то есть ниже уровня в 1.5 стандартных отклонения) согласно данным таблицы КТЧ, опубликованной в пособии Корнева и Ишимовой (2010). У 69 детей значения КТЧ были в пределах 1 стандартного отклонения (типичное чтение), а у трех детей (один ребенок четвертого класса и два второклассника) — между 1 и 1.5 стандартного отклонения (группа риска). Согласно этим результатам, ни у одного ребенка не была бы ошибочно диагностирована дислексия, а специфичность СМИНЧ может быть оценена как 100 %.

Стоит отметить, что в отсутствие скрининга невербального интеллекта, зрения и слуха неспецифические нарушения чтения могут быть ошибочно отнесены к дислексии, в то время как проблема может быть вызвана другим сопутствующим нарушением (Snowling, Nash, Gooch, Hayiou-Thomas, & Hulme, 2019). В нашей первоначальной когорте было 15 детей (позже исключенных из анализа) без диагноза дислексии, но с зарегистрированными нарушениями слуха или с результатами ниже нормы в тесте на уровень невербального интеллекта (Равен и др., 2012). В случае, если бы мы не проверяли их по этим критериям исключения, четверо детей (двое детей по результатам чтения обоих текстов и двое других детей по результатам чтения второго текста) могли бы быть ошибочно отнесены к группе дислексии на основе их значений КТЧ, что снизило бы показатели специфичности теста.

Мы проверили, насколько изменилась бы специфичность теста при использовании диагностических критериев, основанных на наших новых данных. Из 90 детей, учащихся 1–4-го классов, из нашей группы типично развивающихся школьников показатели по скорости чтения у пяти детей (два второклассника, один ребенок из третьего класса и два ребенка из четвертого класса) оказались ниже уровня в 1.5 стандартных отклонения от среднего. То есть эти дети попали бы в группу дислексии, что снизило бы специфичность теста со 100 % до 94.4 %.

Чувствительность

Чтобы оценить чувствительность теста, мы проанализировали, какой процент детей с клинически диагностированной дислексией попал бы в группу дислексии на основе СМИНЧ (Корнев, Ишимова, 2010). Девять первоклассников не могли быть оценены, так как нормативные данные по СМИНЧ были опубликованы только начиная со второго класса. В таблице 7 представлено процентное распределение

оставшихся детей 2–4-го классов ($n = 41$) по группам с типичным чтением, риском развития дислексии и с дислексией на основе КТЧ для текста I и текста II.

Таблица 7. Распределение детей с клинически диагностированной дислексией по группам на основе СМИНЧ и норм Корнева и Ишимовой (2010) ($N = 41$, 2–4-й классы)

$N = 41$	Скорость чтения в пределах нормы, %	Группа риска, %	Дислексия, %
Текст I	51.2	12.2	36.6
Текст II	43.9	17.1	39.0

Таким образом, согласно результатам чтения текста I, 15 детей из 41 (2–4-й классы) с клинически диагностированной дислексией были бы классифицированы как имеющие дислексию при использовании нормативных данных, опубликованных в 2010 году. Это дает чувствительность теста 36.6%. Согласно результатам чтения текста II, 16 детей были бы классифицированы как имеющие дислексию, что соответствует чувствительности теста, равной 39.0%.

Мы проверили, как ситуация изменилась бы, если бы мы использовали уровни отсечки в соответствии с нашими новыми данными, полученными в 2018 году (они приведены в таблице 5). В таблице 8 представлено процентное распределение всех протестированных детей с клинически диагностированной дислексией по группам на основе результатов в тесте на чтение СМИНЧ и согласно новым данным для текста I и текста II.

Таблица 8. Распределение детей с клинически диагностированной дислексией по группам на основе СМИНЧ и наших новых данных для скорости чтения ($N = 50$, 1–4-й классы)

$N = 50$	Скорость чтения в пределах нормы, %	Группа риска, %	Дислексия, %
Текст I	28.0	0.0	72.0
Текст II	26.0	14.0	60.0

С новыми уровнями отсечки по скорости чтения текста I 34 из 50 детей с клинически диагностированной дислексией были бы классифицированы как имеющие диагноз «дислексия» при использовании наших новых данных. Это дает чувствительность теста 72.0%. Согласно скорости чтения текста II, 30 из 50 детей с диагнозом дислексия были бы классифицированы как имеющие диагноз «дислексия», что соответствует чувствительности теста, равной 60.0%.

Мы также проверили, изменилось ли бы значение чувствительности СМИНЧ, если, помимо скорости чтения, оценивалось бы понимание прочитанного. Мы использовали те же самые диагностические уровни

отсечки, которые были предложены в оригинальном руководстве для скорости чтения: одно стандартное отклонение от среднего значения для результатов, оцениваемых как группа риска, и полтора стандартных отклонения для результатов, относимых к дислексии. Мы использовали только наши новые данные (см. таблицу 6), поскольку авторы СМИНЧ не публиковали нормативные данные для понимания прочитанного. В протестированной нами когорте все участники с нарушением понимания прочитанного имели также нарушение скорости чтения, при этом было 13 детей с пониманием прочитанного в пределах нормы, но с нарушениями скорости чтения. В таблице 9 представлено процентное распределение всех протестированных детей с клинически диагностированной дислексией по группам на основе показателей в тесте СМИНЧ на уровень понимания прочитанного и в соответствии с нашими данными для текста I и текста II.

Таблица 9. Распределение детей с клинически диагностированной дислексией по группам на основе СМИНЧ и наших новых нормативных данных для понимания прочитанного ($N = 50$, 1–4-й классы)

$N = 50$	Понимание прочитанного в пределах нормы, %	Группа риска, %	Дислексия, %
Текст I	78.0	2.0	20.0
Текст II	62.0	0.0	38.0

Среди 50 детей с клинически диагностированной дислексией 10 показали результаты ниже отсечки на уровне 1.5 стандартных отклонения, отвечая на вопросы к тексту I. Эти дети были бы диагностированы как имеющие дислексию, если бы в качестве диагностического критерия использовались только показатели понимания прочитанного. Это дает чувствительность теста 20%. 19 детей показали результаты ниже этого уровня отсечки, отвечая на вопросы к тексту II, что соответствует чувствительности теста, равной 38%.

В поисках объяснения, почему у 14 детей (28% из всех протестированных детей с дислексией), которые находились в пределах нормативного диапазона по скорости чтения и уровню понимания прочитанного на основе СМИНЧ, была клинически диагностирована дислексия, мы выполнили дополнительный анализ. Для этого мы переслушали аудиозаписи того, как дети читали, и подсчитали взвешенные баллы за ошибки, используя систему штрафных баллов. Корнев и Ишимова (2010, с. 9) утверждали, что «ошибки, которые допускают дети с дислексией, качественно не отличаются от ошибок, которые в норме наблюдаются у начинающих овладевать чтением». В нашей выборке дети допускали ошибки разного уровня: от незначительных, таких как повторение одного звука или ослабление звука в слабой позиции, до серьезных, таких как произнесение другого слова вместо представленного или потеря строки и дезориентация. В таблице 10 представлены типы ошибок, которые

совершали как дети из группы нормы, так и дети из группы с дислексией, а также система штрафных баллов за эти ошибки.

Таблица 10. Система штрафных баллов за ошибки разного типа

Типы ошибок	Штрафные баллы	
	Без самокоррекции	С самокоррекцией
Неверное ударение в слове	1	0.5
Пропуск, замена или добавление одного звука	1	0.5
Перестановка местами 2–3 звуков	2	1
Пропуск, добавление или изменение 2–3 звуков	2	1
Пропуск одного слова	2	1
Повтор одного слова	2	1
Добавление одного служебного слова	2	1
Добавление одного полнзначного слова	3	1.5
Замена одного слова (замена 4 или более звуков в одном слове)	3	1.5
Потеря строки и дезориентация	3	1.5

Для каждого ребенка мы рассчитали итоговый взвешенный балл за ошибки, суммируя штрафные баллы за каждую допущенную ошибку. Результаты 90 типично развивающихся детей представлены в таблице 11. Затем мы подсчитали штрафные баллы за ошибки для каждого из тех 14 детей с клинически диагностированной дислексией, у которых не был выявлен дефицит чтения по основным показателям СМИНЧ. Мы обнаружили, что для всех этих детей штрафные баллы по крайней мере для одного из текстов были более чем на 1.5 стандартных отклонения выше, чем средние значения для соответствующего класса. Другими словами, эти дети читают бегло, но с большим количеством ошибок большего веса. Таким образом, они имеют нарушение качества чтения, а не скорости, что остается незамеченным при использовании оригинальных критериев оценки СМИНЧ (Корнев, Ишимова, 2010).

Мы проверили, изменилось ли бы значение чувствительности СМИНЧ, если, помимо скорости чтения, оценивалось бы качество чтения с учетом нового критерия — взвешенных баллов за ошибки. Мы использовали те же самые диагностические уровни отсечки, которые были предложены в оригинальном руководстве для скорости чтения: одно стандартное отклонение от среднего значения для результатов, оце-

ниваемых как группа риска, и полтора стандартных отклонения для результатов, относимых к дислексии. Мы использовали наши данные по 50 протестированным детям с дислексией для текста I, но по 40 протестированным детям с дислексией для текста II, поскольку 10 детей с дислексией (20% участников с дислексией) текст II не смогли прочитать совсем, и данные об ошибках во втором тексте для них отсутствуют. В таблице 12 представлено процентное распределение всех протестированных детей с клинически диагностированной дислексией по группам на основе взвешенной оценки ошибок в тесте СМИНЧ в соответствии с нашими данными для текстов I и II.

Таблица 11. Контрольные уровни для оценки взвешенных баллов за ошибки (на основе наших новых данных)

Типично развивающиеся дети, N=90						
Класс	Текст I			Текст II		
	Средний балл за ошибки	Среднее +1 SD	Среднее +1.5 SD	Средний балл за ошибки	Среднее +1 SD	Среднее +1.5 SD
1	2.75	4.61	5.54	2.72	4.70	5.69
2	3.77	7.15	8.84	4.44	7.69	9.32
3	3.40	5.92	7.18	5.54	9.56	11.57
4	2.95	5.90	7.38	5.76	9.52	11.4

Примечание. SD — стандартное отклонение.

Таблица 12. Распределение детей с клинически диагностированной дислексией по группам на основе взвешенной оценки ошибок в тесте СМИНЧ (в соответствии с нашими данными, 1–4-й классы)

	Взвешенные баллы за ошибки в пределах нормы, %	Группа риска, %	Дислексия, %
Текст I (N=50)	30.0	6.0	64.0
Текст II (N=40)	7.5	15.0	77.5

Среди 50 детей с клинически диагностированной дислексией 32 показали результаты ниже отсечки на уровне 1.5 стандартных отклонения при чтении текста I. Эти дети были бы диагностированы как имеющие дислексию, если бы в качестве диагностического критерия использовались только взвешенные баллы за ошибки при чтении. Это дает чувствительность теста 64%. Среди 40 детей с дислексией, сумевших прочитать второй текст, 31 ребенок показал результаты ниже уровня отсечки, что соответствует чувствительности теста, равной 77.5%.

Заключение

Стандартизированные тесты для диагностических целей должны использоваться с актуальными нормативными данными, поскольку нормативные показатели могут изменяться со временем (Raven, 2000). Тот факт, что нормативные данные необходимо обновлять, касается не только данной конкретной методологии, но и является характерным для большинства стандартизированных тестов. В методическом пособии, опубликованном в 2010 году (Корнев, Ишимова, 2010), авторы также отмечали целесообразность обновления нормативных данных. Важно отметить, что для получения надлежащих уровней отсечки для определения нормативных показателей предпочтительнее тестировать не случайно выбранных детей, а тех, которые соответствуют таким критериям включения, как нормальные интеллектуальные способности, зрение и слух, а также отсутствие диагностированных неврологических расстройств (Ivanova, Hallowell, 2013).

Данные, которые мы собрали для СМИНЧ в 2018 году, обеспечивают обновленные уровни для оценки показателей чтения и проконтролированы с учетом упомянутых критериев включения в тестируемую нормативную выборку. Хотя количество участвовавших в нашем исследовании типично развивающихся детей ($N=90$) недостаточно для полной рестандартизации теста, тот факт, что нормы, рассчитанные на выборке в 90 человек, оказались в диагностическом плане продуктивнее норм из авторского руководства (Корнев, Ишимова, 2010) для выявления клинического феномена, говорит о том, что нормы необходимо пересматривать.

Что касается диагностической силы СМИНЧ, специфичность теста является превосходной, это означает, что любой нормативный набор данных СМИНЧ позволяет идентифицировать типичные показатели чтения и отсутствие дислексии в новой когорте тестируемых с высокой точностью: от 100% (при использовании нормативных данных, полученных в 2007–2008 годах) до 94.4% (при использовании контрольных уровней, рассчитанных на данных 2018 года). Однако чувствительность СМИНЧ существенно зависит от нормативных данных, используемых для расчета уровней отсечки. При использовании нормативных данных, полученных в 2007–2008 годах, чувствительность теста находилась между 36.6% и 39.0% (в зависимости от текста), а при использовании наших новых данных по скорости чтения, полученных в 2018 году, чувствительность теста достигала 60.0–72.0%. Чувствительность теста даже 72.0% означает, что 28.0% детей с нарушениями чтения могут быть ошибочно не диагностированы как таковые. Как было показано во многих исследованиях, следствием непризнания дислексии является отсутствие адекватного вмешательства, приводящее к тревоге и депрессивному поведению (Mugnaini et al., 2009; Willcutt, Pennington, 2000; Nelson, Harwood, 2011; Törö et al., 2018), суицидальным мыслям, неудачам в школе и отсеvu из школы (Barbiero et al., 2019; Daniel et al., 2006; Wilson et al., 2009).

Относительно низкая чувствительность СМИНЧ является результатом использования только одного (хотя и самого важного) диагностического критерия —

скорости декодирования. Однако инструмент, разработанный Корневым и Ишимовой, фактически позволяет использовать два дополнительных диагностических критерия: уровень понимания прочитанного и взвешенную оценку ошибок, допущенных при чтении. Мы предоставляем контрольные уровни для оценки понимания прочитанного и взвешенной оценки ошибок, основанные на результатах протестированной нами когорты типично развивающихся детей, для дальнейшего улучшения клинической валидности СМИНЧ.

Наше исследование показало, что введение в качестве дополнительного диагностического критерия взвешенной оценки ошибок может улучшить диагностические возможности СМИНЧ. Подсчет штрафных баллов за ошибки оказался возможен благодаря изменению процедуры тестирования (аудиозаписи того, как читают дети). Учитывая, что это нововведение и решает проблему отвлечения детей, и позволяет более точно оценить параметры чтения, можно рекомендовать включить в процедуру диагностики аудиозапись чтения.

Что касается оценки понимания прочитанного (которая в СМИНЧ осуществляется с помощью вопросов о содержании, на которые дети отвечают после прочтения текстов), то результаты этих тестов довольно вариативны, но в целом достаточно низки даже у нормы. Это приводит к тому, что как нарушения диагностируются только крайне низкие уровни понимания прочитанного (см. таблицу 6, а также комментарии на с. 16 в авторском руководстве Корнева и Ишимовой, 2010). Актуальные данные показывают, что применение критерия «понимание прочитанного» не позволяет повысить общую чувствительность теста. Это говорит либо о том, что данный тест может быть полезен только в специфических случаях, когда у ребенка при нормальной скорости и точности чтения отмечается очень низкий уровень понимания прочитанного (в нашей выборке таких детей не встретилось), либо о том, что диагностическая ценность данного задания сомнительна. Однако поскольку понимание прочитанного является чрезвычайно важным для школьного обучения в целом, в будущих исследованиях, возможно, стоит рассмотреть варианты более эффективных тестов на понимание прочитанного.

Важно, что три вышеупомянутых дефицита (медленное чтение, недостаточное понимание прочитанного и многочисленные серьезные ошибки при чтении), измеряемые с помощью метрик СМИНЧ (скорость чтения, уровень понимания и взвешенная оценка ошибок соответственно), могут наблюдаться у детей с дислексией как по отдельности, так и в сочетании. Вопрос о том, следует ли для диагностики дислексии рассматривать один или несколько критериев, является предметом соглашения и будущих исследований. Наше исследование показало, что дети с только одним типом дефицита — с высокой скоростью чтения и нормальным пониманием прочитанного, но с большим количеством ошибок — были всё же клинически диагностированы как имеющие дислексию.

Мы предлагаем использовать не один критерий, а два или лучше три критерия для диагностических целей: скорость чтения, понимание прочитанного и взвешенная оценка ошибок, поскольку ни один из критериев

отдельно не обеспечивает близкую к 100% чувствительность теста (см. раздел «Результаты»). Контрольные уровни для их оценки представлены в таблице 5, таблице 6 и таблице 11 соответственно. Если показатели ребенка на полтора стандартных отклонения ниже, чем среднее значение для того же класса, необходимо диагностировать дислексию. Кроме того, скрининг невербальных интеллектуальных способностей, слуха и зрения должен обязательно сопровождать тестирование навыков чтения, поскольку эти параметры связаны с чтением — и при оценке навыков чтения их необходимо учитывать. Стоит отметить, что западными исследователями при диагностике нарушений чтения используются также тесты на чтение слов (аналогичный, например, Brus, Voeten, 1973) и на чтение псевдослов (аналогичный, например, van den Bos et al., 1994). Разработка подобных стандартизированных инструментов для русского языка и проверка их эффективности при диагностике дислексии — одна из задач будущих исследований.

В заключение мы выражаем глубокую признательность А. Н. Корневу и О. А. Ишимовой за создание единственной доступной стандартизированной методики исследования навыков чтения у русскоязычных детей. Мы надеемся, что наши новые данные, клиническая экспертиза и предлагаемые поправки к подсчету результатов (касающиеся более точной оценки понимания прочитанного и введения взвешенной оценки ошибок) дополнят усилия авторов методики и будут способствовать еще более широкому использованию СМИНЧ.

Литература

- Ахутина Т. В., Корнеев А. А., Матвеева Е. Ю., Романова А. А., Агрис А. Р., Полонская Н. Н., Пылаева Н. М., Воронина М. Н., Максименко М. Ю., Яблокова Л. В., Меликян З. А., Кузева О. В. Методы нейропсихологического обследования детей 6–9 лет / Под ред. Т. В. Ахутиной. М.: В. Секачев, 2016.
- Базильчик С. В. Пригодность русифицированных версий детского теста Векслера (WISC) для диагностики умственной отсталости // Психиатрия, психотерапия и клиническая психология. 2016. № 2. С. 12–20. URL: <https://elibrary.ru/item.asp?id=26140287>.
- Давыдов Д. Г., Чмыхова Е. В. Применение теста Стандартные прогрессивные матрицы Равена в режиме ограничения времени // Вопросы психологии. 2016. № 4. С. 129–139.
- Корнев А. Н. Нарушения чтения и письма у детей. Учебно-методическое пособие. СПб.: МиМ, 1997.
- Корнев А. Н. Нарушения чтения и письма у детей. Учебно-методическое пособие. СПб.: Речь, 2003.
- Корнев А. Н., Ишимова О. А. Методика диагностики дислексии у детей. Методическое пособие. СПб.: Издательство Политехнического университета, 2010.
- Панасюк А. Ю. Адаптированный вариант методики Д. Векслера WISC. М.: Медицина, 1973.
- Равен Д. Цветные прогрессивные матрицы серии А, Аb, В. М.: Когито-Центр, 2004.
- Равен Д., Равен Д. К., Корт Д. Руководство к Прогрессивным Матрицам Равена и Словарным Шкалам. Раздел 3. Стандартные Прогрессивные Матрицы (включая Параллельные и Плюс версии). М.: Когито-Центр, 2012.
- Семенович А. В. Нейропсихологическая диагностика и коррекция в детском возрасте. М.: Академия, 2002.
- Alexeeva S., Slioussar N., Chernova D. StimulStat: A lexical database for Russian // Behavior Research Methods. 2018. Vol. 50. No. 6. P. 2305–2315. doi:10.3758/s13428-017-0994-3
- Barbiero C., Montico M., Lonciari I., Monasta L., Penge R., Vio C., ..., Ronfani L. on behalf of the EpiDIt (Epidemiology of Dyslexia in Italy) working group. The lost children: The underdiagnosis of dyslexia in Italy. A cross-sectional national study // PLoS One. 2019. Vol. 14. No. 1. P. e0210448. doi:10.1371/journal.pone.0210448
- Bates D., Mächler M., Bolker B., Walker S. Fitting linear mixed-effects models using lme4 // Journal of Statistical Software. 2015. Vol. 67. No. 1. P. 1–48. doi:10.18637/jss.v067.i01
- van den Bos K. P., Spelberg H. C. L., Scheepstra A. J. M., DeVries J. R. De Klepel. VormAenB. Eentestvoordeeleesvaardigheidvanpseudowoorden. Verantwoording, handleiding, diagnostiekenbehandeling [The Clapper. Form A and B. A test for reading the pseudowords. Accountability, manual, diagnostics treatment]. Nijmegen, the Netherlands: Berkhout, 1994.
- Brus B. T., Voeten M. J. Eén-minuut-test vorm A en B. Verantwoording en handleiding [One-minute word reading test version A and B. Justification and manual]. Nijmegen, the Netherlands: Berkhout, 1973.
- Daniel S. S., Walsh A. K., Goldston D. B., Arnold E. M., Reboussin B. A., Wood F. B. Suicidality, school dropout, and reading problems among adolescents // Journal of Learning Disabilities. 2006. Vol. 39. No. 6. P. 507–514. doi:10.1177/00222194060390060301
- Ivanova M. V., Hallowell B. A tutorial on aphasia test development in any language: Key substantive and psychometric considerations // Aphasiology. 2013. Vol. 27. No. 8. P. 891–920. doi:10.1080/02687038.2013.805728
- Moll K., Landerl K. SLRT-II: Lese- und Rechtschreibtest [Reading and Spelling Test]. Bern: Hans Huber, 2010.
- Mugnaini D., Lassi S., La Malfa G., Albertini G. Internalizing correlates of dyslexia // World Journal of Pediatrics. 2009. Vol. 5. No. 4. P. 255–264. doi:10.1007/s12519-009-0049-7
- Nelson J. M., Harwood H. Learning disabilities and anxiety: A meta-analysis // Journal of Learning Disabilities. 2011. Vol. 44. No. 1. P. 3–17. doi:10.1177/0022219409359939
- Raven J. The Raven's progressive matrices: Change and stability over culture and time // Cognitive Psychology. 2000. Vol. 41. No. 1. P. 1–48. doi:10.1006/cogp.1999.0735
- Snowling M. J., Nash H. M., Gooch D. C., Hayiou-Thomas M. E., Hulme C., Wellcome Language and Reading Project Team. Developmental outcomes for children at high risk of dyslexia and children with developmental language disorder // Child Development. 2019. Vol. 90. No. 5. P. e548–e564. doi:10.1111/cdev.13216
- Torgesen J. K., Wagner R. K., Rashotte C. A. Test of word reading efficiency (TOWRE). Austin, TX: PRO-ED, 1999.
- Törö K. T., Miklósi M., Horanyi E., Kovács G. P., Balázs J. Reading disability spectrum: Early and late recognition, sub-threshold, and full comorbidity // Journal of Learning Disabilities. 2018. Vol. 51. No. 2. P. 158–167. doi:10.1177/0022219417704169
- Wechsler D. Wechsler objective reading dimensions. London: The Psychological Corporation, 1990.
- Wilkinson G. S., Robertson G. J. Wide Range Achievement Test 4 professional manual. Lutz, FL: Psychological Assessment Resources, 2006.
- Willcutt E. G., Pennington B. F. Psychiatric Comorbidity in Children and Adolescents with Reading Disability // Journal of Child Psychology and Psychiatry. 2000. Vol. 41. No. 8. P. 1039–1048. doi:10.1017/S0021963099006368
- Wilson A. M., Deri Armstrong C., Furrie A., Walcot E. The mental health of Canadians with self-reported learning disabilities // Journal of Learning Disabilities. 2009. Vol. 42. No. 1. P. 24–40. doi:10.1177/0022219408326216
- Wimmer H., Mayringer H. SLS 2–9: Salzburger Lese-Screening für die Schulstufen 2–9 [Salzburg Reading and Spelling Test]. Bern: Hans Huber, 2014.
- Woodcock R. Woodcock Reading Mastery Test–Revised. Circle Pines, MN: American Guidance Services, 1999.