

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 В или С (в которых размер эффекта превышал бы показатель 0,13 в классификации Зумбо-Томас). Результаты представляют интерес для междунациональных 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>

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: An exploratory 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.).

Введение

Распространенность явления билингвизма в системе образования Казахстана порождает целый ряд вопросов. Наше исследование посвящено инструменту оценки навыков чтения первоклассников на входе в школу и его функционированию в двух культурах: русскоязычных школах города Алматы (Казахстан) и школах города Новосибирска (Россия).

Предшествующие крупномасштабные исследования навыков чтения казахстанских школьников проводились в рамках международных сравнительных исследований. Казахстан впервые участвовал в Международном исследовании качества чтения и понимания текста (Progress in International Reading Literacy Study in Primary Schools, PIRLS) в 2016 году, причем оценивание проводилось

на русском и казахском языках на выборке 2983 учащихся с казахским языком обучения и 1942 — с русским. PIRLS — это независимое международное исследование читательской грамотности учащихся 4-х классов, которое проводит Международная ассоциация по оценке образовательных достижений IEA (International Association for the Evaluation of Educational Achievement) [17]. По результатам PIRLS-2016 четвероклассники Казахстана заняли 27 место в рейтинге 50 стран (для сравнения, российские школьники заняли первое место в рейтинге PIRLS-2016), девочки показывали более высокие результаты, чем мальчики (как и в 48 из 50 стран-участниц), дети лучше работали с информационными, чем с литературными текстами [4]. На основании результатов исследования PIRLS-2016 был сделан вывод о том, что «казахстанские дети получают недостаточный опыт чтения литературных текстов» [3]. В Международной программе по оценке образовательных достижений учащихся (Programme for International Student Assessment, PISA) 2018 года учащиеся Казахстана набрали в среднем 387 баллов по читательской грамотности, что является ниже среднего балла — 487 по странам ОЭСР [23].

Международные исследования навыков читательской грамотности PIRLS и PISA подняли в Казахстане интерес к вопросам развития навыков чтения у детей [22].

В рамках государственной программы полиязычия русский язык так же, как и казахский, и английский, является обязательным в школьных программах на весь период обучения (1—11 классы) вне зависимости от языка обучения. Из 5807 школ республики 1885 представляют собой школы с русским языком обучения. 2147 из названного общего количества школ относятся к школам со смешанными языками обучения. Для получения объективных данных и возможности сопоставления школ с двумя языковыми базами было принято решение проводить мониторинг формирования читательских навыков в два этапа.

- 1 этап: исследовать с помощью инструмента «Старт» читательские навыки у учащихся 1 класса школ с русским языком обучения г. Алматы, поскольку в настоящее время в

литературе есть информация только об исследованиях четвероклассников (PIRLS) и пятнадцатилетних учащихся (PISA) Казахстана.

- 2 этап: разработать казахскую версию инструмента исследования «СТАРТ/БАСТАУ» и провести тестирование по сформированности навыков чтения у учащихся начальных школ с казахским языком обучения г. Алматы.

В данной статье представлены результаты первого этапа исследования. Для оценивания первоклассников мы использовали шкалу чтения из инструмента «Старт» [1], который был создан на базе локализованной русскоязычной версии британского инструмента iPIPS [25], также используемого, например, в Австралии, Бразилии, Германии, ЮАР [6; 19]. Однако для использования русскоязычного инструмента в школах с русским языком обучения Казахстана необходимо доказать, что в условиях билингвальной культуры Казахстана все задания русскоязычной шкалы функционируют так же, как в русскоязычной среде российских школ.

Обзор литературы: мировой опыт исследований сопоставимости инструментов оценивания

Культурой называют «разделяемые членами группы выученные смыслы и паттерны поведения, которые передаются через социальную активность с целью социального приспособления, роста и развития» [17]. В тестах способностей от эффектов культуры не свободен ни один аспект оценивания: содержание теста, стимульные материалы, построение фраз, содержание инструкций, поведение участников во время тестирования и экспертов при назначении баллов (если это предусмотрено). Таким образом, любой тест содержит в себе культурную специфику. Выяснение того, является ли эта специфика препятствием для справедливого оценивания людей из другой культурной группы, — задача исследований эквивалентности.

Ряд примеров показывает наличие проблем с сопоставимостью результатов для отдельных стран и конструкторов в крупнейших международных исследованиях [9; 10; 15; 21]. Однако даже если в разных странах

говорят на одном языке и тест, используемый в этих странах, не нуждается в переводе, это не устраняет риски культурных различий в результатах оценивания [24]. Например, англоязычная версия инструмента iPIPS адаптировалась для использования в Австралии, несмотря на общий английский язык в этих двух странах, поскольку анализ функционирования заданий по чтению для учащихся австралийских школ по сравнению с аборигенным населением Австралии [27] показал различающееся функционирование отдельных заданий не в пользу учащихся из группы аборигенов Австралии. В рамках того же международного проекта iPIPS (лежащего в основе инструмента «Старт») проводились исследования сопоставимости версий инструмента на русском и английском языках для математической части теста [15], в котором были выявлены различия в функционировании нескольких заданий, из-за чего эти задания были исключены из вторичных анализов.

Кросс-культурная эквивалентность изменений исследуется на трех уровнях [26]:

- уровне конструкта (что измеряется),
- уровне инструмента измерений (чем измеряется),
- уровне шкалы измерений (в каких единицах измеряется).

В нашем исследовании для двух культур — России и Казахстана — использовались один и тот же русскоязычный инструмент и одинаковые процедуры психометрического анализа, поэтому уровни конструкта и шкалы идентичны; это позволяет сосредоточить анализ сопоставимости на уровне инструмента.

Три источника возможных межкультурных искажений (bias) в результатах применения инструмента [14] это:

1) Различия в выборках оценивания. Например, выборки могут быть несопоставимыми, если в одной стране для учащихся с ограниченными возможностями здоровья предусмотрены отдельные школы, в то время как в другой стране — применяется инклюзивный подход.

2) Различия в инструментах. Например, в формулировках заданий могут быть использованы реалии, хорошо знакомые в одной культуре, но экзотические для другой.

3) Различия в процедурах сбора данных. Классическим примером таких процедурных нарушений было тестирование детей из Нигерии матрицами Ровена «на порогах домов, в прихожих, под деревьями» с помощью необученных сотрудников, что сильно отличалось от условий, в которых оценивали детей в европейских странах [11].

В нашем исследовании условия сбора данных были стандартизированы единой компьютерной средой, озвучкой инструмента профессиональным диктором (интервьюерам не требовалось зачитывать инструкции) и одинаковыми процедурами обучения интервьюеров. Выборки детей были сопоставимы по возрасту и статусу: дети 6—7 лет, впервые пришедшие в школу. Таким образом, цель исследования — сделать обоснованный вывод о пригодности теста для использования в двух культурах (Россия и Казахстан) ради возможности напрямую сопоставлять полученные в обеих странах результаты. Задача исследования: провести статистический анализ инвариантности результатов, включая DIF (differential item functioning, различающееся функционирование заданий).

Материалы и методы

Инструмент. «Старт» представляет собой компьютеризированный инструмент оценки навыков чтения с полуадаптивным алгоритмом предъявления заданий [1], поэтому, если ребенок совершает определенное количество ошибок, оценивание прекращается, чтобы не демотивировать ребенка. Задания озвучены профессиональным диктором, что значительно стандартизирует условия оценивания. Проведение оценивания предполагает индивидуальные сессии ребенка со специально обученным интервьюером. Шкала чтения включает в себя 35 заданий и охватывает следующие области:

* распознавание букв русского алфавита — на экране по одной предъявляются буквы, голос диктора просит ребенка назвать буквы. Правильным ответом считается название буквы в официальном и разговорном варианте (например, «эм» и «мэ» будут правильными ответами для «М») или обозначение звука («м-

м»). Неправильным ответом считается название предмета на эту букву (например, если на «М» ребенок говорит: «Мяч»).

- чтение отдельных слов (чтение-декодирование),
- чтение короткой истории («механическое» чтение),
- смысловое чтение. В этой части инструмента ребенок читает текст, в отдельных местах которого ему предлагают выбрать одно из трех слов, наиболее подходящее по контексту.

Инструмент «Старт» стандартизирован с использованием дихотомической модели Раша [1; 2].

Выборка. Оценивание навыков чтения первоклассников проводилось инструментом «Старт» в октябре 2019 года, когда дети только пришли в школы. Выборка русскоязычных школ города Алматы составила 1102 ребенка. Выборка не является репрезентативной. Выборка первоклассников в Новосибирске составила 2247 учеников. Новосибирск был выбран как регион с сопоставимым по отношению к Алматы географическим положением, индустриальным развитием и численностью населения. Выборка учеников в Новосибирске была рандомизирована и стратифицирована по типу школы (обычная СОШ или школа со статусом «гимназия», «лицей», «школа с углубленным изучением отдельных предметов») и по району города.

Аналитический подход. Исследование сопоставимости функционирования инструмента в двух культурных средах (России и Казахстане) проводилось в два этапа.

На первом этапе проводился психометрический анализ инструмента на выборке русскоязычных школ Алматы и Новосибирска по отдельности. Анализ проводился с использованием дихотомической модели Раша. Базовая дихотомическая модель Раша применяется в разработке оценочных инструментов, состоящих из заданий дихотомического типа, где есть один правильный и один неправильный ответ. Для оценки надежности шкал используются показатели классической надежности альфа Кронбаха и Раш-надежности (Separation Reliability).

На втором этапе проводился анализ различного функционирования заданий. Различное функционирование задания (differential item functioning, DIF) фиксируется в том случае, когда испытуемые с одинаковым уровнем подготовленности имеют различные шансы выполнить задание верно [28; 29]. Иначе говоря, наличие DIF подразумевает, что задание функционирует неодинаково в двух исследуемых группах в пользу одной из этих групп несмотря на то, что эти учащиеся имеют один и тот же итоговый балл. Принадлежность детей к разным культурным бэкграундам может быть источником различающегося функционирования заданий, даже если разработка инструмента следовала всем необходимым процедурам в соответствии с международными стандартами [13; 20; 21].

Анализ выявленных с помощью DIF-анализа различий способствует корректной интерпретации результатов оценивания и помогает доработке инструмента для того, чтобы он стал более справедливым ко всем группам учащихся.

Результаты

Психометрический анализ данных обеих версий был проведен сначала отдельно для каждой национальной выборки, а затем на объединенной выборке с целью выяснить, работают ли задания инструмента одинаковым образом для учащихся из России и Казахстана при учете их уровня подготовленности по чтению.

В табл. 1 приведены характеристики заданий для каждой из двух выборок по отдельности. Характеристики включают в себя трудность заданий (1, 2 колонки), ошибку оценки трудности (3, 4 колонки), статистики согласия (5, 6 колонки), корреляцию ответа на задание с уровнем подготовленности испытуемого (дискриминативность, 7, 8 колонки), а также название заданий. Статистики согласия представляют собой среднеквадратичные отклонения эмпирических значений от ожидаемых моделью значений для каждого задания, в данной таблице мы используем одну — взвешенную статистику согласия (Infit MNSQ). Как видно из табл. 1, средние значения статистик согласия лежат в пределах рекомендуемых специалистами значений [0,6; 1,4] для всех

заданий, кроме двух первых (на знание букв) для новосибирской выборки. Ошибка измерения несколько выше для заданий для выборки детей из Казахстана. Корреляция ответа на задание с уровнем подготовленности высокая и положительная.

Таблица 1

**Характеристики заданий для выборки учеников
 из г. Новосибирска, Россия и г. Алматы, Казахстан**

Трудность задания		Взвешенная статистика согласия		Наименование задания
Рос.	Каз.	Рос.	Каз.	
-1,37	-1,92	1,53	1,25	Буква «К»
-1,37	-1,65	1,45	1,36	Буква «Л»
-1,02	-1,35	1,13	1,22	Буква «е»
-1,15	-1,43	1,24	1,23	Буква «Ж»
-1,36	-,95	1,15	1,32	Буква «З»
-,33	-,89	1,36	1,26	Буква «Ш»
-,37	-,31	1,11	1,16	Буква «Ц»
-,71	-,60	1,06	1,21	Буква «ю»
-,05	-,09	1,25	1,24	Буква «ь»
-2,44	-2,24	,92	1,05	Слово «мяч»
-3,15	-3,11	1,06	1,27	Слово «утка»
-2,82	-3,38	,90	,95	Слово «щенок»
-2,66	-2,75	,80	,93	Слово «рука»
-2,99	-3,26	,81	,81	Слово «дом»
-1,20	-,63	1,00	1,17	Слово «конь»
-1,10	-,77	,95	1,00	Слово «корабль»
-1,82	-1,65	,81	,86	Слово «кот»
-,99	-,81	,93	,94	Слово «кролик»
,22	-,02	,76	,54	Мех. чтение 1
,32	,05	,76	,52	Мех. чтение 2
,47	,13	,77	,52	Мех. чтение 3
1,22	1,57	,87	,96	Смысловая ловушка 1
1,06	1,41	,74	,88	Смысловая ловушка 2
1,20	1,35	,71	,82	Смысловая ловушка 3
1,24	1,39	,76	,87	Смысловая ловушка 4
1,75	2,16	,90	1,01	Смысловая ловушка 5
1,76	2,00	,92	1,00	Смысловая ловушка 6
,94	,74	,62	,67	Смысловая ловушка 7
2,29	2,08	1,09	1,03	Смысловая ловушка 8
1,95	2,01	1,00	,97	Смысловая ловушка 9
2,50	2,44	1,03	1,00	Смысловая ловушка 10
3,40	3,36	1,16	1,15	Смысловая ловушка 11
1,73	1,59	,91	,93	Смысловая ловушка 12
2,95	3,33	1,24	1,18	Смысловая ловушка 13
1,91	2,16	,93	,94	Смысловая ловушка 14

Преимуществом Раш-моделей является возможность визуализировать характеристики заданий и респондентов с помощью так называемых «карт переменных». На рис. 1 и 2 вертикальная линия представляет собой континуум навыков чтения (и, соответственно,

трудностей заданий по чтению), измеренных в логитах. Самые легкие задания и наименее подготовленные респонденты расположены внизу континуума, а в верхней части находятся самые трудные задания и самые подготовленные респонденты.

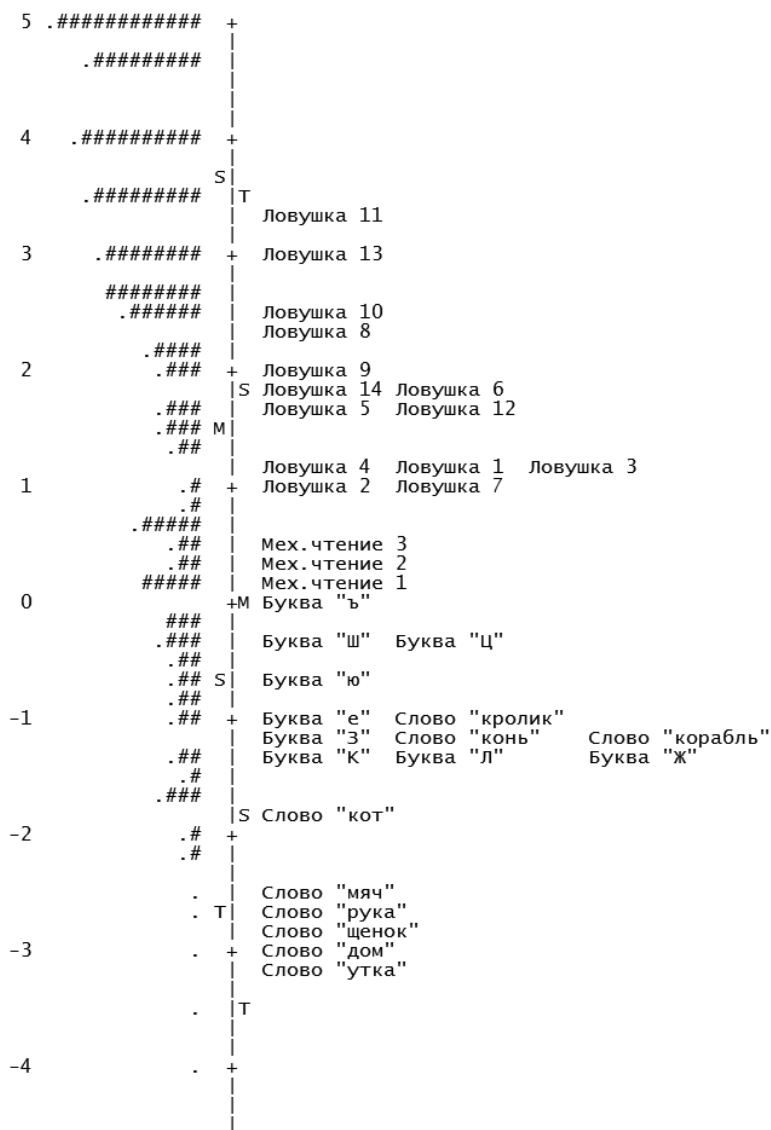


Рис. 1. Карта переменных. Россия

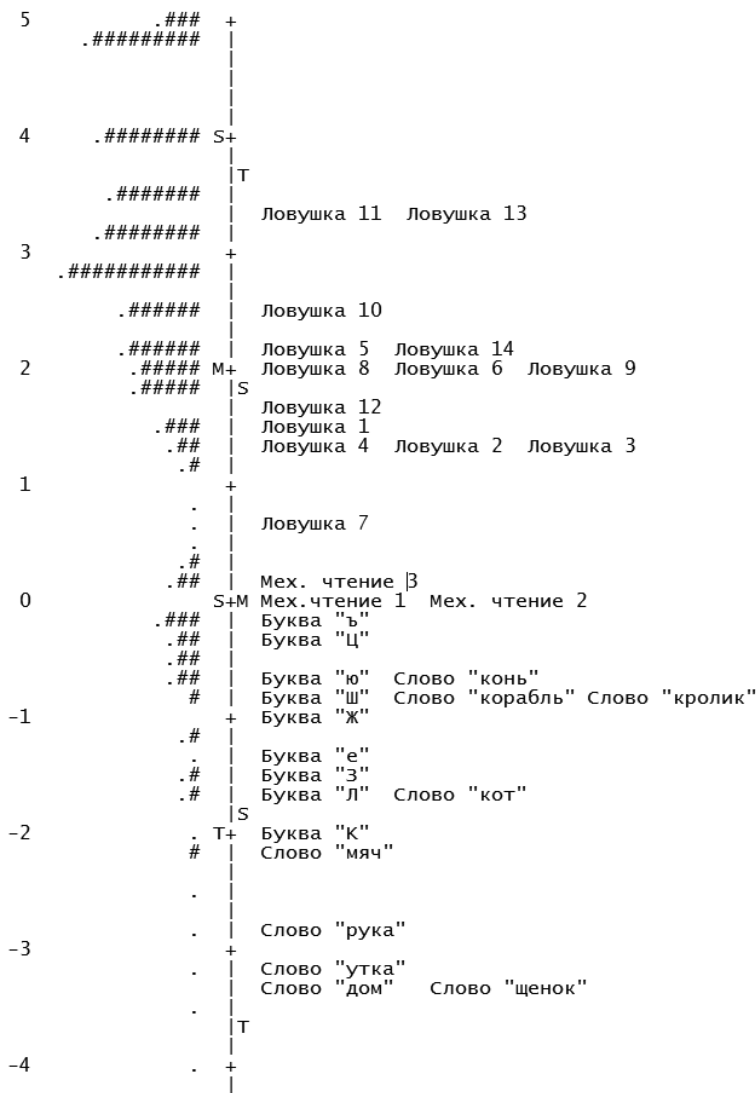


Рис. 2. Карта переменных. Казахстан

Можно заметить, что для обеих выборок характерен сдвиг в сторону высоких баллов для большей части детей, т.е. для большинства учащихся тест оказался достаточно легким. Также мы можем отметить, что в целом иерархия заданий по трудности для двух выборок практически идентична. В нижней части обеих карт расположены задания

на чтение отдельных слов — это короткие частотные слова, графическую оболочку которых многие дети научаются распознавать (запоминают как картинку) еще до того, как полностью изучают буквы и осваивают принцип чтений как декодирования. В средней части карт расположена самая большая группа заданий, включающих в себя распознавание

букв и чтение небольшой истории, которое соответствует навыку механического чтения (чтения-декодирования, по другой терминологической традиции). В верхней части карты расположились наиболее трудные задания на чтение с пониманием.

В табл. 2 представлены общие психометрические показатели надежности шкал чтения для эмпирических данных, полученных на выборках учеников России и Казахстана.

Для обеих версий инструмента характерны высокие показатели надежности — классической альфы Кронбаха и Раш-надежности (Person Reliability), а также достаточно высокий общий уровень чувствительности шкалы (Person Separation), позволяющий различать не менее трех групп испытуемых, выделенных на основании уровня навыка чтения.

Хорошие психометрические качества шкалы чтения отдельно на выборках учащихся из России и Казахстана позволяют построить единую шкалу результатов тестирования первоклассников из двух стран и провести анализ различий в функционировании заданий (DIF-анализ).

В рамках данного исследования DIF-анализ был проведен с помощью метода логистической регрессии. Метод логистической регрессии [7; 29] позволяет обнаружить как однородный (когда статистическая связь между ответом на задание и группирующей переменной постоянна для всех уровней соответствующей переменной), так и неоднородный DIF (когда статистическая связь между ответом на задание и группирующей переменной меняется для разных уровней соответствующей переменной). Метод заключается в статистическом моделировании вероятности верно ответить на задание в зависимости от последовательно вводимых переменных: 1) группирующей переменной «страна» (в нашем случае), 2) суммы

баллов по тесту и 3) взаимодействия между первой и второй переменными. В данной работе для анализа DIF было использовано программное обеспечение R, статистический пакет DIFR [16]. Статистическая значимость параметров моделей оценивалась с помощью LR-теста (теста отношения правдоподобия (Likelihood Ratio Test)).

В методе логистической регрессии задание идентифицировалось как демонстрирующее DIF, когда последовательное добавление переменной страны и переменной взаимодействия дает значимое улучшение модели по сравнению с моделью, которая включает только переменную с суммой баллов [28].

Важно отметить, что различия для групп испытуемых могут быть статистически значимыми, но слишком маленькими, чтобы иметь какое-либо влияние на результаты оценивания. Поэтому исследователям рекомендуется использовать комбинацию показателей — статистическую значимость индикаторов DIF и практическую значимость (или размер эффекта), чтобы принять взвешенное решение о том, что делать, если выявлено различное функционирование заданий.

В данной работе мы используем комбинацию показателей значимости статистики теста отношения правдоподобия (Likelihood Ratio Test) и размера эффекта DIF, представленной в двух вариантах — Зумбо-Томас, а также Джодоин-Гирл [16]. Чтобы задание было классифицировано как демонстрирующее DIF, критерий LR-теста должен иметь значимость, меньшую или равную 0,01, а также размер эффекта должен быть достаточно большим по крайней мере по одному из указанных выше критериев. Согласно критерию Зумбо-Томас можно классифицировать DIF следующим образом: пренебрежимо малый DIF типа A (изменение значений R-квадрат двух вложенных

Таблица 2

Результаты анализа надежности эмпирических данных

Показатели	Надежность альфа Кронбаха	Раш-надежность (Person Reliability)	Индекс чувствительности теста (Person Separation)
Россия	0,96	0,91	3,17
Казахстан	0,94	0,89	2,89

моделей ниже 0,13), умеренный DIF типа В (изменение значений R-квадрат двух вложенных моделей от 0,13 до 0,26) и большой DIF типа С (изменение значений R-квадрат двух вложенных моделей выше 0,26). Согласно более строгому критерию Джодоин и Гирл DIF можно классифицировать DIF следующим образом: пренебрежимо малый DIF типа А (изменение значений R-квадрат двух вложенных моделей ниже 0,035), умеренный DIF типа В (изменение значений R-квадрат двух вложенных моделей от 0,035 до 0,07) и большой DIF типа С (изменение значений R-квадрат двух вложенных моделей выше 0,07). В табл. 3 приведены результаты DIF-анализа с помощью метода логистической регрессии.

Таким образом, проведенный анализ показал, что, несмотря на статистическую значимость LR-теста для ряда заданий по чтению,

размер эффекта DIF настолько мал, что им можно пренебречь для практического использования результатов. Иными словами, для учащихся из России и Казахстана все задания теста по чтению работают скорее одинаково.

Основные итоги

Результаты любого исследования, включающего кросс-культурную оценку, можно корректно сравнивать, если мы готовы предположить, что аналогичные баллы участников тестирования в двух странах означают аналогичный уровень владения проверяемым навыком.

Данное исследование было посвящено возможности напрямую сопоставлять результаты теста чтения, полученные с помощью одного и того же теста в двух культурах: русскоязычных школах города Алматы (Казахстан) и школах города Новосибирска (Россия). Про-

Таблица 3

Анализ различного функционирования заданий, метод логистической регрессии

Задание	LR-тест	P-value	R ²	Размер DIF (Зумбо-Томас)	Размер DIF (Джодоин-Гирл)
Буква «К»	24,36	0,00 ***	0,01	А	А
Буква «Л»	7,91	0,02 *	0,00	А	А
Буква «е»	7,14	0,03 *	0,00	А	А
Буква «З»	6,78	0,03 *	0,00	А	А
Буква «Ш»	22,19	0,00***	0,00	А	А
Слово «мяч»	14,57	0,01 ***	0,00	А	А
Слово «утка»	11,75	0,00 **	0,01	А	А
Слово «щенок»	7,71	0,02 *	0,00	А	А
Слово «конь»	18,64	0,00 ***	0,00	А	А
Слово «корабль»	6,99	0,03*	0,00	А	А
Мех. чтение 1	24,56	0,00 ***	0,00	А	А
Мех. чтение 2	29,62	0,00***	0,00	А	А
Мех. чтение 3	35,69	0,00 ***	0,00	А	А
Смысловая ловушка 1	17,62	0,00 ***	0,00	А	А
Смысловая ловушка 2	26,22	0,00 ***	0,00	А	А
Смысловая ловушка 3	9,85	0,01 **	0,00	А	А
Смысловая ловушка 4	8,82	0,01 *	0,00	А	А
Смысловая ловушка 5	24,47	0,00 ***	0,00	А	А
Смысловая ловушка 6	9,26	0,01 **	0,00	А	А
Смысловая ловушка 13	18,33	0,00 ***	0,00	А	А
Смысловая ловушка 14	10,32	0,01 **	0,00	А	А

Примечание. р: 0 (****), 0.001 (***), 0.01 (**).

веденный нами психометрический анализ показал, что инструмент оценки ранних навыков чтения на входе в школе хорошо функционирует не только в среде русскоязычных детей в России, но и в среде русскоязычных детей в Казахстане. Шкала демонстрирует аналогичные показатели для выборок детей в двух странах, включая согласие данных с моделью измерения, показатели надежности, распределение трудностей заданий и уровней подготовленности учащихся.

Мы показали, что эффект культуры не вмешивается в трудности заданий теста чтения для первоклассников в двух странах. Это открывает перспективы международных исследований в Казахстане с помощью русскоязычных инструментов, уже стандартизированных на российских выборках.

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. Амангазы, 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., Сабырлы 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. Kazakhstan [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

При этом важно обратить внимание, что российские и казахстанские дети одинаково успешно справились с большинством заданий теста на чтение. Возникает вопрос: если на старте школы дети из России и Казахстана обладают сопоставимыми навыками чтения на русском языке, почему к концу четвертого класса возникает разрыв в навыках, фиксируемый в PIRLS? Лонгитюдное исследование в начальных школах Казахстана с русским языком обучения с использованием инструментов с доказанными психометрическими свойствами может помочь ответить на этот вопрос.

В будущем планируется адаптация инструмента «Старт» на казахский язык с целью проведения сопоставительных исследований образовательных достижений школьников из школ с русским и казахским языками обучения.

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 Vijver 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.rrediris.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., Ericikan 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).
- 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 mezhranovogo 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 razvitiu 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. Amanrazay, G. Nogaibaeva, 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.* «Rezultaty 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 pre-school 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/educ.2019.00120/full> (Accessed 02.11.2021). DOI:10.3389/educ.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 Vijver 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).

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

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

Получена 11.04.2022

Received 11.04.2022

Принята в печать 30.08.2022

Accepted 30.08.2022