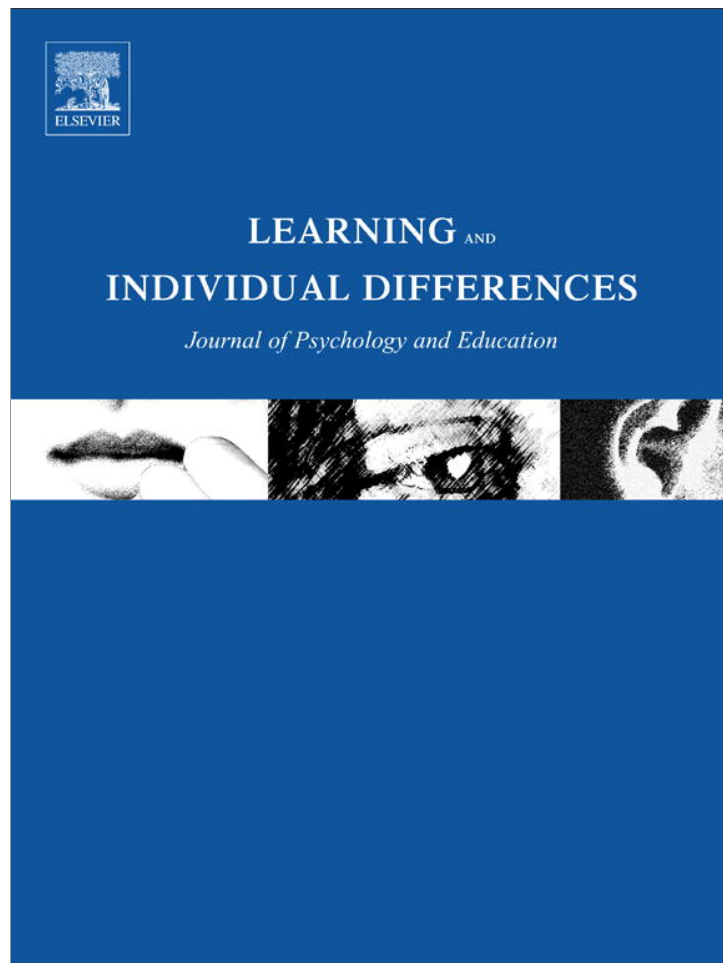


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

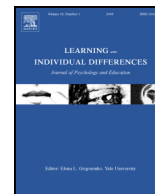
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Between PIRLS and PISA: The advancement of reading literacy in a 10–15-year-old cohort

Galina A. Zuckerman^{a,*}, Galina S. Kovaleva^b, Marina I. Kuznetsova^c^a Psychological Institute, Russian Academy of Education, 9V Mokhovaya ul., Moscow 125009, Russian Federation^b Center for Evaluating the Quality of Education, Russian Academy of Education, 8, Pogodinskaya ul., Moscow 119435, Russian Federation^c Institute of Content and Methods of Education, Russian Academy of Education, 8, Pogodinskaya ul., Moscow 119435, Russian Federation

ARTICLE INFO

Article history:

Received 17 November 2012

Received in revised form 30 April 2013

Accepted 2 May 2013

Available online xxxx

Keywords:

PISA

PIRLS

Assessing reading literacy

Push–Pull diagnostic method

Middle school

ABSTRACT

While reading literacy in Russian 10-year-old schoolchildren was exceptionally high in the PIRLS assessment, reading literacy of 15-year-old students as measured by PISA stayed low. To elucidate this controversy, we developed the “Push–Pull” diagnostic method combining PIRLS and PISA approaches. This new tool compares reading literacy across at least five years of schooling. With Push–Pull, we assessed grades 4, 6 and 9 (3110 students) and demonstrated that two years of study in Russian middle school (grades 5 and 6) failed to promote students' ability to comprehend informational texts, bringing modest improvement over five years of schooling. By measuring and monitoring the dynamics of reading literacy, Push–Pull spots the dead ends of Russian educational approach to reading and comprehension of informational texts in 10–15-year-old readers.

© 2013 Published by Elsevier Inc.

1. Why we need new ways to evaluate reading literacy?

Over the recent decade, our understanding of the importance of reading for the development of individuals and society as a whole has grown significantly (OECD, 2010a). High levels of reading literacy as assessed by the Programme for International Student Assessment (PISA) were shown to predict with considerable certainty that a student would graduate from high school and successfully continue his or her education. High levels of reading literacy based on PISA results were found to predict future economic well-being of an individual more reliably than his/her Grade Point Average (*ibid.*). The average literacy level in a country was found to be a better indicator of future economic growth than any generalized measure of educational achievement (*ibid.*).

These and similarly important messages are derived from countries with developed economies and established democracies. For Russia, where the relationship between education and annual income is not that evident, concern about the problem of reading literacy in middle school graduates is growing due first and foremost to our belief in the value of the written word. The people who are currently deciding the fate of the Russian education system grew up in a bookish milieu, and most of them believe that writing is the cornerstone of both past and future of civilization. The authors of this article agree with Joseph Brodsky when he writes that language and literature

are “things that are more ancient and inevitable, more durable than any form of social organization” (Brodsky, 1987). Hence ensuring the proficiency of youth in written language skills is both a time-honored pedagogical necessity and a moral and social obligation. We mention this first to emphasize the cultural significance of the value of literacy as one of the common threads that has bound generations together for centuries.

The objective of this research was to answer the question that many have asked lately: why is there such a huge gap between the results reached by the Russian students in two international comparative studies that evaluate literacy – the Progress in International Reading Literacy Study (PIRLS) and PISA. PIRLS assesses literacy at the end of elementary school, at the transition from mastering the process of reading itself to using reading as one of the main ways to further extend and expand one's education (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009). PISA assesses literacy of 15-year-olds who are in transition from using reading as a way to learn to employing reading as a means to adapt to the real life. In fact, both surveys evaluate students' readiness for a new way of using texts to meet the challenges of forthcoming developmental and educational tasks (OECD, 2010a). In PIRLS, Russian fourth graders thrice (in 2001, 2006 and 2011) performed with results well above average international levels (Mullis, Martin, Foy, & Drucker, 2012; Mullis, Martin, Gonzales, & Kennedy, 2003; Mullis, Martin, Kennedy, & Foy, 2007). In 2006, Russia significantly improved its results and was at the top of the leading group of countries in PIRLS performance. Meanwhile, 15-year-old Russians who participated in the PISA study of literacy significantly lagged behind the average scores for OECD countries. This situation has been observed already at four time points, in 2000, 2003, 2006

* Corresponding author. Tel.: +7 8 495 4179262.

E-mail addresses: galina.zuckerman@gmail.com (G.A. Zuckerman), galina_kovaleva_rao@mail.ru (G.S. Kovaleva), bernin@mail.ru (M.I. Kuznetsova).

and 2009 (OECD, 2001, 2004, 2006, 2010b). Instead of comparing scores across countries, here we focus on the trends in the Russian samples. Between 2000 and 2009, Russian students failed to make gains considerably in the PISA reading literacy scores (OECD, 2010c).

On the world-wide scale, Russia manifests the largest gap between PIRLS and PISA reading scores. Therefore the Russian educators are deeply concerned about the lack of progress in reading literacy of middle school students. This problem is not specifically Russian. In many countries, such as Austria, Croatia, Israel, Italy and Spain, PIRLS results are above the international average, while PISA reading scores are below the international average. The problem of stagnation or even decay of reading literacy in middle school students has been widely discussed as crucial for future national benefit (Luyten, Peschar, & Coe, 2008; NCES 2010–458, 2010).

The authors of this paper have been trying to figure out why there is such a large gap in reading literacy between the elementary and middle schools in Russia as manifested by the PIRLS and PISA surveys. To explain this gap, we propose two complementary hypotheses.

Hypothesis 1. Despite the overall success in literacy in the elementary school as evidenced by their high PIRLS results, there are weak spots in literacy education at this stage that later chronically retard the acquisition of reading skills, resulting in poor PISA scores.

An in-depth analysis of the performance of the Russian fourth graders in PIRLS 2006 showed that, amid high overall results, the non-fiction or informational texts presented more serious difficulties for students than the fiction or storybook texts. Most problematic were the questions addressed to the basic reading skill of retrieving explicitly stated information and making straightforward inferences (Kuznetsova, 2009).

This phenomenon can be explained by the present-day practice of teaching children to read. By analyzing textbooks that are currently in use in the Russian elementary schools, it was convincingly demonstrated that the textbooks on science, the main source of informational texts, do not develop reading literacy. Instead, they encourage students to memorize and reproduce the author's text (Frumin, 2010; Pinskaya, 2009).

Difficulty in searching for the necessary information and making the simplest inferences based on given information would likely be aggravated in middle school. In other words, **Hypothesis 1** needs to be explored further.

Hypothesis 2. The reasons for Russia's poor performance in the PISA test lie outside the elementary school.

The success of the Russian students in PIRLS shows that they are well prepared to read for learning. It does not imply that fourth graders have already mastered the skill to use texts as a means of self-education. We presume to test whether the school provides sufficient educational environment for developing this skill in 10–15-year old students.

2. Material and methods: measuring reading literacy with a PIRLS-PISA hybrid

2.1. The Push–Pull diagnostic method

To understand what is happening to reading literacy levels in the interval between the PIRLS and PISA tests, it is essential to examine the trends in literacy in the 10–15-year-old students. To approximate these trends, we assessed reading literacy at three points of the educational trajectory: in grades 4, 6 and 9. The fourth-grade scores are linked to PIRLS, the ninth grade to PISA, and grade 6 was chosen as a random point between them.

To keep the assessment consistent, we applied the same measure at all three time points. Diagnostic materials were constructed along

the lines of PIRLS and PISA, which are similar, both conceptually and methodologically. To prove that PIRLS and PISA measure the same construct – reading literacy – we compared the perception of the goals and tasks of reading as defined by the authors of the PIRLS and PISA. In PIRLS framework, reading literacy is defined as “The ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment.” (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009, p. 11). In PISA framework reading literacy is defined as “...the capacity to understand, use and reflect on written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society” (OECD, 2010a, pp. 23–25). It is worth noting that when specifying the object of measurement, the authors of both tests refer to almost the same body of theories describing the whole process of reading in all its complexity – from the recognition of letters to building mental representations of the text content. Both surveys attempt to capture the *practical* and *functional* aspects of the skills acquired in school, including the ability to understand text and use it to solve problems. This implies that both surveys assessed comparable subjects.

We also compared the views of the PIRLS and PISA authors on what constitutes reading activity as embodied in the structure of reading skills and assessed by tests. In PISA, there are three scales to independently measure various reading skills: (1) the ability to retrieve information from the text; (2) the ability to integrate (to combine into a single picture) and interpret (clarify it to oneself and other people) the message from a particular text; and (3) the ability to reflect on a text and evaluate it from one's own perspective (OECD, 2010a). Every diagnostic question in the test is designed to assess a particular ability. To address younger readers, PIRLS combines similar scales, and adds one more scale: straightforward inference based on the information retrieved from the text (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009). For younger readers, the minimal processing of literal meaning from the text is a separate task; for more mature readers, this process is automatic (Kintsch, 1998).

Our analysis of the theoretical views on reading activity basic for PIRLS and PISA has shown that two tests are essentially comparable and compatible.

To construct a diagnostic tool for measuring and monitoring the dynamics of reading literacy, we used texts and questions from PIRLS and PISA materials that were in the public domain (Russian Academy of Education). When selecting reading materials, we ensured that the PIRLS texts were also interesting for the sixth and ninth graders, and that the PISA texts were comprehensible for the younger students. We also took into consideration that the Russian students have a harder time understanding discontinuous texts (Kasprzhak, 2006; Kovaleva, 2004, 2006). As a result, two discontinuous texts (“The River Nord Trail” and “Searching for Food”) were chosen from the released PIRLS materials, with verbal information alternating with pictures, graphs and tables, together with two PISA continuous texts (“Graffiti” and “Flu”).

When choosing PISA texts for our experiment, we made minor revisions and additions to adapt them to the experience and knowledge of 10-year-olds. For example, we added pictures to the text “Graffiti” to illustrate the concept for fourth graders. Footnotes were also added to both tests to define words possibly unfamiliar to younger students.

We dubbed the new diagnostic method “Push–Pull” (after Dr. Doolittle's Push–Me–Pull–Me) to emphasize its two inseparable features: (1) the common nature of the PIRLS and PISA approaches to assessing reading literacy and (2) the difference between the surveyed age groups. When preparing materials for the Push–Pull survey, the most difficult part was constructing extra questions for each text. Here we were prompted by two factors. First, to meet statistical requirements for making the measurements reliable, we needed a considerable number of questions per test as a whole and separately for each process

of comprehending. Second, we sought to balance the difficulty level for the entire set of questions through the tests. To achieve this goal, we added more difficult questions to the PIRLS tests that require the ability to work with contradictory or incompletely defined information, to consider the problem from several perspectives, and to describe other scenarios for the probable unfolding of the described events. On the other hand, we added simpler questions to the PISA tests requiring the retrieval of information explicitly stated in the text and making straightforward inferences. As a result, we added 20 new questions to the initial set of 31 from PIRLS and PISA. A guide for scoring children's answers was composed for these extra questions.

To describe the Push–Pull method, we will first classify reading skills commonly assessed for in PIRLS and PISA items. We applied three scoring scales, based on the PISA design instead of the four-scale design used in PIRLS. Following the authors of PIRLS, we combined two processes of comprehension: retrieving the explicitly stated information and making the straightforward inferences based on this information. In Table 1, we present data on the number of questions for each group of processes in the Push–Pull survey. These figures indicate the prospects for further upgrading this instrument. To create a true PIRLS–PISA hybrid, we had to increase the number of questions that encouraged readers to retrieve information from the text, especially from PISA texts.

Materials for the Push–Pull method were verified in a pilot project, and the necessary adjustments were made. For further testing, the chosen texts and questions were divided into four booklets, so that each booklet contained two texts.

Special questionnaires were given to students and teachers in order to collect contextual information for identifying readers' habits and preferences at different ages, and other factors affecting students' levels of reading literacy.

2.1.1. Sample characteristics

We accounted for the following factors in our experimental sample:

- To match the stratification of the sample with the pattern typical of Russian schools, we focused on such characteristics as the type of educational institution (elementary school, middle school), form of schooling (general education, vocational, boarding school, grammar school, school with in-depth study of specific subjects), and location (type of community).
- To amass the required performance targets, no less than 3000 students participated in the survey (250 students for each of the four booklets in each of the three grades under comparison) ($250 \times 4 \times 3$).

Numerous studies imply that the type of community is among the most important factors determining the nature of the sample. We therefore focused on this factor when constructing our sample. At present, 29% of all elementary school students in Russia are taught in rural areas and 71%, in towns, while 65% of schools are located in rural areas and 35%, in towns. In the Tambov region chosen for this study, 49% of students study in rural areas and 51%, in towns, while rural schools make up 69% of the total and urban schools account for 31%.

Based on the ratio of rural to urban schools, and considering that one of every three schools has no fewer than 1000 students (250 for each booklet), it was necessary to test 290 students in rural schools and 710 students from urban schools in each grade. We were thus able to choose 30 schools: 15 rural and 15 urban. In each of the sampled schools, the tests were run in grades 4, 6 and 9. A total of 3110 participating students included 907 fourth graders, 1069 sixth graders and 1134 ninth graders. The overall sample for each grade retained the ratio of urban to rural schools typical of Russian demographics.

The Tambov region was also chosen for our research because its results in the national examinations and in other monitoring studies correspond to the country average.

2.2. Procedure

When coming up with protocols for test administration in schools, it was important to mark out time allotted to complete the test. The factor of time is not instrumental in measuring reading literacy, so we decided to reserve two class periods (45 min each) in order to guarantee that even the slowest readers had enough time to complete the test. The instructions stated: "The main goal of this study is to uncover whether you are able to read various types of texts, how well can you answer questions that demonstrate your understanding of what you have read, and how you express your opinion of the material. Quality is important here, not speed. There is plenty of time for even the slowest reader to complete the test". Teachers were instructed not to keep students for the full two class sessions; students were allowed to leave as soon they completed all their work.

The study was conducted in May 2010. Logistics for the application of the survey were provided by the Tambov region government agency, The Center of Expertise in Educational Activities. This Center has wide experience in administering national examinations, evaluating the quality of education at the regional level and participating in the monitoring of studies at the federal and regional levels. The test was administered in accordance with the specially developed guide,

Table 1
Types of questions used in the Push–Pull method.

	Reader's ability to		
	Retrieve information and make straightforward inferences	Interpret and integrate information	Reflect on the content and form of the text and evaluate them
<i>Distribution of question types in initial surveys (% of total number)</i>			
PIRLS	50	30	20
PISA	25	50	25
<i>Number of questions in the Push–Pull method (in parentheses, % of the total number)</i>			
Questions to four texts	15 (29)	24 (47)	12 (24)
Questions to two PISA texts	6	8	6
Questions to two PIRLS texts	9	16	6
Questions taken from PIRLS and PISA	10	12	9
New questions, constructed for the Push–Pull method	5	12	3
<i>Maximum points for answers in the Push–Pull method</i>			
Questions to four texts	16 (22%)	37 (51%)	20 (27%)
Questions to two PISA texts	6	10	11
Questions to two PIRLS texts	10	27	9
Questions taken from PIRLS and PISA	11	18	14
New questions, constructed for the Push–Pull method	5	19	6

with two teachers in each classroom who were not teaching the particular class. These teachers put on record: (1) the time when the first student had completed the test and turned in his or her booklet; (2) the time when the last student finished the test and handed in his or her booklet and 3) the time it took for most of the class to complete the test. We noted that the average time it took to complete the test was unrelated to the student's age.

All booklets were submitted to the Center of Expertise. A group of experts, which included teachers of Russian language and literature with experience in scoring the national examination papers, was trained to score the tests. Before we began our experiment, we held a seminar to acquaint these experts with the specific research and the guide describing the procedure for scoring the constructed response questions. Scoring the constructed response test questions together with the specialists who developed the testing procedure facilitated the scoring decisions in difficult cases, e.g., when a student gave an answer unforeseen by the manual, or when the experts had a difficult time evaluating the borderline responses of the students. On the whole, the experts attested to the high quality of the scoring guide, and it minimized the cases of debatable answers.

Data processing was carried out by the Center for Evaluating the Quality of Education using the CONQUEST program (<http://www.rasch.org/rmt/rmt133o.htm>). The reliability of the test based on the Cronbach's alpha coefficient was 0.80 for fourth graders, 0.82 for sixth graders and 0.83 for ninth graders. The average item discrimination of the Push–Pull method was within the range of 0.37 to 0.49 (Tables 2 and 3). For some questions, the ratio did not reach the threshold level of 0.30. The number of such instances among the additional questions was half that of the original PIRLS and PISA questions (6 and 13%, respectively). Improvements to the additional questions with low discrimination are another area of follow-up modifications for the Push–Pull method.

3. Results and discussion

3.1. General results of the Pull–Push assessment

The principal research goal of the reported survey was to study the progress in the acquisition of reading literacy in the period spanning grades 4 to 9. Reading literacy is known to develop through cognitive and personal maturation and pedagogical efforts which are only partially derived from formal schooling (Vygotsky, 1992). Later studies discriminate the particular effect of schooling in the general picture of students' development (Cahan & Davis, 1987; Luyten, 2006).

Table 2
Successful answers to the test questions used in the Push–Pull method.

	Four texts	Two PIRLS texts	Two PISA texts	Two PIRLS texts		Two PISA texts	
				Searching for food	The River Nord trail	Graffiti	Flu
<i>The total number of questions</i>							
	51	31	20	16	15	10	10
<i>Answered correctly, % of the maximum score</i>							
Grade 4	54.3 ^a	61.1 ^a	48.4 ^a	54.2 ^a	66.3 ^a	47.8 ^a	49.1 ^a
Grade 6	55.2 ^a	63.1 ^a	48.5 ^a	56.4 ^a	67.8 ^a	49.7 ^a	47.2 ^a
Grade 9	65.9 ^b	70.0 ^b	62.4 ^b	64.1 ^b	74.5 ^b	64.2 ^b	60.6 ^b
<i>Standard deviations</i>							
Grade 4	20.2	18.1	20.1	17.4	16.8	20.8	19.4
Grade 6	21.8	19.1	21.6	18.0	18.4	22.7	20.3
Grade 9	21.0	17.4	22.3	16.6	16.7	23.7	20.7
<i>Average item discrimination for questions</i>							
Grade 4	0.41	0.42	0.40	0.42	0.41	0.43	0.37
Grade 6	0.43	0.44	0.42	0.41	0.46	0.44	0.40
Grade 9	0.44	0.42	0.46	0.39	0.44	0.49	0.43

Note. Within each column, means with different subscripts differ significantly by the χ^2 criterion at $p < 0.01$.

Table 3
Successful answers to original PIRLS and PISA test questions and those added for the Push–Pull method.

	Questions taken from PIRLS and PISA			New questions, constructed by the authors of the Push–Pull method		
	Four texts	Two PIRLS texts	Two PISA texts	Four texts	Two PIRLS texts	Two PISA texts
<i>The total number of questions</i>						
	31	22	9	20	9	11
<i>Answered correctly, % of the maximum score</i>						
Grade 4	54.1 ^a	70.9 ^a	34.3 ^a	54.7 ^a	42.8 ^a	61.6 ^a
Grade 6	55.0 ^a	71.9 ^a	35.6 ^a	55.4 ^a	46.5 ^a	60.4 ^a
Grade 9	65.4 ^b	78.6 ^b	50.0 ^b	66.6 ^b	53.9 ^b	74.0 ^b
<i>Average item discrimination for questions</i>						
Grade 4	0.40	0.40	0.39	0.42	0.45	0.40
Grade 6	0.42	0.42	0.42	0.44	0.47	0.42
Grade 9	0.44	0.40	0.49	0.45	0.47	0.44

Note. Within each column, means with different subscripts differ significantly by the χ^2 criterion at $p < 0.01$.

We presumed that the results of our study would vary by age group, with the lowest scores in fourth graders, the highest, in ninth graders, and at the mid-point in sixth graders. Such expectation was based on two arguments. First, in some countries (e.g., in Norway and Iceland) one year of schooling provides significant growth in reading literacy between grades 4 and 5 (Martin, Mullis, & Foy, 2011). Our second argument is derived from the educational content in grades 5 and 6 of the Russian schools. In these two years, students begin systematic and regular work with textbooks, and using these informational texts in their studies becomes a daily routine of their school life. Students regularly receive homework assignments that require them to read a paragraph in the textbook and answer questions about its content. Teachers control these assignments practically on each lesson. The main purpose of the textbooks is to remind students about what they learned in the previous lesson, provide additional topical information on the lesson and occasionally offer an alternative line of thinking about the same topic. The questions in the textbook are designed to help students check whether they understood the content of the lesson as it is represented in the textbook.

The Push–Pull diagnostic method is aimed at assessing how students understand texts that only partly resemble material from the school textbooks. Thus, the Push–Pull scores cannot be interpreted as a direct result of working with textbooks in school. Nonetheless, for the majority of Russian fifth and sixth graders, the use of textbooks is the principal activity in which the systematic and purposeful development of reading as a way to acquire and use information may be developed. That is why we were expecting sixth graders to understand the informational texts better than fourth graders, but more poorly than ninth graders.

The overall results of testing reading literacy among the 10–15-year-olds using the Push–Pull method are presented in Table 2. This table lists the average scores (as a proportion of the maximum attainable score) when the students answered 51 questions on four texts.

The data in Table 2 does not support our initial expectation: in understanding informational texts, the sixth graders were not at the mid-point between grades 4 and 9. Sixth graders did not move ahead of fourth graders in understanding the PIRLS texts designed for younger children or the PISA texts designed for older students. Ninth graders performed better than the younger students; however, their progress was rather modest: two thirds of the standard deviation. This summary presents an overall result, and to reach a conclusion, we must analyze these data in more detail. The goal of these further analyses is to search for trends revealing the qualities of reading literacy among sixth graders.

3.2. Questions used for assessing reading literacy

The first step in the analysis of the overall results was to divide the diagnostic questions into two groups: those from original PIRLS and PISA tests, and those that supplemented the Push–Pull method. The results of this division are presented in Table 3, listing the average scores received by the students for each type of question (as a percentage of the maximum attainable scores).

When the questions were grouped as shown in Table 3, we failed to discern a particular group in which sixth graders' performance exceeded that of fourth graders. Nevertheless, this analysis helped us verify our measurement tool. The data presented in Table 3 show that the Push–Pull method is in line with the original design and really assesses reading literacy. In this way we bridge the gap between the PIRLS questions addressing younger schoolchildren and the PISA questions designed for older students. Based on these results, we can claim that the questions we added to the PIRLS texts were more difficult to answer than the original PIRLS questions, and those added for PISA texts were easier than the original PISA questions. Meanwhile, on the whole, the added and genuine questions were well balanced by their difficulty.

3.3. Reading skills assessed through the Push–Pull method

The next step in analyzing the overall results was to differentiate the array of diagnostic questions for the tested processes of comprehension. The data thus obtained are presented in Fig. 1.

The evidence presented in Fig. 1 clearly demonstrates that fourth and sixth graders did not differ in their ability to understand informational texts. The results manifested by ninth graders were significantly better than those attained by the two younger groups of students.

The data in Fig. 2 bring a magnifying glass to this case. We see an odd disproportion in developing reading skills in ninth graders. The progress in retrieving information from the text is less evident than in the two other reading skills.

Let us remind the readers that three processes of comprehension differ dramatically in their psychological nature and most probably, in the ways they are developed in the process of education. The ability to retrieve information from the text is more closely related to the reader's imagination and ability to mentally reconstruct what is written in the text. The ability to integrate and interpret the text message mainly addresses the reader's cognitive faculties. Connecting the text message to the whole picture and clarifying the meaning depend, above all, on the abilities of the reader to identify any uncertainty or

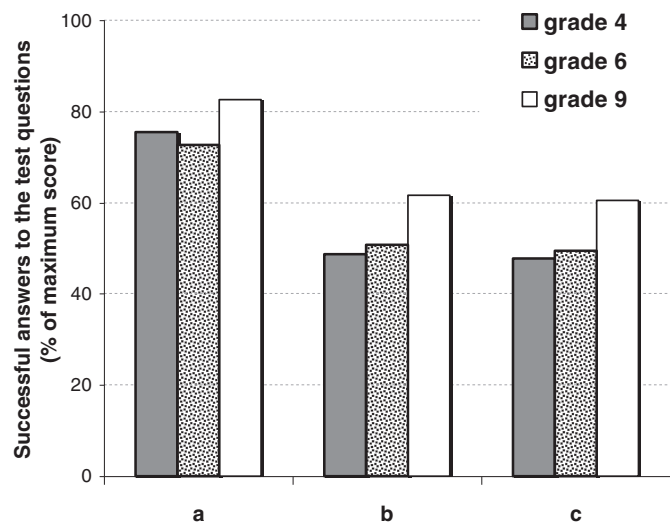


Fig. 1. Assessment of reading literacy with the Push–Pull method. Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

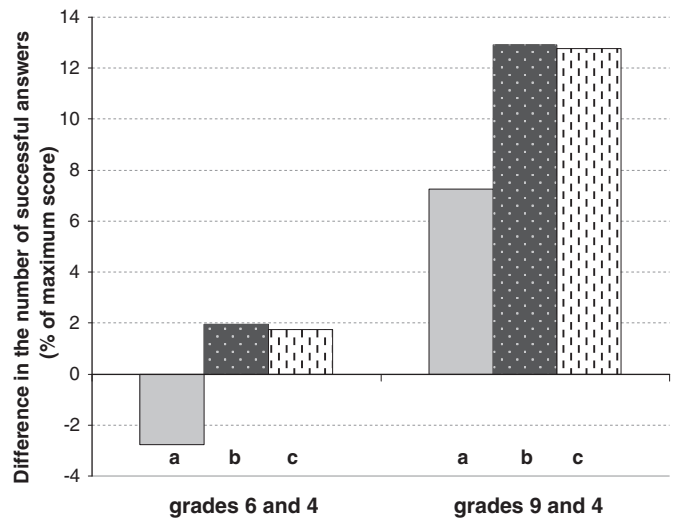


Fig. 2. Progress in reading skills when answering the Push–Pull questions. Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

ambiguity of the message provided by the text. The ability to reflect on the messages in the text and comprehend them is connected largely with such personal characteristics of the reader as his or her critical mind or independent judgments.

When surveying comprehension processes in fourth, sixth, and ninth graders, we found that the processes based primarily on readers' intellectual abilities developed over the years more rapidly than the basic, initial process of constructing an imaginary picture of the reality described by the text. In other words, ninth graders, when compared to fourth graders, excel in the mental processing of the messages retrieved from the texts. The expected growth in the ability to retrieve these messages from the text is slower. Based on our experimental data, the major ability of the readers to retrieve information from the text begins to lag in grades 5 or 6. The data collected by Kuznetsova (2009) presume that this slowdown starts earlier, in elementary school.

It would have been surprising if 14–15-year-olds did not surpass 9–10-year-olds intellectually. Similar advantage was also observed when comparing 10- and 15-year-olds who never experienced formal schooling (Cole & Cole, 1993). In other words, it is uncertain whether the improvements in reading literacy discerned in our experiment resulted from reading instruction in particular, from school education in general or from normal intellectual development.

3.4. Texts used for assessing reading literacy

We further broke down the data in Fig. 2 into two groups: the questions based on PIRLS texts and those on PISA texts. We remind the reader that the PIRLS texts are written to match the life experience of a 10-year-old student, whereas the PISA texts reflect on some aspects of youth and adult life. None of these texts is particularly age-specific – they do not assume any special academic knowledge, but they differ in lexical simplicity.

Fig. 3 presents a pattern that we have already discussed: results of the Push–Pull method to assess reading literacy did not differentiate fourth and sixth graders, while ninth graders performed better than the younger students. The most striking way in which ninth grade scores differed from those of fourth graders was in the ability of the former to integrate and interpret the message in the texts. This progress was more pronounced when the students worked with the “adult” PISA texts. The material in these texts allowed the ninth graders to show their ability to reflect on the information. The progress in the ability to retrieve information from texts (especially from the simpler PIRLS texts) was less considerable. The drop in this

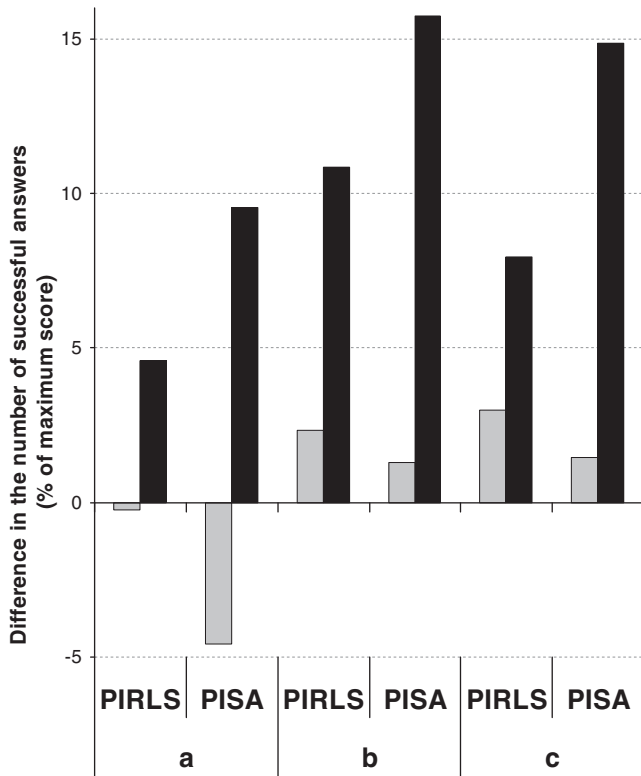


Fig. 3. Progress in reading skills when answering the Push–Pull questions, based on PIRLS and PISA texts. Gray columns present the difference between grades 6 and 4, black – the difference between grades 9 and 4. Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

basic reading skill among sixth graders is particularly notable when they worked with the “adult” PISA texts.

3.5. The most and least successful readers

Another way to analyze our results is to range the students by the level of achievement. We subdivided the total sample into the following three groups:

- A quarter of the sample comprising the most successful students, or the 75th percentile.
- A quarter of the sample comprising the least successful students, or the 25th percentile.
- All other students with the average level of success, or half the sample.

Figs. 4 and 5 illustrate the results of such analysis.

Such grouping of the assessment data (Figs. 4 & 5) highlights the conclusion from our previous analysis of the experimental data:

- The results of sixth graders did not surpass those of fourth graders.
- Ninth graders demonstrated significantly better reading skills, as shown by all indices measured by our method.
- The ability to retrieve information from text was the Achilles' heel for sixth graders: here, three quarters of the group were inferior to younger students.
- The fourth and ninth graders differed in their ability to retrieve information from text. Yet this difference was smaller compared to other processes of reading comprehension.

3.6. Integration and interpretation of textual information in grades 4 and 6

In the last attempt to hit upon improvements in reading literacy in grade 6, we turned to the questions that sixth graders answered

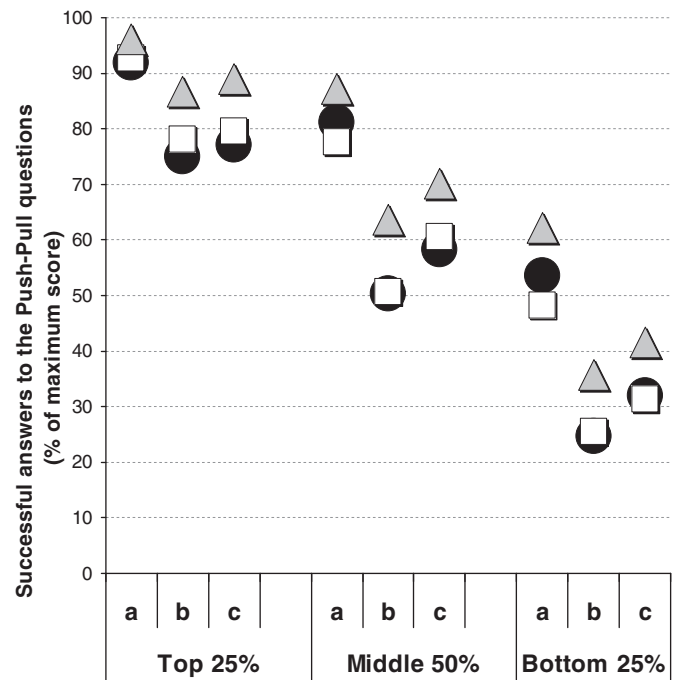


Fig. 4. Difference in reading skills in three groups of students from grades 4 (•), 6 (□) and 9 (△). Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

better than fourth graders. The number of such questions is presented in Table 4. Questions were grouped according to statistically significant differences (by the criterion χ^2 at $p < 0.05$) between grades 4 and 6 when assessed by the Push–Pull method.

The data presented in Table 4 presume some advantage of sixth graders over fourth graders as regards their reading skills involving higher-order intellectual processes: integrating and interpreting the messages from the text, reflecting on these messages and evaluating them from a personal point of view. Nonetheless, such analysis did not return clear-cut results. We therefore attempted to look at these data from another angle. Fig. 6 illustrates the profiles of the ability to integrate and interpret messages from the text in grades 4 and 6: a skill apparently prevalent in older students. A high level of skill in

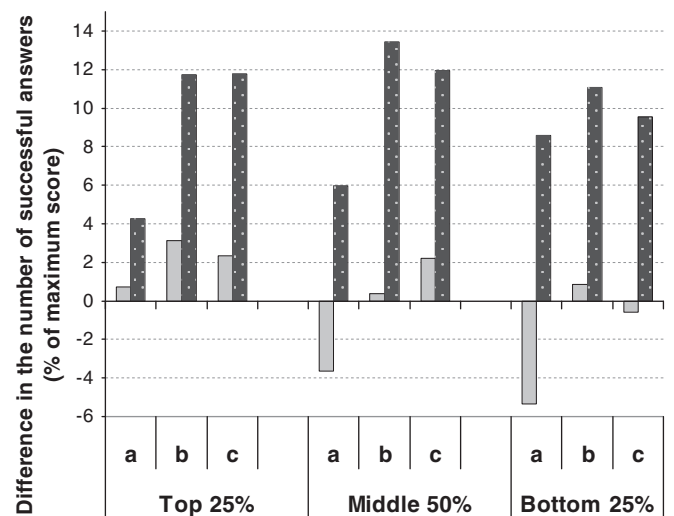


Fig. 5. Progress in reading skills in three groups of students. Gray columns present the difference between grades 6 and 4, black – the difference between grades 9 and 4. Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

Table 4
Classification of the Push–Pull questions, the number of questions (in parentheses, % of total number of questions for each reading skill).

Grades 4 vs. 6	Reader's ability to			All questions
	Retrieve information and make straightforward inferences	Interpret and integrate information	Reflect on the content and form of the text and evaluate them	
4 = 6	10 (67)	16 (67)	8 (67)	34 (67)
4 < 6	1 (7)	6 (25)	3 (25)	10 (19)
4 > 6	4 (27)	2 (8)	1 (8)	7 (14)

this case is presented as the interval between 80% and 100% success in answering the Push–Pull questions that diagnose the ability to integrate and interpret the meaning in texts. In this case, the above-average group scored 60–79%, the average group, 40–59%, and the below-average group, 20–39%, with the bottom-group score at 0–19%.

We see that somewhere between grades 4 and 6, our readers split into two groups:

- students who were generally successful in combining separate parts of textual information and explaining them on their own (high and above-average levels);
- students who were generally unsuccessful in these operations (low and below-average levels).

The number of successful readers remained largely unchanged through two years of survey: 39% of fourth graders and 41% of sixth graders, which answered questions involving the retrieving of information and integration and interpretation of the message, received 60–100% of allotted scores. The proportion of unsuccessful students (0–39%) increased significantly over these two years of schooling, from 36% for grade 4 to 43% for grade 6.

This result is rather paradoxical. On one hand, all manuals for IQ tests insist on a significant advance between the ages of 10 and 12 (Cattell, 1971). On the other hand, the number of sixth graders who seemingly did not use their improved intellects to understand the message in texts also increased, at least regarding the texts students read at their school desks.

3.7. Are problems of adolescence alone to be blamed for the test results?

Here we present several attempts to elucidate any advantage of the Russian sixth graders as regards understanding texts, and these attempts failed. The most disappointing fact we elucidated was the lack of progress through two years of education in the middle school (grades 5 and 6) when students' ability to understand informational texts did not improve. The first thing we consider is the psychological characteristics of 12-year-olds. Their withdrawal from school pursuits, internal resistance to activities imposed by school authority, our tests in particular, basic disloyalty to adults in power in the educational settings (Simmons & Blyth, 1987; Zuckerman, 2007) – all these characteristics vividly recognized in sixth graders, could affect the results of our study.

It is well known that fourth graders usually respond to the instructions by an adult (in this particular case, to the task to read the text and answer the questions) in a serious, diligent and businesslike manner. Sixth graders often respond selectively to school assignments: not every task motivates them to work in full force and to full capacity, and sometimes the work is done carelessly. Neglect of adolescents towards externally imposed assignments would easily explain why the answers to diagnostic questions to the texts in the grade 6 were no better than in the grade 4. Before accepting such an explanation, we divided our experimental sample comprising 30 schools into two parts. The criterion for such subdivision was the overall success of fourth graders at each school in answering the diagnostic questions of the Push–Pull texts. The data presented in Fig. 7 show that in grade 4, two groups of schools differed most notably in their understanding of information texts. Grades 6 and 9 did not differ by these indices. In other words, students entered middle school at two considerably different levels in reading comprehension, and left middle school at a uniform lower level.

Can we attribute this erasing effect of middle school education to adolescent characteristics? The data presented in Fig. 8 do not conclusively support such a presumption. In the schools with fourth graders manifesting high levels of understanding informational texts, sixth graders behaved like typical teenagers: failed to answer the diagnostic questions at their best, with due consideration and thought. In the schools where fourth graders demonstrated a low level of understanding informational texts, sixth graders behaved quite differently:

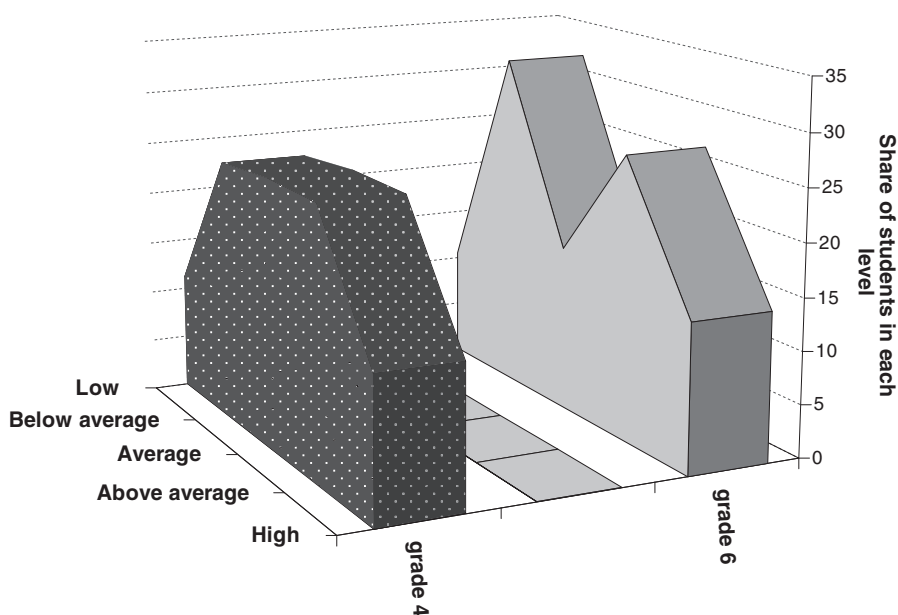


Fig. 6. The profiles of the ability to integrate and interpret the messages from four informational texts in the Push–Pull test assessed in the grades 4 and 6.

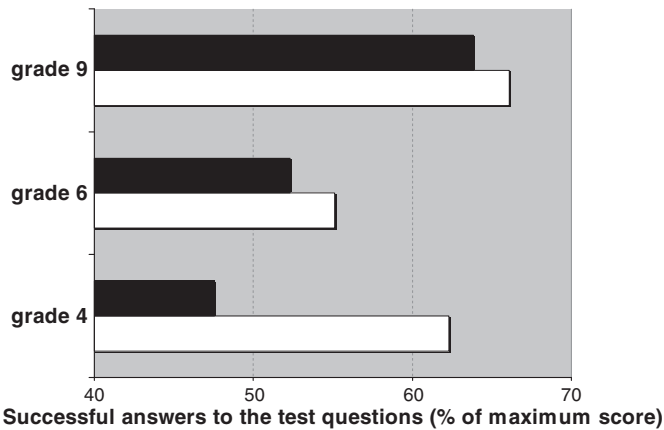


Fig. 7. The changes in reading competence in the schools that differed in the level of students' understanding of informational texts. White columns: when graduating from the elementary school students scored in Push–Pull above average in our sample (14 schools). Black columns: when graduating from the elementary school students scored in the Push–Pull below average in our sample (16 schools).

they responded to all types of questions slightly better than the diligent and zealous elementary schoolchildren. These differences between schools with different levels of text understanding at the end of the elementary education do not allow us to accept the apparently self-evident and convenient explanation of our observations by attributing the educational failures to teenage years.

We assume that the lack of progress in understanding of informational texts during the initial two years in middle school followed by minor progress through the next three years of schooling was due to the inadequate teaching for the comprehension of informational texts rather than to the psycho-physiological characteristics of students.

Obviously, many dark sides in the school life of a sixth grader can be attributed to puberty. Nonetheless, in discussing the development of reading literacy, we cannot underestimate the educational factors, and above all, learning from and with the textbooks (Polivanova, 2005). If education does not improve reading literacy through two initial years of the middle school, something is wrong with the textbooks and the ways teachers use these textbooks in the classrooms!

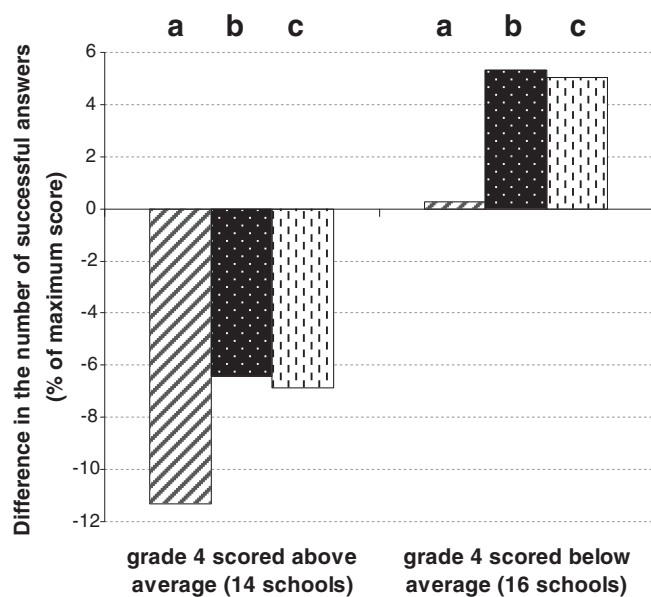


Fig. 8. Difference between grades 6 and 4 when answering the questions to Push–Pull texts. Reading skills: a – retrieve information, b – integrate and interpret, c – reflect and evaluate.

Let us ponder on what takes place, for example, at history and geography lessons. The Russian students start out on these subjects in grade 5. Do they commence at the same time learning to read historical and geographical materials that, unlike the texts used in the elementary school, extend far beyond the knowledge of a 10-year-old reader and are aimed at expanding this knowledge? If not, then in what way do students develop the general ability and practice of retrieving new information from texts, putting separate pieces of information into a new comprehensive pattern, reflecting on the messages from the text, and evaluating the completeness and accuracy of their understanding the ideas that transgress the limits of intellectual, social or personal scope of the reader? We remind the reader that, judging by our data, it is the progress in the basic reading skill of retrieving information from texts that causes the most serious concern in the Russian students.

In other words, our search for growth points of literacy in sixth graders and for their advantages over fourth graders reached a dead end. As regards the understanding of informational texts, we failed to discern in sixth graders any clear advantage over fourth graders. It goes without saying that the tools used for this search can be improved: our method for measuring literacy is far from perfect (see Table 1). By expanding the test sample, we will most probably find schools with sixth graders who understand the texts better than fourth graders. However, despite the obvious need for further research, this study has already unearthed a substantial fact beyond purely academic domain.

3.8. The crucial evidence from our longitude experiment

633 grade 4 students who participated in the Pull–Push study in 2010 were tested with the same diagnostic tool in 2012 when ending grade 6. Table 5 presents the results of this longitudinal study. It gives additional support to the fact revealed in our initial study: stagnation of reading literacy through two years of schooling.

4. Conclusions

Readiness to move on to the next level of education depends largely on an individual's level of reading literacy. The following steps are usually highlighted in PIRLS and PISA as the development trajectory for this skill:

- Learning to read, which according to the global community of educators is mostly completed by the end of the 4th year of formal schooling. (In Russia, this stage coincides with graduating from the elementary school). The PIRLS test evaluates the completeness of this development and the readiness of the students to move on to the next level.
- Learning to “read for learning” until the students need to choose their own path by transition from the compulsory education to the next,

Table 5 Longitudinal data for 633 students.

Grades	Reader's ability to			All questions
	Retrieve information and make straightforward inferences	Interpret and integrate information	Reflect on the content and form of the text and evaluate them	
<i>Answered correctly, % of the maximum score</i>				
Grade 4	76.0	52.7	50.9	54.2
Grade 6	74.0	54.6	53.3	55.8
<i>Standard deviations</i>				
Grade 4	27.1	25.7	32.5	19.3
Grade 6	28.0	24.9	32.5	20.2
Effect size (Cohen's d)	0.073	0.075	0.074	0.081

vocational or general level. (In Russia, this stage corresponds to the middle school.) The PISA test evaluates how well completed is this stage of developing reading literacy and preparing students for transition to next developmental tasks.

- Learning to “read for life”, i.e. to solve a wide range of practical, social and educational tasks as an imminent stage towards successful self-realization in the modern world.

The Russian students demonstrated the highest level of readiness for the first educational transition (as manifested in the PIRLS survey) and the low level of readiness for the next transition (based on PISA evidence). Indeed, the educational transition depends heavily on intellectual and personal development of students. An indirect evidence for successful educational transition of Russian students from grade 4 to grade 8 is provided by TIMSS: a gap between younger and elder students in mathematics and science is very tiny (Martin, Mullis, Foy, & Stanco, 2012; Mullis, Martin, Foy, & Arora, 2012). This pattern is dramatically different from what we observe in the PIRLS–PISA case. We therefore assume that some problems in the educational process tend to restrain and retard the development of reading literacy per se at this stage of schooling (grades 5 to 9), the period when the academic work using textbooks is progressively mounting up.

With the help of the method developed by fusing PIRLS and PISA, we assessed literacy in grades 4, 6 and 9 and established at least three clear-cut facts:

- Sixth graders do not differ from fourth graders when it comes to their understanding of informational texts.
- We cannot attribute this failure to move forward in understanding texts to pubescence problems.
- Ninth graders performed better than fourth and sixth graders; nevertheless the progress in reading literacy over five years was only 5 to 16% when assessed by selected indicators (see Figs. 3 & 5).

To apprehend the progress in development of reading literacy in the Russian middle school, let us compare it with the elementary school. Just imagine that after two initial years of schooling, the elementary schoolchildren demonstrated the same reading skills as when they started school and that when graduating from the elementary school, their reading ability improved only by 16%. Everybody would perceive it as a failure.

Our experimental data certainly require additional validation and a deeper analysis. Presently the initial surmise that teaching literacy in the middle school is burdened with serious problems has gained considerable experimental support. Here we took a passing look into this black hole of school education, wherein the best achievements of the Russian elementary schools in producing good readers disappear year after year. This black hole gapes at the transition step from the elementary to middle school. Lost in this hole are at least two years of growth in understanding texts.

The main message of this paper was to present an efficient Push–Pull method for monitoring reading literacy in the age interval between PIRLS and PISA assessments using indices compatible with these two prominent tests. In this study, we elucidated crucial limitations in middle school reading education.

Acknowledgements

The authors express their sincere gratitude to A.V. Postulgin, Director of the Tambov region government agency “Center of Expertise in Educational Activities”, and the staff of this center for hosting the experiment by the Push–Pull method. We thank V.Y. Baranova, N. Nurminsky, N.G. Koshelenko, and the staff of the Center for Evaluating the Quality of Education of the Institute of Content and Methods of Education of the Russian Academy of Education, who helped run the “Push–Pull” survey and statistically process the data.

The study was conducted in 2010 within the framework of the fundamental research program “Development of an integrated toolkit for evaluating school strategies to achieve reading literacy among students in primary and middle schools” (74/07-10, dated June 24, 2010) launched by the Higher School of Economics (HSE) and the Center for International Cooperation in Educational Development (CICED) at the Academy of National Economy.

References

- Brodsky, J. A. (1987). Nobel Lecture. http://www.nobelprize.org/nobel_prizes/literature/laureates/1987/brodsky-lecture.html
- Cahan, S., & Davis, D. (1987). A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal*, 24(1), 1–12.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.
- Cole, M., & Cole, S. (1993). *The development of children*. N.Y.: Scientific American Books.
- Frumin, I. D. (Ed.). (2010). *Neozhidannaya pobeda: Rossiyskiye shkol'niki chitayut luchshe drugikh [Unexpected victory: Russian students are better at reading than others]*. Moscow: Vysshaya Shkola Ekonomiki Publ.
- Kasprzhak, A. G. (Ed.). (2006). *Rossiyskaya shkola: ot PISA 2000 k PISA 2003 [Russian schools: From PISA 2000 to PISA 2003]*. Moscow: Logos.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kovaleva, G. S. (Ed.). (2004). *Novyi vzglyad na gramotnost' po resul'tatam mezhdunarodnogo issledovaniya: PISA 2000 [A new look at literacy based on international research: PISA 2000]*. Moscow: Logos.
- Kovaleva, G. S. (Ed.). (2006). *Kachestvo obshchego obrazovaniya v rossiyskoi shkole po resul'tatam mezhdunarodnogo issledovaniya [The quality of education in Russian schools: results of international research]*. Moscow: Logos.
- Kuznetsova, M. I. (2009). Sil'nye i slabye storony chitatel'skoi deyatelnosti vypusknikov rossiyskoi nachal'noi shkoly po resul'tatam PIRLS 2006 [The strong and weak sides of reading among Russian elementary school graduates based on PIRLS 2006]. *Voprosy Obrazovaniya [Journal of Educational Studies]*, 1, 107–136 (Moscow).
- Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: Regression-discontinuity applied to TIMSS-95. *Oxford Review of Education*, 32(3), 397–429.
- Luyten, H., Peschar, J., & Coe, R. (2008). Effects of schooling on reading performance, reading engagement, and reading activities of 15-year-olds in England. *American Educational Research Journal*, 45(2), 319–342.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2011). Age distribution and reading achievement configurations among fourth-grade students in PIRLS 2006. *IERI Monograph Series Issues and Methodologies in Large-Scale Assessments*, 4, 9–34.
- Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievements in primary school in 35 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011. Assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- NCES 2010-458 (2010). Alliance for excellent education analysis of U.S. Department of Education Statistics. *The Nation's Report Card: Reading 2009*. Washington, DC: Government Printing Office.
- OECD (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris, France: OECD Publications.
- OECD (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: OECD Publications.
- OECD (2006). *PISA 2006: Science competencies for tomorrow's world*. Paris, France: OECD Publications.
- OECD (2010a). *PISA 2009 framework: Key competencies in reading, mathematics and science*. Paris, France: OECD Publications.
- OECD (2010b). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (Volume I)*. Paris, France: OECD Publications.
- OECD (2010c). *PISA 2009 results. Learning trends: Changes in student performance (Volume V)*. Paris, France: OECD Publications.
- Pinskaya, M. A. (2009). Analiz uchebnykh posobii dlya nachal'noi shkoly [An analysis of elementary school textbooks]. *Voprosy Obrazovaniya [Journal of Educational Studies]*, 1, 137–161 (Moscow).

- Polivanova, K. N. (Ed.). (2005). *Novye trebovaniya k sodержaniyu i metodike obucheniya v rossiyskoi shkole v kontekste rezultatov mezhdunarodnogo issledovaniya PISA 2000. [New requirements for content and teaching methods in the Russian schools in the context of PISA-2000 results.]* Moscow: Universitetskaya Kniga.
- Russian Academy of Education. Center for educational quality assessment. Information on PIRLS and PISA results. <http://www.centeroko.ru/>
- Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York: de Gruyter.
- Vygotsky, L. S. (1992). In M. O'Connor, & R. Silverman (Eds.), *Educational psychology*. Winter Park, FL: PMD Publications.
- Zuckerman, G. (2007). School students age ten to twelve: A "no man's land" in developmental psychology. *Journal of Russian and East European Psychology*, 45(3), 70–93.