Психологическая наука и образование
2022. Т. 27. № 5. С. 107—120
DOI: https://doi.org/10.17759/pse.2022270509
ISSN: 1814-2052
ISSN: 2311-7273 (online)

Psychological Science and Education
2022. Vol. 27, no. 5, pp. 107—120
DOI: https://doi.org/10.17759/pse.2022270509
ISSN: 1814-2052
ISSN: 2311-7273 (online)

Reading Skills of First Graders in Russia and Kazakhstan: a Cross-Cultural Study

Alina E. Ivanova
National Research University "Higher School of Economics", Moscow, Russia
ORCID: https://orcid.org/0000-0003-3340-7651, e-mail: aeivanova@hse.ru

Inna V. Antipkina
National Research University "Higher School of Economics", Moscow, Russia
ORCID: https://orcid.org/0000-0003-4865-3433, e-mail: iantipkina@hse.ru

Saule J. Abisheva
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
ORCID: https://orcid.org/0000-0002-4497-0805, e-mail: s.abisheva@mail.ru

Madina K. Abayeva
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
ORCID: https://orcid.org/0000-0001-6463-9354, e-mail: madina_abaeva@mail.ru

This article assesses the intercultural comparability of reading assessment results taking into account the specifics of the test content in relation to the child's cultural environment. The reading skills of first graders in two countries were assessed using the reading scale of the computerized instrument "Start". The sample of students from Kazakhstan included 1102 first-graders from Russian-language schools in the city of Almaty. The sample of students from Russia included 2247 first-graders from the city of Novosibirsk. Pearson reliability and Chronbach's alpha were in the range from 0.89 to 0.96. Subsequently, Differential Item Functioning analysis was carried out on a combined sample in order to investigate whether the scale tasks work identically for the students from Russia and Kazakhstan when the levels of their reading skills are taken into consideration. Logic regression showed that there are no items with DIF effect size reaching beyond 0.13 (under Zumbo-Thomas classification). The research outcomes may be of interest to international comparative studies of reading skills development.

Keywords**:** international comparative studies, early reading skills, intercultural comparability, primary school, adaptation.

For citation: Ivanova A., Antipkina I., Abisheva S., Abayeva M. Reading skills of first graders in Russia and Kazakhstan: a Cross-Cultural Study. *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2022. Vol. 27, no. 5, pp. 107—120. DOI: https://doi.org/10.17759/pse.2022270509 (In Russ.).

Навыки чтения первоклассников в России и Казахстане: кросс-культурное исследование

Иванова А.Е.
ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0003-3340-7651, e-mail: aeivanova@hse.ru

Антипкина И.В.
ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0003-4865-3433, e-mail: iantipkina@hse.ru

Абишева С.Д.
Казахский национальный педагогический университет имени Абая,
г. Алматы, Казахстан
ORCID: https://orcid.org/0000-0002-4497-0805, e-mail: s.abisheva@mail.ru

Абаева М.К.
Казахский национальный педагогический университет имени Абая,
г. Алматы, Казахстан
ORCID: https://orcid.org/0000-0001-6463-9354, e-mail: madina_abaeva@mail.ru

Представлены материалы исследования межкультурной сопоставимости заданий теста по чтению и культурной среды ребенка. Оценивание навыков чтения первоклассников в двух странах происходило с использованием шкалы чтения компьютеризированного инструмента «Старт». Выборка учащихся из Казахстана представлена первоклассниками из русскоязычных школ города Алматы, N=1102 ребенка. Выборка учащихся из России представлена первоклассниками города Новосибирска, N=2247 учеников. Авторы показали, что задания теста навыков чтения в целом функционируют одинаково для первоклассников из билингвальной среды Казахстана и первоклассников города Новосибирска. Психометрический анализ данных обеих версий был проведен отдельно для каждой национальной выборки. Показатели классической и Раш-надежности для версий двух стран варьировали от 0,89 до 0,96. Затем на объединенной выборке был проведен DIF-анализ с целью выяснить, работают ли задания инструмента одинаковым образом для учащихся из России и Казахстана при учете их уровня подготовленности по чтению. Использованный метод логистической регрессии показал, что в тесте нет заданий с различием функционирования типа B или C (в которых размер эффекта превышал бы показатель 0,13 в классификации Зумбо-Томас). Результаты представляют интерес для международных сопоставительных исследований развития навыков чтения.

***Ключевые слова*:** международные сравнительные исследования, ранние навыки чтения, межкультурная сопоставимость, начальная школа, адаптация.

Для цитаты: *Иванова А.Е., Антипкина И.В., Абишева С.Д., Абаева М.К.* Навыки чтения первоклассников в России и Казахстане: кросс-культурное исследование // Психологическая наука и образование. 2022. Том 27. № 5. С. 107—120. DOI: https://doi.org/10.17759/pse.2022270509

Introduction

Reading is the main technology of cognition [1], so reading research is crucial for education systems. Bilingualism in the education system of Kazakhstan raises a number of questions about the comparability of student learning outcomes. This study will focus on the entry-level reading assessment test for the first-graders and its functioning in two cultures: Russian-language schools in Almaty (Kazakhstan) and Russian-language schools in Novosibirsk (Russia).

Previous large-scale studies of the reading skills of Kazakhstani schoolchildren were carried out in the framework of international comparative studies. For the first time, Kazakhstan participated in the Progress in the International Reading Literacy Study in Primary Schools (PIRLS) in 2016, with assessments conducted in Russian and Kazakh languages on the sample of 2,983 students with the Kazakh language of instruction and 1,942 students with the Russian language of instruction. PIRLS is an independent international study of the reading literacy of the 4th grade students, which is conducted by the International Association for the Evaluation of Educational Achievement (IEA) [2; 3; 4]. According to the results of PIRLS-2016, the fourth-graders of Kazakhstan took the 27th place in the ranking of 50 countries (for comparison, Russian students took the first place in the PIRLS-2016 ranking), girls showed higher results than boys (the same trend as in 48 out of 50 other participating countries), children worked better with information than with literary texts [5]. Based on the results of the PIRLS-2016 study, it was concluded that "Kazakhstani children receive insufficient experience in reading literary texts" [6]. In the Program for International Student Assessment (PISA)-2018, students in Kazakhstan scored an average of 387 in reading literacy, below the OECD average of 487 [7; 8].

International studies of reading skills PIRLS and PISA have raised interest to reading studies in Kazakhstan [9]. In recent years, the country has paid considerable attention to the issues of children's reading [10; 11]. 2019 was declared the National Year of Reading in the country, in 2020 the project "Reading School — Reading Nation" was launched, 2021 became a year of children's reading support in Kazakhstan.

Within the framework of the state program of multilingualism, the Russian language, like Kazakh and English, is mandatory in school curricula for the entire period of study (grades 1—11), regardless of the language of instruction. Of the 5,807 schools in the republic, 1,885 are schools with Russian as the language of instruction. 2,147 out of the total number of schools are mixed language schools. In order to obtain objective data and be able to compare schools with two language bases, it was decided to monitor the formation of reading skills in two stages.

• Stage 1: using the START tool to study the reading skills of students in the 1st grade of schools with the Russian language of instruction in Almaty. Today, only the information about reading skills of the fourth-graders (PIRLS) and fifteen-year-old students (PISA) available in Kazakhstan only.

• Stage 2: develop the Kazakh version of the START / BASTAU test and assess reading skills among students of primary schools with the Kazakh language of instruction.

This article presents the results of the first stage of the study. To assess first-graders, we used the reading scale from the START instrument [12], which was developed on the basis of a localized Russian version of the British iPIPS tool [13; 14], which is also used in Australia, Brazil, Germany, South Africa [15; 16; 17; 18].

In order to use the Russian-language instrument in schools with the Russian language of instruction in Kazakhstan, it is necessary to prove that in the bilingual culture of Kazakhstan, all items of the Russian-language scale function the same way as for the Russian-speaking students of Russian schools.

Literature review: world experience in research on the comparability of assessment tools

The cultural environment can have a significant impact on assessment results that are standardized on other groups [19]. Culture is defined as "learned meanings and patterns of behavior shared by members of a group, which are transmitted through social activity for the purpose of social adaptation, growth and development" [20]. In ability tests, not a single aspect of assessment is free from cultural effects: the content of the test, stimulus materials, construction of phrases, content of instructions, behavior of participants during testing and experts when assigning scores (if needed) [21]. Thus, any test contains cultural specificity. Finding out whether this specificity is an obstacle to the fair assessment of people from different cultural groups is the goal of equivalence studies.

A number of examples shows the lack of comparability of results for individual countries and constructs in major international studies [22; 23; 24; 25]. Even if countries use the same language and the test does not need to be translated, this does not eliminate the risks of cultural differences in assessment results [26]. For example, the English version of the iPIPS tool has been adapted for use in Australia despite the common language (English), because DIF analysis demonstrated functioning of some items not in favor of students from the Aboriginal Australian group [18]. Within the framework of the same international iPIPS project (underlying the Start), studies were carried out on the comparability of the Russian and English versions of the instrument for the mathematical part of the test [25], in which differences in the functioning of several items were revealed, due to which these items were excluded from secondary analyses.

Cross-cultural equivalence of measurements is investigated at three levels [27]:

- construct level (what is being measured),
- the level of a measurement instrument (how is being measured)
- and the level of a measurement scale (in what units it is measured).

In our study, for two cultures — Russia and Kazakhstan — the same Russian-language instrument and the same psychometric analysis procedures were used, so the construct and scale levels are expected to be identical; this allows comparability analysis to be focused on the instrument level.

There are three possible sources of intercultural bias in the assessment results [28]:

1) Sample differences. For example, samples may be not comparable if countries have different rules for disability inclusion at schools.

2) Test differences. For example, in item wordings, realities that are well known in one culture, can be exotic for another country.

3) Differences in data collection procedures. A classic example of such procedural violations was the testing of children from Nigeria with Rowan's matrices "on the thresholds of houses, in hallways, under trees" with the help of untrained staff, which was very different from the conditions in which children were assessed in European countries [29].

In our study, the conditions for data collection were standardized by a single computer environment, voiceover of the instrument by a professional speaker (interviewers were not required to read instructions), and the same training procedures for interviewers. The samples of children were comparable in age (6-7 years old) and status (just started schooling). Thus, the goal of this study was to use statistical methods to assess the threats to intercultural comparability of assessment results due to the content of test items.

Materials and Methods

Assessment instrument. "Start" is a computerized reading assessment tool with a semi-adaptive task presentation algorithm [12; 30], so if a child makes a certain number of mistakes, the assessment stops so as not to demotivate the child. The assessment involves individual sessions of a child with a specially trained interviewer. The reading scale includes 35 tasks and covers the following areas:

* Letter recognition. Russian letters are presented one by one on the screen, the voice of the actor asks the child to name the letter. The correct answer is the name of the letter in the official and colloquial form (for example, “em” and “m” would be the correct answers for “M”) or the sound (“m-m”). The wrong answer is the name of the object with this letter (for example, if the child says “Meat” for “M”).

* Reading individual words

* Reading a short story (“mechanical” reading)

* Reading comprehension test. In this part of the instrument, the child reads the text, in some places of which he or she is asked to choose one of the three words that is most appropriate for the context.

The Start tool is standardized using the dichotomous Rasch model [12; 14].

Sample. The reading skills of the first graders were assessed by the Start tool in October 2019, when the children had just started school. The sample of Russian-language schools in the city of Almaty consisted of 1102 children. The sample is not representative. The sample of the first-graders in Novosibirsk consisted of 2247 students. Novosibirsk was chosen as a region with comparable geographical location, industrial development and population compared to Almaty. The sample of students in Novosibirsk was randomized and stratified by the type of school and city district.

Analysis. The comparability study of the instrument in two cultural environments (Russia and Kazakhstan) was carried out in two stages.

At the first stage, a psychometric analysis of the instrument was carried out on a sample of Russian-speaking schools in Almaty and Novosibirsk separately. The test was analyzed under the dichotomous Rasch model [31]. The model is used for dichotomous items with one correct and one incorrect answer. To assess the reliability of the scales, Cronbach’s alpha and Rasch Person reliability (Separation Reliability) are used. Cronbach’s alpha is a measure of internal consistency and captures the interrelation of the scale items [32]. Person reliability shows the reproducibility of the hierarchy of measures [33].

At the second stage, differential item functioning analysis was carried out. Differential item functioning (DIF) reveals cases when students with the same level of preparedness have different chances to complete an item correctly [34; 35]. The presence of DIF implies that an item functions in favor of one of the groups of students, despite the fact that these students have the same ability/final score. Different cultural backgrounds can be a source of differential item functioning, even if the development of the tool followed all necessary procedures in accordance with international standards [23; 36; 37].

DIF analysis contributes to the correct interpretation of the assessment results and helps to refine the tool so that it becomes fairer to all groups of students.

Results

Psychometric analysis of the data from both versions was conducted first separately for each national sample, and then on a combined sample to find out if the instrument items work in the same way for students from Russia and Kazakhstan, taking into account their level of reading proficiency.

Table 1 shows the characteristics of items for each of the two samples separately. Characteristics include task difficulty measures (columns 1, 2), fit statistics (columns 3, 4). Item codes are presented in column 5. The goodness of fit statistics are the standardized residuals which are based on estimates of response probabilities for each item. In this table we use one fit statistic — weighted goodness of fit (Infit MNSQ) statistic. As can be seen from the table, the mean values of the goodness-of-fit statistics lie within the range recommended by psychometricians [0.6; 1.4] for all items, except for the first two (letter knowledge) for the Novosibirsk sample [33]. The measurement error is somewhat higher for tasks for a sample of children from Kazakhstan. The correlation of the response to the task with the level of preparedness is high and positive.

Table 1

Item characteristics for a sample of students from Novosibirsk, Russia and Almaty, Kazakhstan

Difficulty measures (logits)		MNSQ Infit		Item
Rus	Kaz	Rus	Kaz	
-1.37	-1.92	1.53	1.25	Letter «K»
-1.37	-1.65	1.45	1.36	Letter «L»
-1.02	-1.35	1.13	1.22	Letter «e»
-1.15	-1.43	1.24	1.23	Letter «Zh»
-1.36	-.95	1.15	1.32	Letter «Z»
-.33	-.89	1.36	1.26	Letter «Sh»
-.37	-.31	1.11	1.16	Letter «Ts»
-.71	-.60	1.06	1.21	Letter «Yu»
-.05	-.09	1.25	1.24	Sign «’»
-2.44	-2.24	.92	1.05	Word «myach»
-3.15	-3.11	1.06	1.27	Word «utka»
-2.82	-3.38	.90	.95	Word «shchenok»
-2.66	-2.75	.80	.93	Word «ruka»
-2.99	-3.26	.81	.81	Word «dom»
-1.20	-.63	1.00	1.17	Word «kon»
-1.10	-.77	.95	1.00	Word «korabl»
-1.82	-1.65	.81	.86	Word «kot»
-.99	-.81	.93	.94	Word «krolik»
.22	-.02	.76	.54	Reading decoding 1
.32	.05	.76	.52	Reading decoding 2
.47	.13	.77	.52	Reading decoding 3
1.22	1.57	.87	.96	Comprehension 1
1.06	1.41	.74	.88	Comprehension 2
1.20	1.35	.71	.82	Comprehension 3
1.24	1.39	.76	.87	Comprehension 4
1.75	2.16	.90	1.01	Comprehension 5
1.76	2.00	.92	1.00	Comprehension 6
.94	.74	.62	.67	Comprehension 7
2 .29	2.08	1.09	1.03	Comprehension 8
1 .95	2.01	1.00	.97	Comprehension 9
2 .50	2.44	1.03	1.00	Comprehension 10
3 .40	3.36	1.16	1.15	Comprehension 11
1 .73	1.59	.91	.93	Comprehension 12
2 .95	3.33	1.24	1.18	Comprehension 13
1.91	2.16	.93	.94	Comprehension 14

Rasch modelling allows visualizing the characteristics of the items and respondents using the so-called “variable maps” (or “Wright maps”). In Figures 1 and 2, the vertical line represents the continuum of reading skills measured in logits. The easiest items and the least prepared students are at the bottom of the continuum, while the most difficult items and the most prepared students are at the top.

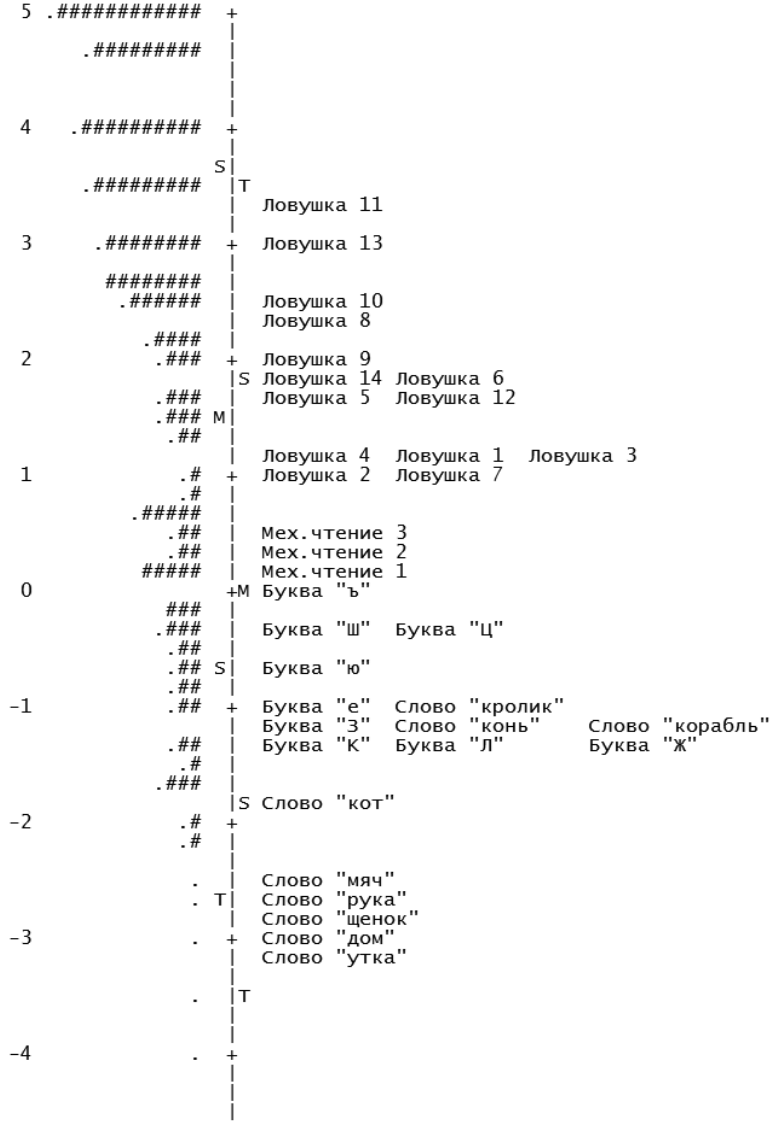

Figure 1. Item and Person map. Russia

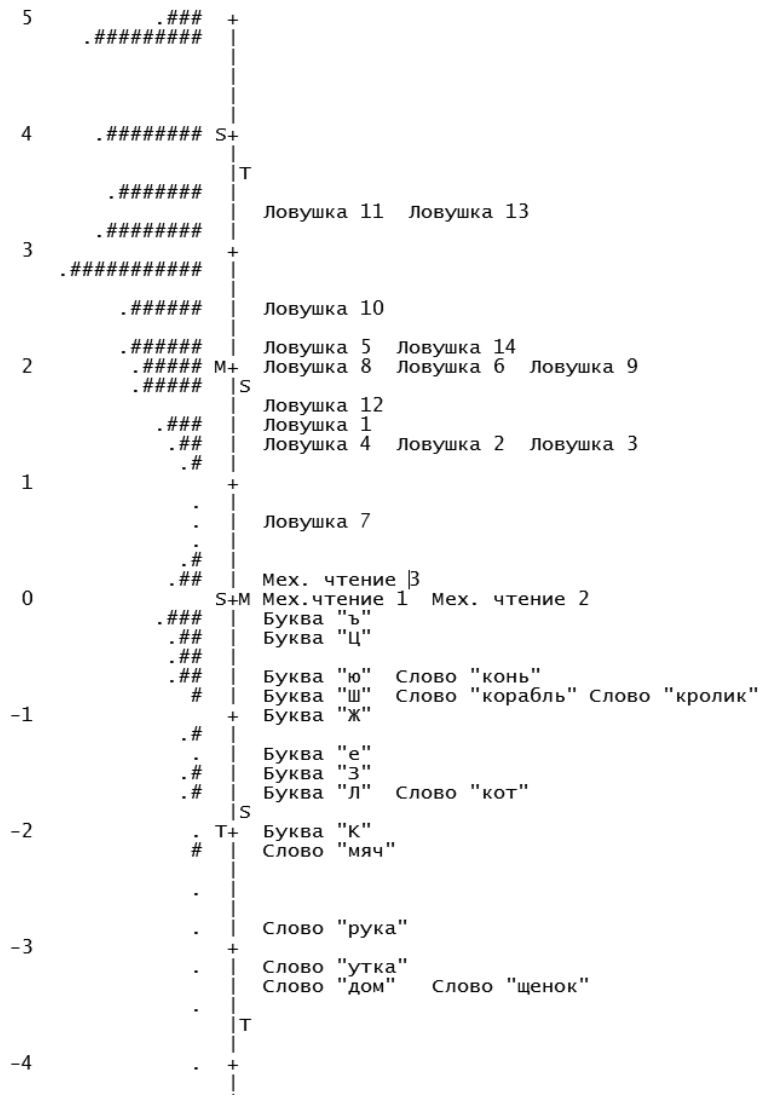

Figure 2. Item and Person map. Kazakhstan

It can be seen that both samples are characterized by a shift towards high scores for most of the children, i.e. for most students, the test was easy enough. We can also note that, in general, the hierarchy of items in terms of difficulty for the two samples is almost identical.

Table 2 presents the overall psychometric indicators of the reliability of reading scales for empirical data obtained from samples of students in Russia and Kazakhstan.

Both versions of the Start are characterized by high levels of reliability: Cronbach's

Table 2

Reliability of the scales

	Chronbach's alpha	Person Reliability	Person Separation
Russia	0.96	0.91	3.17
Kazakhstan	0.94	0.89	2.89

alpha and Person Reliability are over 0.8 which stands for 'excellent' reliability; Person Separation makes it possible to distinguish at least three groups of students according to their level of reading skill.

The good psychometric qualities of the reading scale separately on samples of students from Russia and Kazakhstan make it possible to build a single scale of test results for the first-graders from the two countries, and conduct DIF-analysis.

Within the framework of this study, DIF analysis was carried out using the logistic regression method. The method of logistic regression [35; 38] allows to detect both homogeneous (when the statistical relationship between the response to the item and the grouping variable is constant for all levels of the corresponding variable), and heterogeneous DIF (when the statistical relationship between the response to the item and the grouping variable changes for different levels of sum of scores). The method implies the statistical modeling of the probability of correctly answering the item, depending on the sequentially introduced variables: 1) the grouping variable "country" (in our case), 2) the sum of test scores, and 3) the interaction between the first and second variables. In this work, the R software, the DIFR statistical package, was used to analyze DIF [39]. The statistical significance of the model parameters was assessed using the LR test (Likelihood Ratio Test).

In the logistic regression method, an item is identified as showing a certain type of DIF when the sequential addition of a country variable and an interaction variable gives a significant improvement in the model compared to a model that includes only a variable with the sum of scores [34].

It is important to note that differences between groups of subjects may be statistically significant, but too small to have any effect on the results of the assessment. Therefore, researchers are encouraged to use a combination of indicators: the statistical significance of DIF and practical significance (or effect size) in order to make an informed decision about what to do if differential item functioning is identified.

In this paper, we use a combination of the Likelihood Ratio Test statistic significance and DIF effect size, presented in two versions — Zumbo-Thomas and Jodoin-Girl [39]. For an item to be classified as exhibiting DIF, the LR test criterion must be less than or equal to 0.01, and the effect size must be large enough. According to the Zumbo-Thomas criterion, DIF can be classified as follows: negligible type "A" DIF (change in R-squared values of two nested models is below 0.13), moderate type "B" DIF (change in R-squared values of two nested models is from 0.13 to 0.26) and a large Type "C" DIF (change in R-squared values of two nested models is above 0.26). According to a more stringent Jodoin and Girl DIF criteria, DIF can be classified as negligible type "A" DIF (change in the R-squared values of the two nested models is below 0.035), moderate type "B" DIF (change in the R-square values of the two nested models is from 0.035 to 0.07), and a large type "C" DIF (change in R-squared values of two nested models is above 0.07). Table 3 shows the results of DIF analysis using the logistic regression method.

The analysis showed that despite the statistical significance of the LR-test for a

Table 3

Differential item functioning analysis using logistic regression model

Item	LR-test	P-value	R^2	DIF size (Zumbo-Thomas)	DIF size (Jodoin -Girl)
Letter «K»	24.36	0.00 ***	0.01	A	A
Letter «L»	7.91	0.02 *	0.00	A	A
Letter «e»	7.14	0.03 *	0.00	A	A
Letter «Z»	6.78	0.03 *	0.00	A	A
Letter «Sh»	22.19	0.00***	0.00	A	A
Word «myach»	14.57	0.01 ***	0.00	A	A
Word «utka»	11.75	0.00 **	0.01	A	A
Word «shchenok»	7.71	0.02 *	0.00	A	A
Word «kon»	18.64	0.00 ***	0.00	A	A
Word «korabl»	6.99	0.03*	0.00	A	A
Reading decoding 1	24.56	0.00 ***	0.00	A	A
Reading decoding 2	29.62	0.00***	0.00	A	A
Reading decoding 3	35.69	0.00 ***	0.00	A	A
Comprehension 1	17.62	0.00 ***	0.00	A	A
Comprehension 2	26.22	0.00 ***	0.00	A	A
Comprehension 3	9.85	0.01 **	0.00	A	A
Comprehension 4	8.82	0.01 *	0.00	A	A
Comprehension 5	24.47	0.00 ***	0.00	A	A
Comprehension 6	9.26	0.01 **	0.00	A	A
Comprehension 13	18.33	0.00 ***	0.00	A	A
Comprehension 14	10.32	0.01 **	0.00	A	A

Note: *** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$

number of reading items, the size of the DIF effect is so small that it can be neglected for the practical use of the results. In other words, for students from Russia and Kazakhstan, all items of the reading test work rather the same way.

The results of any study that includes cross-cultural assessment can be correctly compared if similar scores of test participants in two countries mean a similar level of proficiency [40; 41]. This study focused on the first-graders' reading assessment tool and its functioning in two cultures: Russian-language schools in the city of Almaty (Kazakhstan) and Russian schools in the city of Novosibirsk (Russia). Our psychometric analysis showed that the tool for assessing early reading skills at the start of school functions well not only among Russian-speaking children in Russia, but also among Russian-speaking children in Kazakhstan. The scale shows similar measures for samples of children in the two countries, including model fit, reliability scores, distribution of item difficulty, and student achievement levels.

Ensuring the psychometric quality of assessment tools is a priority for the evi-

dence-based decisions in education [42]. We have shown that the cultural effect does not interfere with the difficulty of first grade reading test items in two countries. This opens up prospects for international research in Kazakhstan, with the help of Russian-language tools already standardized on Russian samples.

At the same time, it is important to note that Russian and Kazakhstani children in an equally successful way coped with most of the reading test items. The question arises: if at the start of school children from Russia and Kazakhstan have comparable reading skills in Russian, why by the end of the fourth grade there is a gap in skills demonstrated in PIRLS? A longitudinal study in Russian-language primary schools in Kazakhstan using instruments with proven psychometric properties may help answer this question.

In the future, it is planned to adapt the START tool to the Kazakh language in order to conduct a comparative study of the educational achievements of schoolchildren from schools with Russian and Kazakh as the language of instruction.

The co-authors declare no conflict of interest in the present study.

It seems that part of Conclusion / Discussion is missing. Normally, it follows the part Results and summarizes the results you got in the study. So, please arrange it accordingly.

References

1. Ivanova A.E., Kardanova-Biryukova K.S. Sozdanie russkoyazychnoi versii mezhdunarodnogo instrumenta otsenivaniya rannikh navykov chteniya [Constructing a Russian-Language Version of the international early reading assessment tool]. *Voprosy obrazovaniya [Educational Studies. Moscow]*, 2019, no. 4, pp. 93—115. DOI:10.17323/1814-9545-2019-4-93-115
2. Ivanova A.E., Kardanova E.Yu. Izuchenie vozmozhnosti provedeniya mezhstranovogo sravnitel'nogo issledovaniya navyka chteniya u uchashchikhsya na vkhode v shkolu [Checking the Possibility of an International Comparative Study of Reading Literacy Assessment for Children Starting School]. *Voprosy obrazovaniya [Educational Studies. Moscow]*, 2020, no. 4, pp. 8—36. DOI:10.17323/1814-9545-2020-4-8-36
3. Natsional'nyi doklad o sostoyanii i razvitii sistemy obrazovaniya Respubliki Kazakhstan (po itogam 2018 goda) [National report on the state and development of education systems of the Republic of Kazakhstan (based on the results of 2018)]. M. Atanaeva, M. Amanғazy, G. Nogaĭbaeva, A. Akhmetzhanova, M. Shakenova, G. Karbaeva, Zh. Dzhumabaeva, N. Kasymbekova, M. Dauliev, D. Abdrasheva, A. Kusidenova. Nur-Sultan: MON RK, AO «Informatsionno-analiticheskii tsentr», 2019, pp. 40—41.
4. Nurlanov E.B., Baigulova A.D., Kartpaev B.A., Amangazy M., Sabyruly E., Nogaibaeva G.A. «Rezul'taty Kazakhstana v PIRLS-2016»: natsional'nyi otchet ["Results of Kazakhstan in PIRLS-2016": national report]. Astana: AO «Informatsionno-analiticheskii tsentr», 2018. 155 p.
5. Avvisati F., Echazarra A., Givord P., Schwabe M. PISA Country note. Kazhakstan [Elektronnyi resurs]. *OECD Publishing*, 2019. URL: https://www.oecd.org/pisa/publications/PISA2018_CN_KAZ.pdf (Accessed 07.09.2021).
6. Bartholo T.L., Koslinski M.C., Costa M.D., Barcellos T. What do children know upon entry to preschool in Rio de Janeiro? *Ensaio: Avaliação e Políticas Públicas em Educação*, 2020. Vol. 28. Issue 107, pp. 292—313. DOI:10.1590/S0104-40362019002702071
7. Belzak W.C.M. Testing differential item functioning in small samples. *Multivariate behavioral research*, 2020. Vol. 55. Issue 5, pp. 722—747. DOI:10.1080/00273171.2019.1671162
8. Bond T.G., Yan Z., Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th ed.). New York: Routledge, 2020. 376 p. DOI:10.4324/9780429030499
9. Cheema J.R. Cross-country gender DIF in PISA science literacy items. *European Journal of Developmental Psychology*, 2019. Vol. 16, Issue 2, pp. 152—166. DOI:10.1080/17405629.2017.1358607
10. Ercikan K., Guo H., He Q. Use of Response Process Data to Inform Group Comparisons and Fairness Research. *Educational Assessment*, 2020. Vol. 25, Issue 3, pp. 179—197. DOI:10.1080/10627197.2020.1804353
11. Fahrmeier E.D. The effect of school attendance on intellectual development in northern Nigeria. *Child Development*, 1975. Vol. 46, pp. 281—285.
12. Goodrich S., Ercikan K. Measurement Comparability of Reading in the English and French Canadian Populations: Special Case of the 2011 Progress in International Reading Literacy Study. *Frontiers in Education*, 2019. Vol. 4, Issue 120. URL: https://www.frontiersin.org/articles/10.3389/feduc.2019.00120/full (Accessed 02.11.2021). DOI:10.3389/feduc.2019.00120
13. Grisay A., Gonzalez E., Monseur C. Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. In M. von Davier & D. Hastedt (Eds.). IERI monograph series: Issues and methodologies in large scale assessments, 2009. Vol. 2.

14. He J., van de Vijer F. Bias and equivalence in cross-cultural research. *Online readings in psychology and culture*, 2012. Vol. 2, Issue 2.
15. Ivanova A., Kardanova E., Merrell C., Tymss P., Hawker D. Checking the possibility of equating a mathematics assessment between Russia, Scotland and England for children starting school. *Assessment in Education: Principles, Policy and Practice*, 2018. Vol. 25, Issue 2, pp. 141—159. DOI:10.1080/0969594X.2016.1231110
16. Magis D., Beland S., Raiche G. Collection of Methods to Detect Dichotomous Differential Item Functioning (DIF). Package 'difR'. 2020. URL: http://cran.rediris.es/web/packages/difR/difR.pdf (Accessed 01.12.2021).
17. Marsella A.J., Yamada A.M. Culture and mental health: An introduction and overview of foundations, concepts, and issues. In I. Cuéllar & F.A. Paniagua (Eds.). Handbook of multicultural mental health. Academic Press, 2000, pp. 3—24. DOI:10.1016/B978-012199370-2/50002-X
18. Mullis I.V.S., Martin M.O., Sainsbury M. PIRLS 2016 reading framework. *PIRLS*, 2016, pp. 11—29. URL: https://timss.bc.edu/pirls2016/downloads/P16_FW_Chap1.pdf (Accessed 10.11.2021).
19. Niklas F., Schneider W. Home literacy environment and the beginning of reading and spelling. *Contemporary Educational Psychology*, 2013. Vol. 38, Issue 1, pp. 40—50. DOI:10.1016/j.cedpsych.2012.10.001
20. Oliveri M.E., Ercikan K., Zumbo B. Analysis of sources of latent class differential item functioning in international assessments. *International Journal of Testing*, 2013. Vol. 13, Issue 3, pp. 272—293. DOI:10.1080/15305058.2012.738266
21. Oliveri M.E., von Davier M. Toward increasing fairness in score scale calibrations employed in international large-scale assessments. *International Journal of Testing*, 2014. Vol. 14, Issue 1, pp. 1—21. DOI:10.1080/15305058.2013.825265
22. Sarmurzin Y., Amanzhol N., Toleubayeva K. et al. The impact of OECD research on the education system of Kazakhstan. *Asia Pacific Education Review*, 2021. Vol. 22, pp. 757—766. DOI:10.1007/s12564-021-09715-8
23. Schleicher A. PISA 2018: Insights and Interpretations [Elektronnyi resurs]. *OECD Publishing*, 2019. URL: https://eric.ed.gov/?id=ED601150 (Accessed 07.07.2021).
24. Stubbe T.C. How do different versions of a test instrument function in a single language? A DIF analysis of the PIRLS 2006 German assessments. *Educational Research and Evaluation*, 2011. Vol. 17, Issue 6, pp. 465—481. DOI:10.1080/13803611.2011.630560
25. Tymms P. Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading*, 1999. Vol. 22, Issue 1, pp. 27—36.
26. Van de Vijver F.J.R., Poortinga Y.H. Towards an integrated analysis of bias in cross-cultural assessment. *European journal of psychological assessment*, 1997. Vol. 13, Issue 1, pp. 29—37.
27. Wildy H., Styles I. Measuring what students entering school know and can do: PIPS Australia 2006-2007. *Australian Journal of Early Childhood*, 2008. Vol. 33, Issue 4, pp. 43—52. DOI:10.1177/183693910803300407
28. Woitschach P., Zumbo B.D., Fernández Alonso R. An ecological view of measurement: Focus on multilevel model explanation of differential item functioning. *Psicothema*, 2019. Vol. 31, pp. 194—203. DOI:10.7334/psicothema2018.303
29. Zumbo B.D. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense, 1999. URL: https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf (Accessed 01.12.2021).

Литература

1. *Иванова А.Е., Карданова-Бирюкова К.С.* Создание русскоязычной версии международного инструмента оценивания ранних навыков чтения // Вопросы образования. 2019. № 4. С. 93—115. DOI:10.17323/1814-9545-2019-4-93-115
2. *Иванова А.Е., Карданова Е.Ю.* Изучение возможности проведения межстранового сравнительного исследования навыка чтения у учащихся на входе в школу // Вопросы образования. 2020. № 4. С. 8—36. DOI:10.17323/1814-9545-2020-4-8-36
3. Национальный доклад о состоянии и развитии системы образования Республики Казахстан (по итогам 2018 года) / М. Атанаева, М. Аманғазы, Г. Ногайбаева, А. Ахметжанова, М. Шакенова, Г. Карбаева, Ж. Джумабаева, Н. Касымбекова, М. Даулиев, Д. Абдрашева, А. Кусиденова. Нур-Султан: МОН РК, АО «Информационно-аналитический центр», 2019. С. 40—41.
4. *Нурланов Е.Б., Байгулова А.Д., Картпаев Б.А., Амангазы М., Сабырулы Е., Ногайбаева Г.А.* «Результаты Казахстана в PIRLS-2016»: национальный отчет. Астана: АО «Информационно-аналитический центр», 2018. 155 с.
5. *Avvisati F., Echazarra A., Givord P., Schwabe M.* PISA Country note. Kazhakstan [Электронный ресурс] // OECD Publishing. 2019 URL:https://www.oecd.org/pisa/publications/PISA2018_CN_KAZ.pdf (дата обращения: 07.09.2021).
6. *Bartholo T.L., Koslinski M.C., Costa M.D., Barcellos T.* What do children know upon entry to pre-

school in Rio de Janeiro? // Ensaio: Avaliação e Políticas Públicas em Educação. 2020. Vol. 28, issue 107. P. 292—313. DOI:10.1590/S0104-40362019002702071
7. *Belzak W.C.M.* Testing differential item functioning in small samples // Multivariate behavioral research. 2020. Vol. 55. Issue 5. P. 722—747. DOI:10.1080/00273171.2019.1671162
8. *Bond T.G., Yan Z., Heene M.* Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th ed.). New York: Routledge, 2020. 376 p. DOI:10.4324/9780429030499
9. *Cheema J.R.* Cross-country gender DIF in PISA science literacy items // European Journal of Developmental Psychology. 2019. Vol. 16. Issue 2. P. 152—166. DOI:10.1080/17405629.2017.1358607
10. *Ercikan K., Guo H., He Q.* Use of Response Process Data to Inform Group Comparisons and Fairness Research // Educational Assessment. 2020. Vol. 25. Issue 3. P. 179—197. DOI:10.1080/10627197.2020.1804353
11. *Fahrmeier E.D.* The effect of school attendance on intellectual development in northern Nigeria // Child Development. 1975. Vol. 46. P. 281—285.
12. *Goodrich S., Ercikan K.* Measurement Comparability of Reading in the English and French Canadian Populations: Special Case of the 2011 Progress in International Reading Literacy Study // Frontiers in Education. 2019. Vol. 4. Issue 120. URL: https://www.frontiersin.org/articles/10.3389/feduc.2019.00120/full (дата обращения: 02.11.2021). DOI:10.3389/feduc.2019.00120
13. *Grisay A., Gonzalez E., Monseur C.* Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments / In M. von Davier & D. Hastedt (Eds.) // IERI monograph series: Issues and methodologies in large scale assessments. 2009, Vol. 2.
14. *He J., van de Vijer F.* Bias and equivalence in cross-cultural research // Online readings in psychology and culture. 2012. Vol. 2. Issue 2.
15. *Ivanova A., Kardanova E., Merrell C., Tymss P., Hawker D.* Checking the possibility of equating a mathematics assessment between Russia, Scotland and England for children starting school // Assessment in Education: Principles, Policy and Practice. 2018. Vol. 25. Issue 2. P. 141—159. DOI:10.1080/0969594X.2016.1231110
16. *Magis D., Beland S., Raiche G.* Collection of Methods to Detect Dichotomous Differential Item Functioning (DIF). Package 'difR'. 2020. URL: http://cran.rediris.es/web/packages/difR/difR.pdf (дата обращения: 01.12.2021).
17. *Marsella A.J., Yamada A.M.* Culture and mental health: An introduction and overview of foundations, concepts, and issues / In I. Cuéllar & F.A. Paniagua (Eds.) // Handbook of multicultural mental health. Academic Press. 2000. P. 3—24. DOI:10.1016/B978-012199370-2/50002-X
18. *Mullis I.V.S., Martin M.O., Sainsbury M.* PIRLS 2016 reading framework // PIRLS. 2016. P. 11—29. URL:https://timss.bc.edu/pirls2016/downloads/P16_FW_Chap1.pdf (дата обращения: 10.11.2021).
19. *Niklas F., Schneider W.* Home literacy environment and the beginning of reading and spelling // Contemporary Educational Psychology. 2013. Vol. 38. Issue 1. P. 40—50. DOI:10.1016/j.cedpsych.2012.10.001
20. *Oliveri M.E., Ercikan K., Zumbo B.* Analysis of sources of latent class differential item functioning in international assessments // International Journal of Testing. 2013. Vol. 13. Issue 3. P. 272—293. DOI:10.1080/15305058.2012.738266
21. *Oliveri M.E., von Davier M.* Toward increasing fairness in score scale calibrations employed in international large-scale assessments // International Journal of Testing. 2014. Vol. 14. Issue 1. P. 1—21. DOI:10.1080/15305058.2013.825265
22. *Sarmurzin Y., Amanzhol N., Toleubayeva K.* et al. The impact of OECD research on the education system of Kazakhstan // Asia Pacific Education Review. 2021. Vol. 22. P. 757—766. DOI:10.1007/s12564-021-09715-8
23. *Schleicher A.* PISA 2018: Insights and Interpretations [Электронный ресурс] // OECD Publishing. 2019. URL: https://eric.ed.gov/?id=ED601150 (дата обращения: 07.07.2021).
24. *Stubbe T.C.* How do different versions of a test instrument function in a single language? A DIF analysis of the PIRLS 2006 German assessments // Educational Research and Evaluation. 2011. Vol. 17. Issue 6. P. 465—481. DOI:10.1080/13803611.2011.630560
25. *Tymms P.* Baseline assessment, value-added and the prediction of reading // Journal of Research in Reading. 1999. Vol. 22. Issue 1. P. 27—36.
26. *Van de Vijver F.J.R., Poortinga Y.H.* Towards an integrated analysis of bias in cross-cultural assessment // European journal of psychological assessment. 1997. Vol. 13. Issue 1. P. 29—37.
27. *Wildy H., Styles I.* Measuring what students entering school know and can do: PIPS Australia 2006-2007 // Australian Journal of Early Childhood. 2008. Vol. 33. Issue 4. P. 43—52. DOI:10.1177/183693910803300407
28. *Woitschach P., Zumbo B.D., Fernández Alonso R.* An ecological view of measurement: Focus on multilevel model explanation of differential item functioning // Psicothema. 2019. Vol. 31. P. 194—203. DOI:10.7334/psicothema2018.303
29. *Zumbo B.D.* A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense. 1999. URL: https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf (дата обращения: 01.12.2021).

Information about the authors
Alina E. Ivanova, PhD in Education, Senior Researcher, Centre for Psychometrics and Measurements in Education, National Research University "Higher School of Economics", Moscow, Russia, ORCID: https://orcid.org/0000-0003-3340-7651, e-mail: aeivanova@hse.ru

Inna V. Antipkina, PhD, Researcher, Centre for Psychometrics and Measurements in Education, National Research University "Higher School of Economics", Moscow, Russia, ORCID: https://orcid.org/0000-0003-4865-3433, e-mail: iantipkina@hse.ru

Saule J. Abisheva, Doctor of Philology, Head of the Department of Russian Language and Literature, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan, ORCID: https://orcid.org/0000-0002-4497-0805, e-mail: s.abisheva@mail.ru

Madina K. Abayeva, PhD in Philology, Associate Professor, Department of Russian Language and Literature, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan, ORCID: https://orcid.org/0000-0001-6463-9354, e-mail: madina_abaeva@mail.ru

Информация об авторах
Иванова Алина Евгеньевна, кандидат наук об образовании, старший научный сотрудник Центра психометрики и измерений в образовании Института образования, ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0003-3340-7651, e-mail: aeivanova@hse.ru

Антипкина Инна Вениаминовна, кандидат наук об образовании, научный сотрудник Центра психометрики и измерений в образовании Института образования, ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0003-4865-3433, e-mail iantipkina@hse.ru

Абишева Сауле Джунусовна, доктор филологических наук, заведующая кафедрой русского языка и литературы, Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан, ORCID: https://orcid.org/0000-0002-4497-0805, e-mail: s.abisheva@mail.ru

Абаева Мадина Кабуловна, кандидат филологических наук, доцент кафедры русского языка и литературы, Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан, ORCID: https://orcid.org/0000-0001-6463-9354, e-mail: madina_abaeva@mail.ru

Получена 11.04.2022 Received 11.04.2022
Принята в печать 30.08.2022 Accepted 30.08.2022