МРНТИ 14.35.09

Ж.Ж. Бейсенова[1], Л.Е. Шинетова[2], Г.К. Жабаева[3]
[1,2,3]РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Республика Казахстан

[1]ORCID 0000-0002-7347-1538, [2]ORCID 0000-0003-4280-7999, [3]ORCID 0009-0006-1003-6316

ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ ТЕСТОВЫХ ЗАДАНИЙ ПО ОЦЕНКЕ ЗНАНИЙ ПЕДАГОГОВ В РЕСПУБЛИКЕ КАЗАХСТАН (ПО РЕЗУЛЬТАТАМ АПРОБАЦИИ)

В статье представлены детальные результаты психометрического анализа тестовых заданий по математике, которые использовались в рамках апробации процедуры оценки знаний педагогов (ОЗП) в Казахстане. Исследование основано на сравнительном анализе трех различных вариантов теста, подготовленных для аттестации. Для оценки качества заданий и выявления их сильных и слабых сторон были задействованы методы как классической теории тестов (КТТ), так и современной теории тестов (IRT), в частности, модель Раша. Целью работы являлась не только оценка существующих оценочных инструментов, но и разработка практических рекомендаций по их усовершенствованию. Анализ апробации показал, что, несмотря на общую высокую надежность тестов (коэффициент альфа-Кронбаха превышает 0,84) и их хорошую способность дифференцировать испытуемых по уровню подготовленности, в каждом варианте были обнаружены отдельные проблемные задания. Было выявлено наличие заданий с низкой или даже отрицательной дискриминативностью, а также дистракторов, которые не работали должным образом. Все замечания, выявленные в ходе апробации, были учтены и устранены в последующей работе по корректировке банка заданий. Это позволило повысить точность и валидность оценочных инструментов в целом.

Ключевые слова: оценка знаний педагогов, психометрический анализ, апробация, надежность, валидность, классическая теория тестов, модель Раша.

Введение

В современной образовательной системе Казахстана процедура оценки знаний педагогов (ОЗП) представляет собой важный инструмент, который направлен на повышение качества обучения и профессионального статуса учителей. В соответствии с приказом Министра образования и науки Республики Казахстан от 27 января 2016 года №83 «Об утверждении Правил и условий проведения аттестации педагогов», ОЗП проводится для определения уровня профессиональной компетентности педагога с помощью тестов, разработанных уполномоченным органом в области образования [1, 2]. Этот процесс распространяется на педагогов дошкольных организаций, предшкольных классов, учителей начального, среднего и общего среднего образования, а также на работников дополнительного, специального и технического образования. ОЗП также затрагивает руководителей организаций образования, методистов, лиц без стажа и тех, кто желает возобновить педагогическую деятельность [3].

В отличие от прошлых форматов, ОЗП больше сфокусирована на практических навыках и методических компетенциях [4]. Такой подход позволяет получать объективные данные об уровне подготовки учителей, выявлять системные пробелы и находить лучшие способы для развития всей системы образования [5]. Ключевым компонентом ОЗП является тестирование, которое должно обладать высокой надежностью и валидностью. Успешно пройденный тест – это не просто подтверждение квалификации. Это возможность для карьерного роста и увеличения заработной платы, что, несомненно, мотивирует учителей постоянно совершенствоваться. Использование электронного формата делает оценку еще более объективной и прозрачной [1].

В этом исследовании мы провели всесторонний анализ тестовых заданий по математике, чтобы оценить их качество и соответствие современным требованиям. Результаты, которыми мы делимся, были получены в ходе апробации этого инструмента.

Материалы и методы исследования

Для анализа были выбраны три варианта теста по математике, использованные при апробации аттестации педагогов: 369168, 369169 и 369694. Выборка составила 190, 185 и 182 участника соответственно. Психометрический анализ проводился с использованием двух ключевых теоретических подходов. В рамках классической теории тестов (КТТ) анализ включает расчет показателей надежности (коэффициенты *KR-20* и *альфа-Кронбаха*), средней трудности и распределения заданий, а также дискриминативности (*pBis*) и эффективности дистракторов [6]. В рамках современной теории тестов (IRT) в рамках модели Раша оценивались надежность, разделяющая способность (*Person Separation*) и соответствие модели (*Infit MNSQ и Outfit MNSQ*), где допустимый диапазон значений установлен в пределах 0,7-1,3. Для расчетов использовались специализированные программы, такие как TAP, Jamovi и Winsteps [7].

Результаты и обсуждение

Результаты анализа по КТТ

Анализ данных был проведен нами с помощью специальных программ TAP и Jamovi. В ходе анализа было выяснено следующее. Все три варианта теста продемонстрировали высокий уровень надежности, что подтверждается значениями коэффициентов KR-20 и альфа-Кронбаха, превышающими 0,84. Это отличный показатель, который говорит о том, что результаты тестов стабильны и внутренне согласованы [8]. Проще говоря, если бы участники прошли этот тест еще раз, их результаты, скорее всего, были бы очень похожими (таблица 1).

Таблица 1 - Основные результаты анализа (КТТ)

Показатель	Вариант 369168	Вариант 369169	Вариант 369694
Число участников	190	185	182
Средний балл	27,53	28,62	31,1
Стандартное отклонение	7,70	8,38	8,64
Минимум / Максимум	9 / 47	9 / 46	10/ 47
Асимметрия	0,033	-0,095	-0,413
Эксцесс	-0,557	-0,677	-0,764
Надежность (KR-20 и альфа-Кронбаха)	0,84	0,87	0,89
Средняя трудность	0,551	0,572	0,622
Число заданий трудности < 0,2	3	0	0
Число заданий трудности > 0,8	4	4	11
Проблемные задания с дискриминативностью < 0	3 (№2, 22, 34)	1 (№44)	1 (№34)
Задания с дискриминативностью < 0,1	5 (№2, 9, 17, 22, 34)	3 (№ 1, 44, 45)	2 (№34, 49)

Таблица представляет собой сводку основных результатов психометрического анализа, выполненного в рамках КТТ с использованием программного обеспечения TAP.

Распределение баллов по всем вариантам в целом соответствует нормальному, с приемлемыми показателями асимметрии и эксцесса. Вариант 369694 показал небольшое смещение в сторону более высоких баллов (отрицательная асимметрия), что может быть обусловлено его чуть меньшей сложностью (рисунок 1). В целом, такое распределение говорит о том, что тесты успешно дифференцируют испытуемых по уровню подготовки.

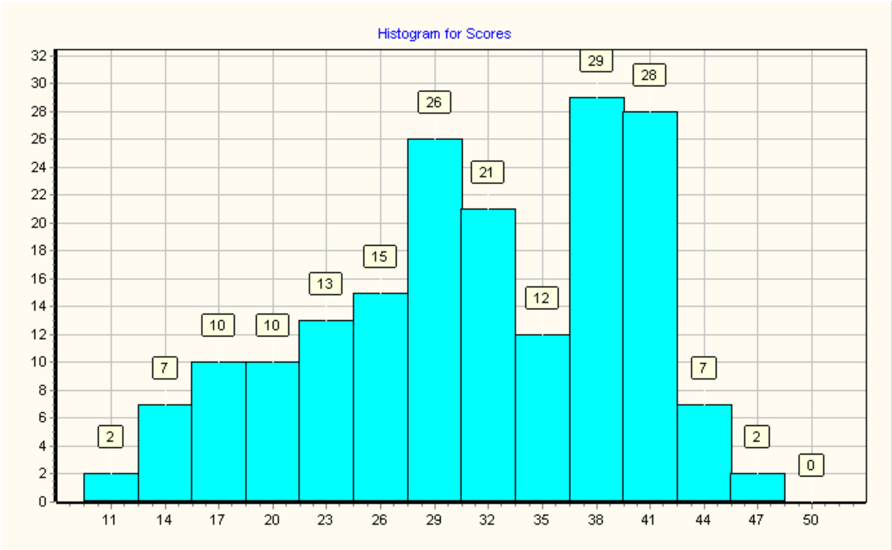

Рисунок 1 - Гистограмма распределения баллов по варианту 369694 (данные TAP)

Несмотря на высокие общие показатели, углубленный анализ выявил отдельные задания, требующие дополнительного внимания. Так, было обнаружено, что некоторые вопросы имеют низкую или даже отрицательную дискриминативность (таблица 1), что может свидетельствовать о слабой корреляции с общим уровнем подготовленности испытуемых. Иными словами, такие задания не всегда эффективно дифференцируют сильных и слабых участников, что может быть обусловлено их формулировкой или возможностью случайного угадывания. Эти особенности могут оказывать влияние на точность измерения и, как следствие, на валидность оценочного инструмента. Кроме того, анализ дистракторов показал, что функционирование некоторых из них не всегда соответствовало ожиданиям: отдельные из них были слишком очевидными, а другие, напротив, вводили в заблуждение даже высококвалифицированных испытуемых.

Дистракторный анализ проблемных заданий в КТТ

Дистракторный анализ является ключевым этапом в процессе психометрической оценки качества заданий. Он позволил нам выявить и идентифицировать проблемные задания, которые не соответствуют установленным критериям эффективности (таблицы 2) [9]. В отличие от традиционного анализа, фокусирующегося только на общем проценте правильных ответов, дистракторный анализ дает возможность детально изучить привлекательность каждого дистрактора. Это имеет огромное значение для повышения надежности и валидности теста [10].

Анализ проблемных заданий в КТТ начинается с изучения коэффициентов дискриминации и точечно-бисериальной корреляции (pBis) для каждого варианта ответа. Высокий положительный pBis для правильного ответа (например, задание №9, вариант А: pBis=0,10386) указывает на то, что этот вариант ответа чаще выбирают студенты с высокими общими баллами. В то же время, эффективные дистракторы должны иметь отрицательный pBis (например, задание №9, варианты C и D). Это свидетельствует о том, что они привлекают испытуемых с низкими результатами, что является желаемой характеристикой.

Проблемные задания, напротив, демонстрируют аномальные результаты. Например, в задании №22, правильный ответ C имеет низкий pBis (-0,0681), а его дискриминация (0,0126) практически равна нулю, что указывает на отсутствие связи с общим уровнем знаний испытуемых. Более того, дистрактор D в этом же задании имеет положительный pBis (0,0828), что свидетельствует о его привлекательности для более сильных студентов. Такие результаты требуют пересмотра или исключения задания из теста, так как оно не выполняет своей основной функции по дифференциации испытуемых.

Таблица 2 – Результаты дистракторного анализа тестовых заданий

Вариант 369168

Задание № 2

	correct	key	N	rspP	pBis	discrim	lower	mid66	upper
A		A	51	0,26842	0,0832	0,15441	0,2206	0,2273	0,375
B	*	B	35	0,18421	-0,025	0,0021	0,1765	0,197	0,179
C		C	8	0,04211	-0,1219	-0,0588	0,0588	0,0606	0
D		D	95	0,5	-0,1667	-0,0976	0,5441	0,5	0,446

Задание №9

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	33	0,1737	0,10386	0,1176	0,132	0,152	0,25
B		B	44	0,2316	0,00942	0,0798	0,206	0,212	0,2857
C		C	95	0,5	-0,1707	-0,0977	0,544	0,5	0,4464
D		D	18	0,0947	-0,1437	-0,0998	0,118	0,136	0,0179

Задание №17

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	53	0,27895	-0,2903	-0,2574	0,382	0,303	0,125
B	*	B	42	0,22105	0,033	0,062	0,206	0,197	0,268
C		C	47	0,24737	0,2017	0,2279	0,147	0,2424	0,375
D		D	47	0,24737	-0,1454	-0,0326	0,265	0,2424	0,232

Задание №22

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	16	0,0842	-0,1985	-0,1082	0,1618	0,0303	0,0536
B		B	4	0,0211	-0,1522	-0,0263	0,0441	0	0,0179
C	*	C	109	0,5737	-0,0681	0,0126	0,5588	0,5909	0,5714
D		D	61	0,3211	0,0828	0,1218	0,2353	0,3788	0,3571

Задание №34

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	32	0,16842	-0,0413	0,0494	0,147	0,1667	0,196
B		B	27	0,14211	-0,0908	-0,0546	0,162	0,1515	0,107
C	*	C	45	0,23684	-0,0278	0,0861	0,235	0,1667	0,321
D		D	85	0,44737	-0,0661	-0,0809	0,456	0,5	0,375

Вариант 369169

Задание №1

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	34	0,1838	-0,3332	-0,2051	0,286	0,1833	0,0806
B		B	16	0,0865	-0,07	-0,0305	0,111	0,0667	0,0806
C	*	C	45	0,2432	0,0846	0,0684	0,238	0,1833	0,3065
D		D	90	0,4865	0,0712	0,1672	0,365	0,5667	0,5323

Задание №44

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	61	0,33	-0,045	0,0371	0,302	0,35	0,3387
B		B	73	0,395	0,1419	0,2783	0,254	0,4	0,5323
C		C	28	0,151	-0,2305	-0,1895	0,254	0,133	0,0645
D		D	23	0,124	-0,1429	-0,126	0,19	0,117	0,0645

Задание №45

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	33	0,178	-0,2907	-0,2048	0,302	0,1333	0,0968
B		B	19	0,103	-0,112	-0,0622	0,143	0,0833	0,0806
C	*	C	69	0,373	0,0333	0,1019	0,317	0,3833	0,4194
D		D	64	0,346	0,1002	0,1651	0,238	0,4	0,4032

Вариант 369694

Задание №34

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	17	0,0934	-0,32895	-0,19201	0,209	0,0357	0,0169
B		B	43	0,2363	0,20532	0,29825	0,0746	0,2857	0,3729
C		C	17	0,0934	-0,1850	-0,1004	0,1343	0,1071	0,0339
D	*	D	105	0,5769	-0,0082	-0,0058	0,5821	0,5714	0,5763

Задание №49

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	95	0,522	0,0841	0,258	0,403	0,5179	0,661
B		B	56	0,3077	-0,0408	-0,0402	0,3284	0,3036	0,2881
C		C	12	0,0659	-0,0217	-0,0238	0,0746	0,0714	0,0508
D		D	19	0,1044	-0,2883	-0,194	0,194	0,1071	0

Таблица представляет собой результаты анализа ответов на задания, выполненного с помощью программного обеспечения Jamovi.

Таким образом, в ходе дистракторного анализа были выявлены задания, которые требуют пересмотра для повышения валидности и надежности теста (таблица 2) [12]. На основе полученных данных, были выделены три основные проблемы, влияющие на качество теста.

Низкая и отрицательная дискриминативность ключей. Основной задачей тестового вопроса является дифференциация испытуемых по уровню их знаний. В таких заданиях показатель pBis, либо близок к нулю, либо имеет отрицательное значение. Это означает, что правильный ответ на такой вопрос не позволяет надежно отличить сильных участников от слабых, что ставит под сомнение валидность задания.

Чрезмерно легкие задания. Ряд вопросов показал аномально высокую долю правильных ответов (показатель трудности p>0,9). Это свидетельствует о том, что задания не выполняют своей диагностической функции, так как их дистракторы оказались нерабочими – их практически никто не выбрал. Такие вопросы не вносят существенного вклада в общую оценку.

Ключ выбран реже, чем дистрактор. В ряде случаев правильный ответ оказался менее популярным, чем один из неправильных. Это серьезный сигнал о том, что в задании может быть либо ошибка, либо его формулировка слишком сложна и сбивает с толку даже хорошо подготовленных испытуемых.

Для устранения выявленных проблем и оптимизации качества оценочного инструмента нами были предложены следующие рекомендации:

Обязательная переэкспертиза заданий с отрицательной дискриминативностью (pBis < 0). Эти вопросы необходимо подвергнуть глубокому содержательному анализу, чтобы выявить причины их низкой эффективности;

Пересмотр или удаление заданий, где ключ не позволяет эффективно дифференцировать испытуемых по группам или выбирается реже, чем дистрактор;

Повышение качества дистракторов. Для улучшения диагностической ценности вопросов необходимо усилить правдоподобность неправильных вариантов ответа. Это позволит уменьшить вероятность случайного угадывания и повысить точность измерения знаний.

Для более глубокой оценки качества заданий и выявления их соответствия единой шкале способностей после КТТ проведен анализ в IRT (модели Раша) с использованием программы Winsteps [13]. Основные результаты анализа IRT приведены в таблице 3.

Таблица 3 - Основные результаты анализа по IRT (модель Раша)

Показатель	Вариант 369168	Вариант 369169	Вариант 369694
Количество респондентов	190	185	182
Надежность (альфа-Кронбаха)	0,84	0,87	0,88
Надежность Person Reliability	0,83	0,86	0,86
Разделяющая способность Person Separation	2,20	2,45	2,50
Реальная стандартная ошибка (Real RMSE)	0,34	0,35	0,36
Средняя способность участников (Mean Measure Persons)	0,26	0,40	0,67
Стандартное отклонение SD Measure (Persons)	0,83	0,92	0,97
Число заданий и номера с Corr<0	1 (№2)	0	0
Число заданий и номера с Corr< 0,1	1 (№2)	1 (№44, 45)	1 (№34)
Число заданий и номера с Infit MNSQ <0,7	0	0	0
Число заданий и номера с Infit MNSQ > 1,3	2 (№2, 34)	1 (№44)	1 (№34)
Число заданий и номера Outfit MNSQ<0,7	0	2 (№25, 39)	2 (№32, 43)
Число заданий и номера Outfit MNSQ > 1,3	2 (№2, 34)	4 (№1, 38, 44, 45)	4 (№34, 49, 41, 33)

Таблица представляет собой сводку основных результатов психометрического анализа, выполненного в рамках IRT с использованием программного обеспечения Winsteps.

Полученные результаты демонстрируют высокие показатели надежности и разделяющей способности для всех трех вариантов тестов. Коэффициент альфа-Кронбаха для каждого варианта превышает 0,84, а показатель Person Separation стабильно находится выше 2,0. Это свидетельствует о высокой согласованности заданий и их способности эффективно дифференцировать испытуемых по уровню подготовленности [14].

Однако, несмотря на высокие интегральные показатели, детальный анализ выявил ряд заданий, не соответствующих требованиям модели Раша. В частности, были идентифицированы вопросы с низкими и даже отрицательными значениями корреляции (Corr), а также с нежелательными значениями Infit MNSQ и Outfit MNSQ, выходящими за допустимые пределы (0,7-1,3). Данные показатели свидетельствуют о несоответствии заданий модели и указывают на наличие нестандартных паттернов ответов, что может быть вызвано некорректной формулировкой вопроса, ошибкой в ключе или неоднозначностью содержания.

Наиболее проблемным является Вариант 369168, где задание №2 демонстрирует отрицательную корреляцию, что является критическим индикатором его неработоспособности. В целом, во всех трех вариантах выявлены вопросы с высокими значениями Outfit MNSQ > 1,3, что указывает на непредсказуемость ответов, вероятно, из-за случайного угадывания или невнимательности. Например, в варианте 369694 таких заданий было обнаружено четыре.

Таким образом, несмотря на общую высокую надежность инструмента, наличие отдельных заданий, не соответствующих модели, ставит под вопрос валидность и справедливость оценки, поскольку они искажают результаты и не позволяют получить точное измерение латентной переменной [15].

На основе проведенного анализа, для повышения качества оценочного инструмента были предложены следующие рекомендации.

Проведение экспертной верификации. Все задания, демонстрирующие значительные отклонения от модели Раша, должны быть подвергнуты тщательной экспертной проверке. Это касается вопросов с низкими или отрицательными корреляциями и нестандартными значениями Infit/Outfit MNSQ. По итогам проверки они могут быть переформулированы или исключены из банка заданий.

Усиление контроля при разработке заданий. Необходимо усилить процедуру разработки и содержательной экспертизы тестовых заданий. Это позволит минимизировать ошибки в формулировках и исключить неоднозначность, что в свою очередь повысит их соответствие психометрическим требованиям.

Повторный анализ. После внесения всех корректировок необходимо провести повторную апробацию и повторный психометрический анализ. Это позволит объективно оценить степень улучшения качества теста и подтвердить его соответствие требованиям валидности и надежности.

Заключение

Проведенный психометрический анализ апробации подтверждает общее приемлемое качество тестовых заданий по математике, используемых в рамках ОЗП. Высокие показатели надежности и разделяющей способности тестов свидетельствуют об их пригодности для массовой аттестации. Замечания, выявленные в ходе апробации, были приняты во внимание, и на их основе была проведена необходимая работа по улучшению качества банка заданий. В частности, была проведена дополнительная экспертная проверка заданий, показавших отрицательную дискриминативность, и были пересмотрены или исключены задания, где дистракторы вводили в заблуждение сильных испытуемых. Также был усилен контроль за содержательной экспертизой на этапе разработки тестовых заданий и проведена обязательная апробация новых заданий. Проведенная работа по корректировке банка тестовых заданий и повторный психометрический анализ позволили обеспечить высокое качество оценочных инструментов, что, в свою очередь, способствует повышению эффективности всей системы аттестации педагогических кадров в Казахстане.

Литература

1. Приказ Министра образования и науки Республики Казахстан от 27 января 2016 года №83 «Об утверждении Правил и условий проведения аттестации педагогов».
2. Закон Республики Казахстан от 27 июля 2007 года № 319-III. Об образовании.
3. Оценка знаний педагогов. Ұлттық тестілеу орталығы. http://new.testcenter.kz/?page_id=15731&lang=ru
4. Методические рекомендации по развитию профессиональной компетенции педагогов. – г. Астана: Национальная академия образования имени И. Алтынсарина, 2024. – 201 стр.
5. Магомедов, А.Г. (2022) Применение тестирования как инструмента педагогических измерений: возможности и ограничения. Современное педагогическое образование, №9, 62-67.
6. Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests // Psychometrika. 1951. № 16. P. 297-334.
7. Карданова Е.Ю., Иванова А.Е. (2023) Психометрические исследования: современные методы и новые возможности для образования. Вопросы образования, № 3, 8-19.
8. Корнилов, С.А. Проблема надежности психологических шкал и ее значение для психодиагностики в научных и практических целях. Экспериментальная психология в России: традиции и перспективы, 101-105.
9. Crocker, L., & Algina, J. (2008). Introduction to Classical and Modern Test Theory. Cengage Learning.
10. Статистика обучения с jamovi. https://lsj.readthedocs.io/ru/latest/.
11. Thompson, N. Classical Statistics and their Interpretation. https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation.
12. Evaluating Item Difficulty and Discrimination in Knowledge Tests, January 4, 2024. https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/.
13. Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice), Mike Linacre, instructor - July 2011, www.winsteps.com.
14. Е.Б. Белов, М.В. Алексеев, Н.П. Китаев, А.И. Кучумов. Анализ влияния первичных баллов заданий на уровни подготовленности респондентов в дихотомической модели Раша на основе взвешенного метода максимального правдоподобия. Статистика и математические методы в экономике, 2019, №2.
15. Математико-статистические методы в эмпирических социально-экономических исследованиях: учеб. пособие / И.Н. Дубина. – М.: Финансы и статистика; ИНФРА-М. – 2010. – 416с.: ил.

Ж.Ж. Бейсенова, Л.Е. Шинетова, Г.К. Жабаева

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНДАҒЫ ПЕДАГОГТЕРДІҢ БІЛІМІН БАҒАЛАУ БОЙЫНША ТЕСТ ТАПСЫРМАЛАРЫНЫҢ ПСИХОМЕТРИЯЛЫҚ ТАЛДАУЫ (АПРОБАЦИЯ НӘТИЖЕЛЕРІ БОЙЫНША)

Бұл мақалада Қазақстанда педагогтердің білімін бағалау рәсімінде (ОББ) қолданылатын математика пәні бойынша тест тапсырмаларының апробациясы барысындағы толық психометриялық талдауы ұсынылған. Зерттеу аттестаттау кезінде қолданылған тесттің үш түрлі нұсқасын салыстырмалы талдауға негізделген. Тапсырмалардың сапасын бағалау және олардың күшті, әлсіз жақтарын анықтау үшін классикалық тест теориясы (КТТ) және қазіргі тест теориясы (IRT), атап айтқанда, Раш моделінің әдістері қолданылды. Жұмыстың мақсаты қолданыстағы бағалау құралдарын бағалаумен қатар, оларды жетілдіру бойынша нақты ұсыныстар әзірлеу болды. Апробация талдауы тесттердің жалпы сенімділігінің жоғары болуына (Cronbach's

Alpha коэффициенті 0,84-тен жоғары) және олардың тестіленушілерді дайындық деңгейі бойынша жақсы ажырата білуіне қарамастан, әр нұсқада бірқатар проблемалық тапсырмалар анықталғанын көрсетті. Атап айтқанда, дискриминативтілігі төмен немесе тіпті теріс тапсырмалар, сондай-ақ тиімсіз дистракторлар (қате жауап нұсқалары) анықталды. Апробация барысында анықталған барлық ескертулер ескеріліп, тест тапсырмаларының қорын түзету бойынша кейінгі жұмыстарда жойылды. Бұл бағалау құралдарының жалпы дәлдігі мен валидтілігін арттыруға ықпал етті.

Түйін сөздер: педагогтердің білімін бағалау, психометриялық талдау, апробация, сенімділік, валидтілік, классикалық тест теориясы, Раш моделі.

Zh.Zh. Beisenova, L.Y. Shinetova, G.K. Zhabayeva

PSYCHOMETRIC ANALYSIS OF TEST ITEMS FOR ASSESSING TEACHER KNOWLEDGE IN THE REPUBLIC OF KAZAKHSTAN (BASED ON APPROBATION RESULTS)

This article provides a detailed psychometric analysis of mathematics test items used during the approbation of the teacher knowledge assessment procedure (OZP) in Kazakhstan. The research is based on a comparative analysis of three different test versions administered during this piloting phase. To evaluate the quality of the items and identify their strengths and weaknesses, the study employed methods from both classical test theory (CTT) and modern item response theory (IRT), specifically the Rasch model. The purpose of the work was not only to assess the existing evaluation instruments but also to develop practical recommendations for their improvement. The approbation analysis showed that, despite the overall high reliability of the tests (Cronbach's Alpha coefficient exceeding 0,84) and their good ability to differentiate test-takers by proficiency level, a number of problematic items were found in each version. Specifically, items with low or even negative discrimination and ineffective distractors were identified. All issues discovered during the approbation have since been taken into account and rectified in the subsequent work on refining the test item bank. This has contributed to improving the overall accuracy and validity of the assessment instruments.

Keywords: teacher knowledge assessment, psychometric analysis, approbation, reliability, validity, classical test theory, Rasch model.

References

1. Prikaz Ministra obrazovaniya i nauki Respubliki Kazakhstan ot 27 yanvarya 2016 goda №83 «Ob utverzhdenii Pravil i usloviy provedeniya attestatsii pedagogov».
2. Zakon Respubliki Kazakhstan ot 27 iyulya 2007 goda № 319-III. Ob obrazovanii.
3. Otsenka znaniy pedagogov. Ұlttyq testileu ortalygy.
4. Metodicheskiye rekomendatsii po razvitiyu professionalnoy kompetentsii pedagogov. – g. Astana: Natsionalnaya akademiya obrazovaniya imeni I. Altynsarina, 2024. – 201 str.
5. Magomedov, A.G. (2022) Primeneniye testirovaniya kak instrumenta pedagogicheskikh izmereniy: vozmozhnosti i ogranicheniya. Sovremennoye pedagogicheskoye obrazovaniye, №9, 62-67.
6. Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests // Psychometrika. 1951. № 16. P. 297-334.
7. Kardanova, Ye.Yu., Ivanova A.Ye. (2023) Psikhometricheskiye issledovaniya: sovremennyye metody i novyye vozmozhnosti dlya obrazovaniya. Voprosy obrazovaniya, № 3, 8-19.
8. Kornilov, S.A. Problema nadezhnosti psikhologicheskikh shkal i yeye znacheniye dlya psikhodiagnostiki v nauchnykh i prakticheskikh tselyakh. Eksperimentalnaya psikhologiya v Rossii: traditsii i perspektivy, 101-105.
9. Crocker, L., & Algina, J. (2008). Introduction to Classical and Modern Test Theory. Cengage Learning.
10. Statistika obucheniya s jamovi. https://lsj.readthedocs.io/ru/latest/.
11. Thompson, N. Classical Statistics and their Interpretation. https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation.
12. Evaluating Item Difficulty and Discrimination in Knowledge Tests, January 4, 2024. https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/.
13. Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice), Mike Linacre, instructor - July 2011, www.winsteps.com.

14. Belov, Ye.B., Alekseyev, M.V., Kitayev, N.P., & Kuchumov, A.I. (2019). Analiz vliyaniya pervichnykh ballov zadaniy na urovni podgotovlennosti respondentov v dikhotomicheskoy modeli Rasha na osnove vzveshennogo metoda maksimalnogo pravdopodobiya. Statistika i matematicheskiye metody v ekonomike, №2.
15. Matematichesko-statisticheskiye metody v empiricheskikh sotsialno-ekonomicheskikh issledovaniyakh: ucheb. posobiye / I.N. Dubina. – M.: Finansy i statistika; INFRA-M. – 2010. – 416s.: il.

Сведения об авторах:
Бейсенова Жанар Жумагазиновна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий научный сотрудник, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: zhanar-beisenova@mail.ru.
Шинетова Ляззат Ермековна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, заведующий лабораторией, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: shinetovalyazzat24@gmail.com.
Жабаева Гульдана Куатовна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий научный сотрудник, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: danka1919@mail.ru.

Авторлар туралы мәлімет:
Бейсенова Жанар Жумагазиновна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші ғылыми қызметкер, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: zhanar-beisenova@mail.ru
Шинетова Ляззат Ермековна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, зертхана меңгерушісі, Родниковая көшесі, 1/1, 010008, Астана, Қазақстан, e-mail: shinetovalyazzat24@gmail.com.
Жабаева Гульдана Куатовна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші ғылыми қызметкер, Родниковая көш., 1/1, 010008, Астана, Қазақстан, danka1919@mail.ru.

Information about authors:
Beisenova Zhanar Zhumagazinovna – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Researcher, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: zhanar-beisenova@mail.ru.
Shinetova Lyazzat Yermekovna – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Head of Laboratory, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: shinetovalyazzat24@gmail.com.
Zhabaeva Guldana Kuatovna – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Leading Researcher, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: danka1919@mail.ru.