

Ж.Ж. Бейсенова* , Л.Е. Шинетова ,
Н.А. Диарбекова , Г.К. Жабаева 

Национальный центр тестирования, Астана, Казахстан

*e-mail: zhanar-beisenova@mail.ru

МЕТОДЫ RASCH-МОДЕЛИРОВАНИЯ В WINSTEPS: ПОДХОДЫ К АНАЛИЗУ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ

В статье рассматривается применение программного обеспечения Winsteps 5.8.3 для анализа итогов оценивания с использованием методов Rasch-моделирования. Winsteps предоставляет мощные инструменты для обработки и анализа данных, полученных в результате оценивания, что позволяет исследователям и практикам в области педагогики и психометрии глубже понять результаты тестирования. Статья охватывает ключевые функции программы, включая этапы подготовки данных, построение модели и интерпретацию результатов. Исследование подчеркивает важность точного анализа данных для повышения качества образовательных процессов и разработки эффективных оценочных инструментов. Результаты анализа, полученные с помощью Winsteps, могут быть использованы для принятия обоснованных решений в области образования, таких как корректировка учебных программ, разработка новых методик обучения или изменение стратегий оценивания. Возможности визуализации данных в Winsteps помогают исследователям и педагогам лучше понимать результаты тестирования, что способствует более точной интерпретации и коммуникации результатов. Статья имеет высокую практическую ценность для специалистов, занимающихся разработкой и анализом тестовых заданий, а также для тех, кто стремится улучшить качество образовательных программ и методов оценивания. Она предоставляет не только теоретические основы, но и конкретные рекомендации, которые могут быть применены в реальной практике.

Ключевые слова: Winsteps, Rasch-моделирование, анализ итогов оценивания, психометрика, оценочные инструменты.

Zh.Zh. Beisenova*, L.E. Shinetova, N.A. Didarbekova, G.K. Zhabayeva

National Testing Center, Astana, Kazakhstan

*e-mail: zhanar-beisenova@mail.ru

Rasch Modeling Methods in Winsteps: Approaches to the Analysis of Educational Results

This article discusses the application of the Winsteps 5.8.3 software for analyzing assessment results using Rasch modeling methods. Winsteps provides powerful tools for processing and analyzing assessment data, enabling researchers and practitioners in the fields of pedagogy and psychometry to gain a deeper understanding of test results. The article covers the key functions of the program, including the stages of data preparation, model construction, and interpretation of results. The study highlights the importance of accurate data analysis to improve the quality of educational processes and develop effective assessment tools. The results of the analysis obtained with the help of Winsteps can be used to make informed decisions in the field of education, such as adjusting curricula, developing new teaching methods, or changing assessment strategies. Data visualization capabilities in Winsteps help researchers and educators better understand test results, which facilitates more accurate interpretation and communication of results. The article has a high practical value for specialists involved in the development and analysis of test tasks, as well as for those who seek to improve the quality of educational programs and assessment methods. It provides not only theoretical foundations, but also specific recommendations that can be applied in real practice.

Keywords: Winsteps, Rasch modeling, assessment outcome analysis, psychometrics, assessment tools.

Ж.Ж. Бейсенова*, Л.Е. Шинетова, Н.А. Дидарбекова, Г.К. Жабаева

Ұлттық тестілеу орталығы, Астана, Қазақстан

*e-mail: zhanar-beisenova@mail.ru

Winsteps-тері Rasch модельдеу әдістері: білім беру нәтижелерін талдау тәсілдері

Бұл мақалада Rasch модельдеу әдістерін қолдана отырып, бағалау нәтижелерін талдау үшін Winsteps 5.8.3 бағдарламалық жасақтамасын пайдалану қарастырылады. Winsteps педагогика мен психометрика саласындағы зерттеушілер мен тәжірибешілерге тест нәтижелерін тереңірек түсінуге мүмкіндік беретін бағалаудан алынған деректерді өңдеуге және талдауға арналған қуатты құралдарды ұсынады. Мақала мәліметтерді дайындау, модель құру және нәтижелерді интерпретациялау кезеңдері қарастырылған бағдарламаның негізгі функцияларын қамтитын. Зерттеу білім беру процестерінің сапасын арттыру және тиімді бағалау құралдарын өзірлеу үшін деректерді нақты талдаудың маңыздылығын көрсетеді. Winstep көмегімен алынған талдау нәтижелерін оқу бағдарламаларын түзету, оқытудың жаңа әдістерін өзірлеу немесе бағалау стратегияларын өзгерту сияқты білім беру саласында негізделген шешімдер қабылдау үшін пайдалануға болады. Winstep-тері деректерді визуализациялау мүмкіндіктері зерттеушілер мен мұғалімдерге тестілеу нәтижелерін жақсы түсінуге көмектеседі, бұл нәтижелерді дәлірек түсініруге және байланыстыруға ықпал етеді. Мақала тест тапсырмаларын өзірлеумен және талдаумен айналысадын мамандар үшін, сондай-ақ білім беру бағдарламалары мен бағалау әдістерінің сапасын арттыруға ұмытылатындар үшін жоғары практикалық құндылыққа ие. Ол тек теориялық, негіздерді ғана емес, сонымен қатар нақты тәжірибеде қолдануға болатын нақты ұсыныстармен қамтамасыз етеді.

Түйін сөздер: Winsteps, Rasch-модельдеу, бағалау нәтижелерін талдау, психометрика, бағалау құралдары.

Введение

Программное обеспечение Winsteps предназначено для анализа данных о тестировании, использующее метод Раш для анализа и понимания конечных результатов. Winsteps используется практиками, которые должны принимать надежные и быстрые решения на пути к созданию эффективных тестов. Winsteps разработан как инструмент, облегчающий исследование и анализ и применяется в сфере оценки образовательных достижений, социологических исследований и других областях, требующих глубокой аналитики данных.

Метод Раш был изобретен датским ученым Георгом Рашем в 1953 году для анализа ответов на серию тестов на чтение. Он является психологической моделью, применяемой для создания и анализа тестов. Ключевая мысль модели в том, что вероятность правильного ответа на задание обуславливается расхождением между способностями испытуемого и трудностью задания. Модель дает возможность определить как, к примеру, уровень знаний испытуемого, так и характеристики тестовых заданий, такие, например, как трудность (Деменченок, 2010: 3). Самое главное в Rasch Measurement – это то, что создаваемый при этой технологии тест гарантированно становится качественным средством пе-

дагогического измерения. Все прочие известные методы и технологии разработки тестов такой гарантии не давали и не дают. Тем самым теорией Rasch Measurement был открыт новый путь к эффективной технологии разработки тестов более высокого качества (Аванесов, 2011: 16).

В статье нами рассматриваются основные возможности программы Winsteps, а также методы анализа данных результатов. Целью исследования является анализ применения методов Раш в программе Winsteps для оценки результатов обучения и определение аспектов, имеющих воздействие на показатели качества обучения.

Winsteps разработан для типичных образовательных или опросных данных, где набор вопросов, также называемых «тестовыми заданиями», задается выборке лиц. Вопросы могут быть с множественным или единичным выбором. Winsteps анализирует эти данные, чтобы получить «способность», «отношение» или подобную меру для каждого человека и меру сложности для каждого вопроса, а также множество других диагностических показателей (Линакре, 2012: 2).

Основные возможности Winsteps.

1. Для выявления исследователем заданий, которые являются более трудными для испытуемых и для определения уровня способности, Winsteps дает возможность подробного анализа

показателей тестовых элементов с определением их трудности и дискриминационной способности.

2. Winsteps приспосабливается к вариативным разновидностям данных и гарантирует надежные результаты, что является очень важным при использовании инструментов оценивания.

3. Winsteps включает широкий спектр средств для графического отображения результатов исследования, что необходимо, чтобы сделать анализ более визуальным и доступным для понимания.

4. Winsteps создает множество видов отчетов, которые необходимы для интерпретации конечных данных оценивания (Линакре, 2011: 1).

Все это дает возможность исследователям верно определить сложность задания, уровень возможностей и умений участников, пригодность данных конкретной модели и это, в конечном счете, содействует повышению достоверности и результативности средств оценки.

Обзор литературы

В Казахстане в последние годы заметен растущий интерес к методам Rasch-моделирования как среди ученых, аналитиков и научных сотрудников, так и среди практиков в образовательной среде, в связи с тем, что методы Rasch-моделирования являются важным инструментом оценивания, который дает возможность на основе теории вероятностей производить анализ. В связи с этим, отмечается рост научных публикаций в отечественной и зарубежной научной среде в сфере педагогических измерений и психометрики.

Rasch-моделирование широко используется при анализе тестовых заданий и результатов. Так, исследования таких авторов, как В.С. Аванесов (2010) , А.А. Маслак (2009) , В.С. Ким (2008) , Е.Ю. Карданова (2023: 8) посвящаются использованию методов Rasch-моделирования для определения качества тестовых заданий, их надежности и валидности, исследованиям в области психометрики и педизмерений и созданию технологий обнаружения ошибок. Исследования Е.Б. Белова, М.В. Алексеева, Н.П. Китаева (2019: 89) наглядно показывают как Rasch-моделирование помогает в выполнении сравнительного анализа между различными группами обучающихся, а также отслеживать динамику образовательных результатов.

В зарубежной практике, такие исследова-

тели, как, Райт Б.Д., Линакре Дж.М(1994) выявили, что в результате исследований и анализа сотен наборов данных статистика согласия интерпретируется на уровне параметров. При этом правильно определенный интервал статистики согласия составляет (0,5; 1,5) . Также в ходе нашего исследования былоделено большое внимание материалам международных передовых организаций в данной сфере, таких, как ETS, NFER, AQA Global Assessment Services.

Казахстанские исследования с применением методов Rasch-моделирования с каждым годом набирают популярность. Данные исследования привлекают внимание к вопросам повышения надежности и валидности современных инструментов оценивания при соответствии их стандартам Rasch, фокусируются на применении программных обеспечений при анализе итогов национальных тестирований. В отечественной литературе большое внимание уделяется работам, демонстрирующим, как применение Rasch-моделирования при анализе результатов тестирования помогает оценивать и прослеживать прогресс обучающихся, а также как образовательные программы оказывают влияние на достижения обучающихся.

Все это указывает на все большее увеличивающийся интерес в применении методов Rasch-моделирования в Казахстане. Большое внимание казахстанских ученых уделяется повышению качества образовательных результатов путем оптимизации системы оценивания в стране. При этом, как отечественные, так и зарубежные исследователи отмечают актуальность и потенциал программного обеспечения Winsteps для анализа результатов оценки знаний обучающихся.

Материалы и методы

Winsteps работает на платформе Windows и предоставляет графический интерфейс для анализа данных. Выходные данные состоят из множества полезных графиков и таблиц, которые можно импортировать в письменные отчеты.

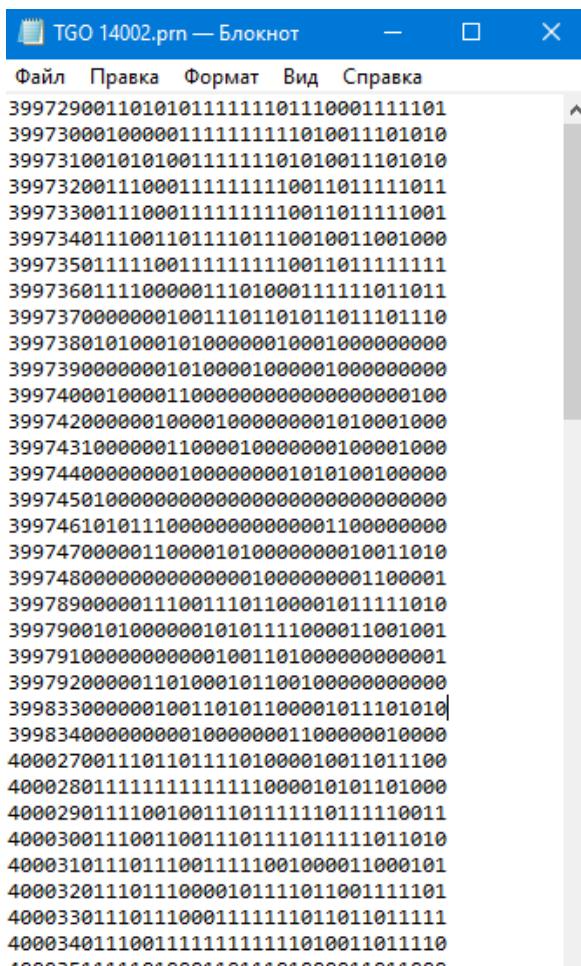
Перед началом анализа необходимо подготовить данные. Формат данных для Winsteps требует, чтобы каждая строка представляла собой ответ респондента, а каждый столбец – отдельный элемент теста. Важно учитывать, что данные должны быть очищены от пропусков и аномалий, так как это может повлиять на результаты анализа (Линакре, 2012: 3) .

За исходные данные мы взяли данные апро-

бации тестовых заданий комплексного тестирования в магистратуру, в частности, теста на определение готовности к обучению, вариант на русском языке. Всего тестируемых 364 обучающихся, количество тестовых заданий – 30.

Комплексное тестирование для поступающих в магистратуру проводится на основании приказов Министра образования и науки Республики Казахстан от 8 мая 2019 года №190 и от 31 октября 2018 года №600. Один из блоков комплексного тестирования в магистратуру тест на определение готовности к обучению состоит из 30 заданий, тест состоит из двух блоков. Первый блок – критическое мышление, второй – аналитическое мышление. Все тестовые задания представляются в закрытой форме (*Об утверждении Типовых правил приема, 2018*), (*Об утверждении Правил проведения, 2019*).

В блокноте исходник выглядит таким образом (рисунок 1). В дальнейшем будем работать с этим документом.



```

TGO 14002.prn — Блокнот
Файл Правка Формат Вид Справка
39972900110101011111101110001111101
399730001000011111111010011101010
3997310010101001111111010011101010
39973200111000111111110011011111011
39973300111000111111110011011111001
399734011100110111101110010011001000
39973501111100111111110011011111111
39973601111000011101000111111011011
3997370000001001110110110111101110
399738010100010100000010001000000000
39973900000010100001000001000000000
3997400010000110000000000000000000100
399742000001000010000000001010001000
399743100000110000100000000100001000
399744000000010000000001010100100000
39974501000000000000000000000000000000
3997461010111000000000000011000000000
3997470000110000101000000010011010
39974800000000000000001000000001100001
39978900000111001110110001011111010
39979001010000001010111100011001001
399791000000000001001101000000000001
399792000011010001011001000000000000
39983300000010011010110001011101010
3998340000000001000000011000000010000
400027001110110111101000010011011100
40002801111111111111000010101101000
40002901111001001110111110111110011
40003001110011001111011111011111010
400031011101110011111001000011000101
40003201110111000101111011001111101
40003301110111000111111011011011111
400034011100111111111111010011011110

```

Рисунок 1 – Исходный документ для анализа данных в Winsteps

Далее открываем главное окно Winsteps. В верхней части экрана находится строка меню Winsteps (Янченко, 2006: 115).

В диалоговом окне Winsteps нажимаем на «Управление + Процедура настройки данных»



Откроется экран Winsteps «Настройка управляющего файла» (рисунок 2). Это один из многих способов создания управляющих файлов и файлов данных Winsteps (Линакре, 2012: 3).

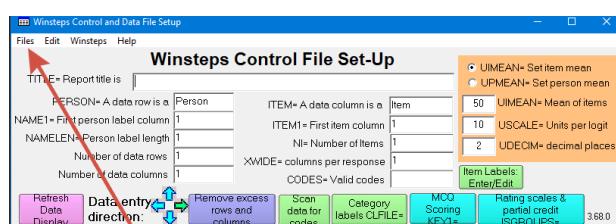
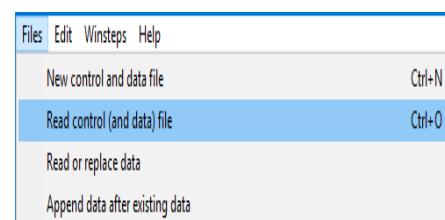


Рисунок 2 – Загрузка файла в Winsteps

Мы рассмотрим один из примеров файлов данных. Нажимаем на файл управления чтением (и данных)



Отобразится список каталогов. Выбираем файл.

Появится следующая электронная таблица данных (рисунок 3).

Информация, идентифицирующая личность, находится в столбцах с 1 по 13. Она выделена синей полосой. Ответы на вопросы приведены в столбцах с 14 по 43. Всего 30 пунктов, они обозначены зеленой полосой.

The screenshot shows the 'Winsteps Control File Set-Up' dialog box. On the right, there are configuration options for item means (UIMEAN= Set item mean, UPMEAN= Set person mean), item scaling (USCALE= Units per logit), and decimal places (UDECIM= decimal places). Below these are sections for 'Category labels', 'CLFILE', 'MCQ Scoring KFV1', and 'Rating scales & partial credit ISGROUPS'. The main area contains a data entry grid with columns labeled 1 through 32. The first few rows of data are as follows:

1	3	9	9	7	2	9	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
2	3	9	9	7	3	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1
3	3	9	9	7	3	1	0	0	1	0	1	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1
4	3	9	9	7	3	2	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
5	3	9	9	7	3	3	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
6	3	9	9	7	3	4	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1
7	3	9	9	7	3	5	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
8	3	9	9	7	3	6	0	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	0	1	1
9	3	9	9	7	3	7	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1	0	1	1	0
10	3	9	9	7	3	8	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Рисунок 3 – Электронная таблица данных в Winsteps

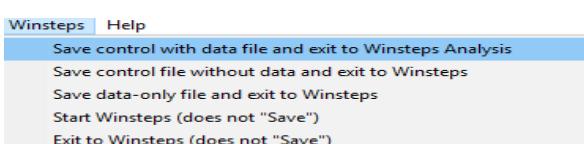
Возвращаясь к экрану настройки Winsteps, нажатие «Проверить данные на наличие кодов»

определяет, сколько столбцов должно быть в каждом ответе (рисунок 4).

The screenshot shows the same 'Winsteps Control File Set-Up' dialog box as in Figure 3. A red arrow points to the 18th column of the data entry grid, which is highlighted in blue. The data grid is identical to the one in Figure 3.

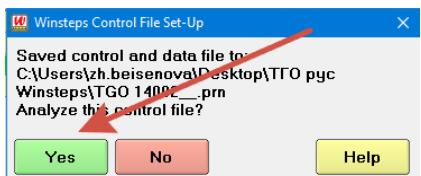
Рисунок 4 – Проверка данных на наличие кодов

Теперь можно приступить к анализу по методу Раша. В строке меню настройки нажать на Winsteps



Нажать на Сохранить управление с файлом данных и перейти к Winsteps Analysis.

Процедура настройки Winsteps подтверждает, что она сохранила ваш файл управления и данных, и спрашивает, хотите ли вы проанализировать их: Нажмите «Да»



Далее Winsteps запрашивает «Дополнительные спецификации». Это дополнительные инструкции по управлению. У нас их нет, нажимаем Enter (2 раза).

Запускается анализ Winsteps (рисунок 5).

```

Temporary Workfile Directory: C:\Users\ZHD524-1.BEI\AppData\Local\Temp\
Reading Control Variables ..
W4: Oops! Item labels = 0, but NI=30
NO ITEM LABELS. NI= 30
Input in process: Input Data Record: Length = 36
39972900110101011111101110001111101
^P ^I ^N
364 Person Records Input

CONVERGENCE TABLE
-Control: \TGO 14002.prn Output: \ZOU907WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION Person Item CATS Person Item MEASURES STRUCTURE|
>=====|
| 1 364 30 2 4.89 2.14 -3.3673 |
>=====|
| 2 364 30 2 5.38 2.47 -.3345 |
| Probing data connection: to skip out: Ctrl+F - to bypass: subset=no |
| Processing unanchored persons ... |
| Data fully connected. No subsets found |
| Control: \TGO 14002.prn Output: \ZOU907WS.TXT |
| JMLE MAX SCORE MAX LOGIT LEAST CONVERGED CATEGORY STRUCTURE |
| ITERATION RESIDUAL* CHANGE Person Item CAT RESIDUAL CHANGE |
>=====|
| 1 4.58 -.0665 32 24* |
| 2 2.18 -.0321 10 24* |
| 3 1.09 -.0166 10 24* |
| 4 .55 -.0085 69 24* |
| 5 .28 -.0046 69 24* |
| |

```

Рисунок 5 – Начало анализа в программе Winsteps. Таблицы сходимости данных

Файл управления и данных обрабатывается (зеленое поле). Отображается первая запись данных. Это полезно для проверки правильности размещения данных (^P = начало метки пользователя, ^I = начало ответов на элементы, ^N = окончание ответов на вопросы).

Указывается количество записей данных. Здесь указано «364 Person», что соответствует количеству тестируемых. Все данные верные. Winsteps начинает делать свои первоначальные оценки (красное поле). Далее результаты анализа суммируются и отображаются в синем поле (рисунок 6) (Смирнова, 2006: 106) [13].

Calculating JMLE Fit Statistics

Time for estimation: 0:0:0.937

Output to C:\Users\zh.beisenova\Desktop\Winsteps\ZOU907WS.TXT

C:\Users\zh.beisenova\Desktop\Winsteps\IGO 14002.prn

	Person		364 INPUT		364 MEASURED		INFIT		OUTFIT		
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD			
MEAN	15.3	30.0	.04	.45	1.00	-.1	1.03	.01			
P.SD	5.9	.0	1.09	.08	.18	1.0	.36	1.01			
REAL RMSE	.46	TRUE SD	.99	SEPARATION	2.14	Person	RELIABILITY	.82			

	Item		30 INPUT		30 MEASURED		INFIT		OUTFIT		
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD			
MEAN	186.1	364.0	.00	.13	.99	-.1	1.03	.31			
P.SD	61.5	.0	.91	.01	.09	1.5	.19	1.81			
REAL RMSE	.13	TRUE SD	.90	SEPARATION	7.16	Item	RELIABILITY	.981			

Output written to C:\Users\zh.beisenova\Desktop\Winsteps\ZOU907WS.TXT

CODES= 01

measures constructed: use "Diagnosis" and "Output Tables" menus

Рисунок 6 – Таблицы сходимости данных

После выбора модели и импорта данных Winsteps проводит калибровку, вычисляя параметры элементов теста и способности респондентов. Таким образом, Winsteps предоставляет инструменты для проверки соответствия данных выбранной модели, включая анализ остаточных значений и графиков.

Результаты и обсуждение

Анализ свойств компонентов. В процессе исследования можно оценить как хорошо каждая составляющая тестового задания выполняет свою задачу и выявить элементы, являющиеся слишком простыми или сложными.

На рисунке 7 (данные таблицы 14.1 в Winsteps) видно, что сложными заданиями здесь являются 1, 5, 19, 22 и 30, на основании данных местоположения предмета (Item Measure), которое показывает трудность задания в логитах. Задания с высокими значениями сложности (например, 1 логит или больше) более сложны для участников. Это может означать о том, что задания диссонируют с уровнем подготовки группы испытуемых или являются сложными для пони-

мания (Фер, 2010: 375) .

Оценка восстановления (Item Fit Statistics) показывает, насколько хорошо задание вписывается в модель Rasch. MNSQ и ZSTD – границы модели теста. Infit и Outfit – это два статистических показателя, которые показывают, насколько «подходит» задание к предполагаемой модели. Средние значения общей (OUTFIT MNSQ) и взвешенной общей статистики согласия (INFIT MNSQ) находятся в пределах 0,5-1,5. Средние значения стандартизированной общей (OUTFIT ZSTD) и стандартизированной взвешенной общей статистики (INFIT ZSTD) находятся в пределах -2-2. Следовательно, задания теста хорошо согласуются с моделью. Если значения слишком высокие или слишком низкие, это может указывать на то, что задание плохо соответствует модели или участники отвечали на него необычным образом (Примбетова, 2014: 51).

Ниже, в таблице 1 представлено распределение тестовых заданий по уровням согласованности. Из таблицы видно, что в пяти заданиях (1, 5, 10, 19, 30) статистики согласия не входят в указанный интервал (16,7%) и коэффициент корреляции в этих заданиях низкий.

Файл Правка Формат Вид Справка														
TABLE 14.1 C:\Users\zh.beisenova\Desktop\ТГО рус ZOU156WS.TXT_ Dec 02 2024 14:40														
INPUT: 364 Person 30 Item REPORTED: 364 Person 30 Item 2 CATS WINSTEPS 5.8.3.0														
Person: REAL SEP.: 2.14 REL.: .82 ... Item: REAL SEP.: 7.16 REL.: .98														
Item STATISTICS: ENTRY ORDER														
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT [MNSQ	OUTFIT ZSTD	[MNSQ	ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	MATCH EXP%	OBS%	EXP%	Item
1	80	364	1.59	.14	1.04	.58	1.32	2.04	.28	.34	76.4	79.0		I0001
2	224	364	-.50	.12	.94	-1.24	.95	-.68	.49	.44	71.7	71.6		I0002
3	257	364	-1.01	.13	.95	-.79	.93	-.68	.48	.44	78.6	76.5		I0003
4	229	364	-.58	.12	.91	-1.63	.89	-1.45	.51	.44	75.3	72.3		I0004
5	111	364	1.07	.12	1.10	1.94	1.23	2.00	.27	.38	70.1	72.7		I0005
6	165	364	.31	.12	.95	-1.16	.96	-.49	.46	.42	73.4	67.7		I0006
7	189	364	-.02	.12	1.05	1.08	1.08	1.14	.39	.43	66.8	68.4		I0007
8	204	364	-.22	.12	.98	-.50	1.02	.34	.45	.44	71.4	69.5		I0008
9	138	364	.68	.12	.95	-1.25	1.11	1.26	.42	.48	73.6	69.2		I0009
10	122	364	.91	.12	1.09	1.92	1.31	2.95	.28	.39	68.4	71.0		I0010
11	250	364	-.90	.13	1.04	.60	1.23	2.41	.38	.44	78.0	75.4		I0011
12	252	364	-.93	.13	.82	-3.11	.74	-3.09	.59	.44	79.7	75.7		I0012
13	233	364	-.64	.12	.87	-2.39	.82	-2.38	.55	.44	75.8	72.8		I0013
14	211	364	-.32	.12	.90	-2.08	.86	-2.14	.52	.44	72.5	70.1		I0014
15	308	364	-2.03	.16	.74	-2.87	.48	-3.60	.61	.39	87.4	85.6		I0015
16	251	364	-.91	.13	1.06	.99	1.09	1.01	.38	.44	75.0	75.6		I0016
17	153	364	.47	.12	1.03	.68	1.04	.56	.39	.41	67.0	68.0		I0017
18	245	364	-.82	.12	.99	-.14	1.01	.13	.44	.44	77.7	74.6		I0018
19	96	364	1.31	.13	1.04	.71	1.29	2.20	.29	.36	76.6	75.5		I0019
20	149	364	.53	.12	1.03	.62	1.09	1.17	.38	.41	67.3	68.3		I0020
21	155	364	.44	.12	.99	-.11	1.00	.06	.42	.42	68.7	68.0		I0021
22	101	364	1.23	.13	1.01	.21	1.17	1.39	.34	.37	74.7	74.5		I0022
23	240	364	-.74	.12	.98	-.38	1.00	-.01	.45	.44	76.6	73.9		I0023
24	296	364	-1.75	.15	.80	-2.40	.69	-2.24	.56	.41	86.0	83.1		I0024
25	172	364	.21	.12	.99	-.13	1.00	-.01	.43	.43	69.2	67.8		I0025
26	186	364	.02	.12	1.08	1.74	1.11	1.61	.36	.43	64.3	68.2		I0026
27	200	364	-.17	.12	1.16	3.31	1.26	3.57	.30	.44	65.4	69.1		I0027
28	125	364	.86	.12	1.00	.11	1.03	.39	.38	.39	73.1	70.6		I0028
29	140	364	.65	.12	1.01	.16	.98	-.15	.40	.40	69.8	69.0		I0029
30	100	364	1.25	.13	1.09	1.51	1.25	1.99	.27	.36	73.4	74.7		I0030
MEAN	186.1	364.0	.00	.12	.99	-.13	1.03	.31			73.5	72.6		
P.SD	61.5	.0	.91	.01	.09	1.52	.19	1.76			5.4	4.4		

Рисунок 7 – Таблица 14.1 с уровнями трудностей заданий, полученная с помощью Winsteps

Таблица 1 – Распределение тестовых заданий по уровням согласованности

№ варианта	Задания со слабой внутренней согласованностью (0-0,3)	Задания с хорошей внутренней согласованностью (0,31-0,80)	Задания с высокой внутренней согласованностью (0,81-1)
Item1	0,28		
Item2		0,50	
Item3		0,50	
Item4		0,52	
Item5	0,28		
Item6		0,45	
Item7		0,40	
Item8		0,46	
Item9		0,41	
Item10	0,29		
Item11		0,42	
Item12		0,60	
Item13		0,55	
Item14		0,52	
Item15		0,65	
Item16		0,42	
Item17		0,39	
Item18		0,47	

Продолжение таблицы

№ варианта	Задания со слабой внутренней согласованностью (0-0,3)	Задания с хорошей внутренней согласованностью (0,31-0,80)	Задания с высокой внутренней согласованностью (0,81-1)
Item19	0,29		
Item20		0,38	
Item21		0,41	
Item22		0,33	
Item23		0,47	
Item24		0,60	
Item25		0,43	
Item26		0,38	
Item27		0,32	
Item28		0,37	
Item29		0,39	
Item30	0,27		

По данным таблицы 1 можно сделать следующие выводы: основная часть тестовых заданий (76,67%) расположена в категории «хорошей внутренней согласованности», это показывает, что они применимы для оценки знаний и навыков. Отсутствие заданий с высокой внутренней согласованностью может сигнализировать о том, что тестовые задания требуют доработки или переработки для достижения более высокого уровня согласованности. Уровень слабой внутренней согласованности (16,67%) также требует внимания, так как эти задания могут вносить неопределенность в результаты тестирования. Следует принять во внимание, что отклонение показателей данных по согласованию от норм в данных условиях незначительное.

Степень компетентности участников. Winsteps помогает определить группы респондентов, которой может потребоваться дополнительное обучение и поддержка, приводя подробные отчеты о способностях участников. Winsteps исследует уровни подготовленности каждого испытуемого, анализируя при этом согласие с моделью измерения результатов каждого участника на все задания теста. При этом для каждого испытуемого рассматриваются общая OUTFIT MNSQ и взвешенная INFIT MNSQ статистики и их стандартизованные версии (OUTFIT ZSTD и INFIT ZSTD).

Статистическим показателем, использующимся при анализе внутренней согласованности, является коэффициент альфа-Кронбаха (α). Он может принимать значения от 0 до 1: значение близкое к 1 показывает высокую степень

внутренней согласованности между элементами шкалы, значение ниже 0,6 – обычно считается низким, от 0,6 до 0,8 – приемлемым и 0,8 и выше – хорошим. В нашем анализе коэффициент альфа-Кронбаха равен 0,84 (табл. 3.1 Summary Statistics в Winsteps), что говорит о том, что тест, используемый для анализа уровня подготовленности тестируемых, имеет высокую надежность. Это значит, что вопросы или элементы теста последовательно измеряют одно и то же качество или характеристику, и результаты в этом случае считаются надежными для дальнейшего исследования или для принятия решений (Примбетова, 2014: 51).

Как видно из таблицы 2, стандартное отклонение равно 1,09. Стандартное отклонение показывает степень отклонения оценок испытуемых относительно среднего значения. В аспекте тестирования это означает: низкое стандартное отклонение (близкое к 0) показывает, что большинство тестируемых получили оценки, близкие к среднему значению, высокое стандартное отклонение показывает, что оценки тестируемых значительно варьируются. В нашем случае стандартное отклонение 1,09 говорит о том, что есть заметное разнообразие в результатах тестируемых.

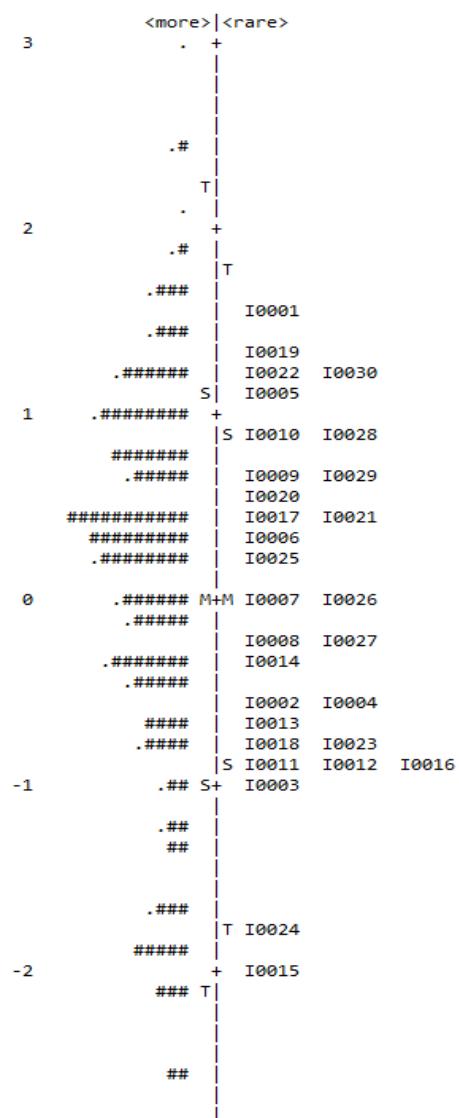
Анализ значений общей (OUTFIT MNSQ) и взвешенной общей статистики согласия (INFIT MNSQ) (таблица 2) позволяет сделать вывод о том, что профиль ответа испытуемых, выполнивших данный вариант теста, находится в хорошем согласии с моделью (Дубина, 2010: 47) [16].

Таблица 2 – Общие характеристики тестируемых

	Total score	Count	Measure	Model error	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	15,3	30,0	0,04	0,44	1,00	-0,06	1,03	-0,03
S.D	5,9	0,0	1,09	0,07	0,18	1,00	0,36	1,02
MAX	28,0	30,0	2,96	1,03	1,55	2,75	2,89	3,26
MIN	1,0	30,0	-3,75	0,40	0,62	-3,02	0,39	-2,75

На рисунке 8 представлена карта переменных, на которой показано распределение испытуемых и заданий относительно друг друга на общей метрической шкале. Слева – шкала логитов (уровень подготовленности испытуе-

мых), справа – задания. Более трудные задания и более сильные испытуемые расположены в верхней части карты, легкие задания и менее подготовленные испытуемые – в нижней части карты.

**Рисунок 8** – Карта переменных

Распределение на карте близко к нормальному. Следовательно, тест ориентирован на данную выборку и соответствует уровню подготовленности испытуемых. Об этом говорит центрированность множества заданий относительно выборки тестируемых. Данный тест оптимален по трудности.

Заключение

На основании проведенного нами исследования можно сделать выводы и привести рекомендации о качестве тестирования, это важно для улучшения образовательных программ и тестовых заданий в дальнейшем.

В результате анализа результатов апробации тестирования по тесту на определение готовности к обучению в рамках комплексного тестирования для поступающих в магистратуру мы пришли к следующим выводам:

1. Тестовые задания составлены корректно и находятся в согласии с моделью.

2. Варианты тестовых заданий составлены должным образом, позволяют выявить уровень подготовленности испытуемых.

3. Профиль ответа испытуемых находится в хорошем согласии с моделью.

4. Отсутствие тестовых заданий с высокой внутренней согласованностью может указывать на то, что возможно тестовые задания требуют доработки для достижения более высокого уровня согласованности.

В качестве рекомендаций можно предложить следующее:

1. Учитывать возможность внесения изменений и улучшения тестовых заданий из категорий

с низкой и хорошей согласованностью с целью повышения их качества.

2. Для обнаружения причин низкой согласованности и для разработки новых тестовых заданий, которые смогут лучше отразить анализируемые системы, провести дополнительные исследования.

3. Для поддержания актуальности и качества тестовых заданий с регулярной периодичностью осуществлять их проверки и пересмотры.

4. Изучить выявленные в процессе анализа сложные тестовые задания, которые могут иметь общие характеристики, такие как, например, необходимость применения сложных аналитических навыков или наличие многослойных заданий. Это поможет определить какие темы требуют более глубокого изучения.

5. При выявлении сложных тестовых заданий, важно оценить, соотносятся ли они с более простыми заданиями и соответствуют ли это общей цели тестирования.

6. Необходимо и важно провести анализ эффективности тестовых заданий при внесении в них изменений, используя различные группы испытуемых.

Таким образом, применение Winsteps 5.8.3 для анализа результатов тестирования предоставляет исследователям и практикам мощные инструменты для глубокого понимания процессов оценивания. Программа позволяет не только анализировать данные, но и визуализировать результаты, что способствует лучшему восприятию информации. Тщательный анализ и интерпретация результатов могут существенно повлиять на принятие решений в области образования и повышения качества оценивания.

Литература

1. Linacre, M. (2011, July). Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice). Retrieved from www.winsteps.com
2. Linacre, M. (2012, June). Winsteps: Tutorial 1. Software operation and basic concepts. Retrieved from <https://www.winsteps.com/tutorials.htm>
3. Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
4. Аванесов, В. (2011). Метрическая система Георга Раша – Rasch Measurement (RM). Статья вторая. *Педагогические измерения*, 3.
5. Белов, Е. Б., Алексеев, М. В., Китаев, Н. П., & Кучумов, А. И. (2019). 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*, 2.
6. Демченок, О. (2010). Математические основы Rash Measurement. *Педагогические измерения*, 1.
7. Дубина, И. Н. (2010). Математико-статистические методы в эмпирических социально-экономических исследованиях. Москва: Финансы и статистика; ИНФРА-М.
8. Карданова, Е. Ю., & Иванова, А. Е. (2023). Психометрические исследования: современные методы и новые возможности для образования. *Вопросы образования*, 3.

9. Ким, В. С. (2008). Обработка результатов тестирования компьютерной программой RUMM-2020. *Педагогические измерения*, 4, 53.
10. Маслак, А. А., & Поздняков, С. А. (2009). Анализ качества тестовых заданий с выбором одного правильного ответа. *Методические рекомендации*. Славянск-на-Кубани: Издательский центр СГПИ.
11. Министерство образования и науки Республики Казахстан. (2018, 31 Октября). Об утверждении Типовых правил приема на обучение в организации образования, реализующие образовательные программы высшего и послевузовского образования. Приказ № 600.
12. Министерство образования и науки Республики Казахстан. (2019, 8 Мая). Об утверждении Правил проведения комплексного тестирования. Приказ № 190.
13. Примбетова, Г. С., & Искакова, А. М. (2014). Психометрический анализ результатов тестирования в рамках внешней оценки учебных достижений в 2014 году. *Качество образования в Евразии*.
14. Смирнова, Г. И., & Смирнов, А. В. (2006). Начало работы с программой Winsteps. *Педагогические измерения*, 3.
15. Фер, Р. М., & Бакарак, В. Р. (2010). Психометрика: Введение (А. С. Науменко, А. Ю. Попова, Trans.; Н. А. Батурина & Е. В. Эйдман, Eds.). Челябинск: Издательский центр ЮУрГУ.
16. Янченко, С. И. (2006). Начало работы в Winsteps с данными статистического пакета SPSS. *Компьютерная программа для педагогических измерений WINSTEPS-MINISTEPS*, 3.

References

- Avanesov, V. (2011). Metricheskaja sistema Georga Rasha – Rasch Measurement (RM). Stat'ja vtoraja [The metric system of Georg Rasch – Rasch Measurement (RM). Part two]. *Pedagogicheskie izmerenija* [Pedagogical Measurements], 3. (in Russian)
- Belov, E. B., Alekseev, M. V., Kitaev, N. P., & Kuchumov, A. I. (2019). Analiz vlijanija pervichnyh ballov zadanij na urovni podgotovlennosti respondentov v dihotomicheskoy modeli Rasha na osnove vzveshennogo metoda maksimal'nogo pravdopodobija [Analysis of the influence of primary task scores on the preparedness levels of respondents in the dichotomous Rasch model based on the weighted maximum likelihood method]. *Statistika i matematicheskie metody v jekonomike* [Statistics and Mathematical Methods in Economics], 2. (in Russian)
- Demenchenok, O. (2010). Matematicheskie osnovy Rash Measurement [Mathematical foundations of Rash Measurement]. *Pedagogicheskie izmerenija* [Pedagogical Measurements], 1. (in Russian)
- Dubina, I. N. (2010). Matematiko-statisticheskie metody v jempiricheskikh social'no-jekonomiceskikh issledovanijah [Mathematical-statistical methods in empirical socio-economic research]. Moskva: Finansy i statistika; INFRA-M. (in Russian)
- Kardanova, E. Ju., & Ivanova, A. E. (2023). Psihometricheskie issledovanija: sovremennye metody i novye vozmozhnosti dlja obrazovanija [Psychometric research: Modern methods and new opportunities for education]. *Voprosy obrazovanija* [Educational Issues], 3. (in Russian)
- Kim, V. S. (2008). Obrabotka rezul'tatov testirovaniya komp'juternoj programmoj RUMM-2020 [Processing test results using the RUMM-2020 computer program]. *Pedagogicheskie izmerenija* [Pedagogical Measurements], 4, 53. (in Russian)
- Maslak, A. A., & Pozdnjakov, S. A. (2009). Analiz kachestva testovyh zadanij s vyborom odnogo pravil'nogo otveta [Analysis of the quality of test tasks with one correct answer]. *Metodicheskie rekomendacii* [Methodical Recommendations]. Slavjansk-na-Kubani: Izdatel'skij centr[Publising house] SGPI. (in Russian)
- Ministerstvo obrazovaniya i nauki Respubliki Kazahstan. (2018, 31 Oktjabrja). Ob utverzhdenii Tipovyh pravil priema na obuchenie v organizacii obrazovanija, realizujushchie obrazovatel'nye programmy vysshego i poslevuzovskogo obrazovanija. Prikaz № 600 [Approval of the typical rules for admission to educational organizations implementing higher and postgraduate educational programs. Order No. 600]. (in Russian)
- Primbetova, G. S., & Iskakova, A. M. (2014). Psihometricheskij analiz rezul'tatov testirovaniya v ramkah vneshej ocenki uchebnyh dostizhenij v 2014 godu [Psychometric analysis of test results within the framework of external evaluation of academic achievements in 2014]. *Kachestvo obrazovanija v Evrazii* [Quality of Education in Eurasia]. (in Russian)
- Smirnova, G. I., & Smirnov, A. V. (2006). Nachalo raboty s programmoj Winsteps [Getting started with the Winsteps program]. *Pedagogicheskie izmerenija* [Pedagogical Measurements], 3. (in Russian)
- Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Fer, R. M., & Bakarak, V. R. (2010). Psihometrika: Vvedenie (A. S. Naumenko, A. Ju. Popova, Trans.; N. A. Baturin & E. V. Jejdman, Eds.) [Psychometrics: An introduction (A. S. Naumenko, A. Ju. Popova, Trans.; N. A. Baturin & E. V. Jejdman, Eds.)]. Cheljabinsk: Izdatel'skij centru JuUrGU. (in Russian)
- Linacre, M. (2011, July). Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice). Retrieved from www.winsteps.com
- Linacre, M. (2012, June). Winsteps: Tutorial 1. Software operation and basic concepts. Retrieved from <https://www.winsteps.com/tutorials.htm>
- Ministerstvo obrazovaniya i nauki Respubliki Kazahstan. (2019, 8 Maja). Ob utverzhdenii Pravil provedenija kompleksnogo testirovaniya. Prikaz № 190 [Approval of the rules for conducting comprehensive testing. Order No. 190]. (in Russian)

Авторлар туралы мәлімет:

Ж.Ж. Бейсенова – магистр, *Фылыми зерттеулер мен психометрика зертханасының жетекші гылыми қызметкері, Үлттық тестілеу орталығы (Астана қ., Қазақстан, e-mail: zhanar-beisenova@mail.ru)*

Л.Е. Шинетова – докторант, *Фылыми зерттеулер мен психометрика зертханасының менгерушісі, Үлттық тестілеу орталығы, Родниковая көшесі (Астана қ., Қазақстан, e-mail: shinetovalyazzat24@gmail.com)*

Н.А. Дидарабекова – филология гылымдарының кандидаты, директордың орынбасары, *Үлттық тестілеу орталығы (Астана қ., Қазақстан, e-mail: nauzhan.didarbekova@mail.ru)*

Г.К. Жабаева – магистр, *Фылыми зерттеулер мен психометрика зертханасының жетекші гылыми қызметкері, Үлттық тестілеу орталығы, Родниковая көшесі (Астана қ., Қазақстан, e-mail: danka1919@mail.ru)*

Сведения об авторах:

Ж.Ж. Бейсенова – магистр, ведущий научный сотрудник лаборатории научных исследований и психометрики, Национальный центр тестирования (г.Астана, Казахстан, e-mail: zhanar-beisenova@mail.ru)

Л.Е. Шинетова – докторант, заведующий лабораторией научных исследований и психометрики, Национальный центр тестирования, (г.Астана, Казахстан, e-mail: shinetovalyazzat24@gmail.com)

Н.А. Дидарабекова – кандидат филологических наук, заместитель директора, Национальный центр тестирования (г.Астана, Казахстан, e-mail: nauzhan.didarbekova@mail.ru)

Г.К. Жабаева – магистр, ведущий научный сотрудник лаборатории научных исследований и психометрики, Национальный центр тестирования (г.Астана, Казахстан, e-mail: danka1919@mail.ru)

Information about authors:

Zh.Zh. Beisenova – Master's degree, *Leading Researcher at the Laboratory of Scientific Research and Psychometrics, National Testing Center (Astana, Kazakhstan, e-mail: zhanar-beisenova@mail.ru)*

L.E. Shinetova – PhD student, *Head of the Laboratory of Scientific Research and Psychometrics, National Testing Center (Astana, Kazakhstan, e-mail: shinetovalyazzat24@gmail.com)*

N.A. Didarbekova – Candidate of Philological Sciences, *Deputy Director, National Testing Center, (Astana, Kazakhstan, e-mail: nauzhan.didarbekova@mail.ru)*

G.K. Zhabayeva – Master's degree, *Leading Researcher at the Laboratory of Scientific Research and Psychometrics, National Testing Center (Astana, Kazakhstan, e-mail: danka1919@mail.ru)*

Поступила 27.01.2025

Принята 01.03.2025