





А.Н. Қуанышбекова^{1*}, А.Е. Қожабаев¹, Р.Н. Зияханова¹,
А.Т. Турганбекова²

¹МНВО РК «Национальный центр тестирования» РГКП на ПХВ

² Специализированный лицей № 82 «Дарын»

010000, г.Астана, Республика Казахстан

*e-mail: quanyszbekova98@mail.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ПСИХОМЕТРИЧЕСКАЯ ВАЛИДАЦИЯ ТЕСТА ЕНТ ПО БИОЛОГИИ: НА ПУТИ К МЕЖДУНАРОДНЫМ СТАНДАРТАМ (НА ПРИМЕРЕ AQA, A-LEVEL)

Аннотация

В статье представлены результаты комплексного исследования качества Единого национального тестирования (ЕНТ) по биологии в Казахстане. Целью исследования является оценить психометрические качества и содержательную валидность теста ЕНТ по биологии на основе эмпирических данных 2025 года, используя комплексный подход, который включает психометрический анализ по модели Раша и внешний содержательный бенчмаркинг с признанным международным эталоном (AQA A-Level Biology). На основе данных 2025 года проведен психометрический анализ с использованием модели Раша, а также содержательный бенчмаркинг с британской программой AQA A-Level Biology. Сравнение проводилось по распределению когнитивных уровней таксономии Блума, целям оценки (Assessment Objectives) и степени интеграции математических навыков. Результаты подтвердили удовлетворительные психометрические характеристики большинства заданий ЕНТ. Однако выявлен структурный дисбаланс: 59,1% теста составляют задания репродуктивного уровня (АО1), тогда как навыки применения знаний и анализа (АО2/АО3) представлены недостаточно. Также отмечено фрагментарное освещение современных направлений, включая генетические технологии. На основе детального анализа полученных данных были сформулированы важные практические рекомендации по комплексной модернизации текущего формата ЕНТ. Авторы предлагают увеличить общую долю тестовых заданий высокого когнитивного уровня, внедрить обязательный базовый математический компонент и значительно усилить механизмы экспертно-академической валидации. Предложенные меры направлены на существенное повышение диагностической ценности государственного экзамена и его полное соответствие современным международным образовательным стандартам, что критически важно для качественного отбора абитуриентов в вузы.

Ключевые слова: ЕНТ, биология, психометрическая валидация, модель Раша, таксономия Блума, AQA A-Level.

Введение

Проблема обеспечения качества национальных систем оценивания в современном образовательном пространстве приобретает характер критического вызова, особенно в государствах, находящихся в стадии активной трансформации образовательных стандартов. В Республике Казахстан Единое национальное тестирование (далее - ЕНТ) выступает ключевым инструментом с высокими ставками (*high-stakes testing*), определяющим не только уровень академических достижений, поступающих в высшие учебные заведения, но и механизмы распределения государственных образовательных грантов.

Процесс трансформации системы высшего образования в Казахстане неразрывно связан с эволюцией национальных инструментов оценки учебных достижений. На современном этапе ЕНТ представляет собой компьютерный экзамен из 120 заданий по пяти предметам с общим временем 240 минут. В последние годы система претерпела модернизацию: был полностью внедрён компьютерный формат (*Computer-Based Testing, CBT*), реализованы системы прокторинга, а также увеличено число попыток сдачи экзамена, что снизило психологическое давление на абитуриентов и повысило объективность итоговых результатов [1].

Настоящее исследование направлено на комплексную оценку качества тестовых материалов ЕНТ по биологии с использованием современной теории тестирования (Item Response Theory, IRT) в рамках модели Раша, а также внешнюю валидацию ЕНТ путём сопоставления с признанным международным образовательным эталоном — британской

программой AQA A-Level Biology. Бенчмаркинг проводится по трем основным направлениям: охват содержания, цели оценки (Assessment Objectives, AO) и требования к математическим навыкам [2].

Проектирование теста и разработка заданий регулируются Государственным общеобязательным стандартом образования (далее - ГОСО), проходя несколько этапов экспертизы и апробации. Профильный предмет «Биология»: содержательная часть и критические вопросы валидности. Биология в системе ЕНТ выступает не только индикатором фактологических знаний, но и инструментом оценки научного мышления абитуриентов. Конструкция теста по биологии включает четыре типа заданий, каждый из которых, согласно спецификациям, нацелен на определенный уровень когнитивной активности по таксономии Блума.

1. Дихотомические задания с выбором одного ответа (25 вопросов): проверяют знание терминологии, базовых концепций и способность к идентификации биологических объектов

2. Контекстные задания (5 вопросов): представляют собой кейсы, где абитуриент должен проанализировать предложенный текст, схему или график и сделать выводы на основе представленных данных. Это прямой аналог заданий АО2 и АО3 в международной практике.

3. Задания на установление соответствия (5 вопросов): оценивают умение систематизировать информацию и выявлять логические взаимосвязи между биологическими процессами и их характеристиками.

4. Задания с множественным выбором (5 вопросов): требуют глубокого понимания предмета, так как поступающий должен выбрать до трех правильных ответов из предложенных шести, что существенно снижает вероятность случайного угадывания.

Цель данного исследования: оценить психометрические качества и содержательную валидность теста ЕНТ по биологии на основе эмпирических данных 2025 года, используя комплексный подход, который включает психометрический анализ по модели Раша и внешний содержательный бенчмаркинг с признанным международным эталоном (AQA A-Level Biology) [3].

Материалы и методы исследования

Дизайн исследования. В работе использовался смешанный исследовательский подход (mixed-methods), предполагающий интеграцию качественных и количественных процедур анализа. Качественная часть была представлена сравнительным контент-анализом, тогда как количественный компонент включал психометрическую обработку данных тестирования. Ключевой объект анализа составили спецификации теста и эмпирические результаты тестов [4].

Выборка и данные для психометрического анализа. Эмпирической основой психометрической части исследования выступили обезличенные (анонимные) данные Национального центра тестирования Республики Казахстан по одному уникальному варианту теста ЕНТ по биологии, предложенному в основной период сдачи в 2025 году. Выборка включила ответы 274 поступающих, сдававших данный конкретный вариант по всей территории Казахстана в единый день тестирования. Данный подход обеспечивает репрезентативность выборки относительно совокупности абитуриентов, сдававших именно этот вариант теста. Однако полученные результаты не могут быть автоматически распространены на все варианты ЕНТ без дополнительных исследований [5-7].

Результаты и их обсуждение

Содержательный бенчмаркинг с программой AQA A-Level. По результатам проведенного исследования внешняя валидацию ЕНТ по биологии, основанная на анализе содержательных разделов, показала, что спецификация AQA A-Level Biology структурирована вокруг восьми тематических блоков, обеспечивающих логическую прогрессию от молекулярного уровня организации живых систем к популяционно экосистемному уровню.

Таблица 1- Сопоставление содержания АQA A-Level Biology и ЕНТ

Тема АQA A-Level	Содержание и фокус	Корреляция со спецификацией ЕНТ
3.1 Biological molecules	Углеводы, липиды, белки, нуклеиновые кислоты, АТФ, вода и ионы.12	Полное соответствие. В ЕНТ эти темы являются базовыми.
3.2 Cells	Структура клеток, методы изучения (микроскопия, фракционирование), клеточный цикл, транспорт.12	Высокая корреляция. В АQA больше внимания уделяется методам исследования.
3.3 Organisms exchange	Поверхности обмена, газообмен, пищеварение, транспортные системы у растений и животных.12	Средняя корреляция. В ЕНТ акцент смещен на анатомию человека.
3.4 Genetic information	ДНК, гены, хромосомы, синтез белка, генетическая вариативность и биоразнообразие.12	Высокая корреляция. Темы ДНК и биоразнообразия идентичны.
3.5 Energy transfers	Фотосинтез, дыхание, энергетические потоки и циклы питательных веществ.12	Высокая корреляция. В АQA более детально рассматриваются биохимические пути.
3.6 Organisms response	Реакции на раздражители, нервная координация, скелетные мышцы, гомеостаз.12	Средняя корреляция. В АQA глубоко изучаются механизмы передачи импульса.
3.7 Genetics & populations	Наследование, популяции, эволюция и экосистемы.12	Высокая корреляция. Популяционная генетика — общая область.
3.8 Control of gene expression	Мутации, регуляция транскрипции/трансляции, генные технологии.12	Низкая корреляция. Данный раздел в ЕНТ представлен фрагментарно.
Примечание: составлено авторами		

Психометрический анализ. Психометрический анализ был выполнен в рамках модели Раша с использованием программного пакета WINSTEPS версии 5.8.3.0. В результате анализа были получены показатели качества измерения, характеризующие как распределение уровней респондентов, так и функционирование заданий теста. Ключевые параметры модели, включая показатели надёжности измерения (Person и Item Reliability), индексы разделимости (Separation Index), а также статистики соответствия модели (Infit), представлены в таблицах 2 и 3.

Таблица 2- Характеристики измерения респондентов

Statistic	Total Score	Count	Measure	S.E.	Infit MNSQ	Infit ZSTD
Mean	28.8	40.0	0.53	0.34	1.01	0.03
SEM	0.5	0.0	0.06	0.00	0.01	0.06
P.SD	9.0	0.0	0.98	0.06	0.21	0.94
S.SD	9.0	0.0	0.98	0.06	0.21	0.95
Max	48.0	40.0	3.51	0.71	1.77	3.27
Min	11.0	40.0	-1.35	0.30	0.58	-1.93
Additional statistics (Person):	Real RMSE	0.36		Person Reliability	0.87 / 0.88	
	Model RMSE	0.34		SE of Person Mean	0.06	
	True SD	0.91 / 0.92		Separation	2.56 / 2.69	
	Cronbach's Alpha (KR-20)	0.87		SEM (raw score)	3.18	
	Raw score-measure correlation	0.99				
Примечание: составлено авторами						

В таблице 2 приведены сводные характеристики измерения 274 респондентов.

1. Способности испытуемых (Measure): Среднее = 0.53 логита, SD = 0.98, диапазон –1.35...3.51. Тест охватывает умеренный спектр способностей.
2. Надёжность и разделимость: Person Reliability = 0.87–0.88, Cronbach’s Alpha = 0.87, Separation = 2.56–2.69. Тест стабилен и различает 3–4 уровня подготовки.
3. Соответствие модели Раша: Infit MNSQ = 1.01, Infit ZSTD = 0.03; отдельные респонденты с Infit MNSQ >1.5 могут иметь непоследовательные ответы.
4. Ошибка измерения: SEM (raw score) = 3.18 балла.

Таблица 3 - Характеристики измерения заданий

Statistic	Total Score	Count	Measure	S.E.	Infit MNSQ	Infit ZSTD
Mean	197.2	274.0	0.00	0.13	1.00	-0.02
SEM	13.4	0.0	0.15	0.00	0.02	0.35
P.SD	83.9	0.0	0.96	0.03	0.14	2.19
S.SD	84.9	0.0	0.97	0.03	0.15	2.22
Max	422.0	274.0	2.02	0.21	1.33	4.97
Min	87.0	274.0	-2.02	0.08	0.73	-2.77
Additional statistics (Item):	Real RMSE	0.14		Separation	6.80 / 6.95	
	Model RMSE	0.14		Item Reliability	0.98	
	True SD	0.95		SE of Item Mean	0.15	
Примечание: составлено авторами						

В таблице 3 представлены сводные показатели измерения 40 заданий теста, включая оценки их трудности, стандартные ошибки и статистики соответствия модели Раша.

1. Трудность заданий (Measure): Среднее = 0.00 логита, диапазон –2.02...2.02. Тест включает лёгкие, средние и сложные задания.
2. Надёжность и разделимость: Item Reliability = 0.98, Separation = 6.80–6.95. Тест надёжен и хорошо дифференцирует задания.
3. Соответствие модели Раша: Средние Infit MNSQ = 1.00, Infit ZSTD = –0.02; отдельные задания имеют |ZSTD| >1.5, что может указывать на несоответствия модели.
4. Ошибка измерения: SEM = 0.15 логита.

Таблица 4 - Параметры трудности и соответствия модели Раша для заданий теста

№	Item	Total Score	Count	Measure (logit)	S.E.	Infit MNSQ	Infit ZSTD	PTMEA Corr.	PTMEA Exp.
1	MC-1	147	274	0.33	0.13	1.14	1.77	0.28	0.40
2	MC-2	213	274	-0.96	0.15	0.99	-0.05	0.34	0.32
3	MC-3	185	274	-0.37	0.14	0.87	-2.77	0.50	0.37
4	MC-4	194	274	-0.55	0.14	0.79	-3.63	0.57	0.35
5	MC-5	142	274	0.41	0.13	1.03	0.69	0.37	0.41
6	MC-6	159	274	0.11	0.13	0.96	-0.82	0.44	0.40
7	MC-7	138	274	0.49	0.13	0.99	-0.19	0.42	0.41
8	MC-8	197	274	-0.61	0.14	0.96	-0.67	0.40	0.35
9	MC-9	87	274	1.43	0.14	1.02	0.34	0.39	0.41
10	MC-10	108	274	1.02	0.14	1.00	0.09	0.40	0.41
11	MC-11	151	274	0.26	0.13	1.26	4.97	0.15	0.40
12	MC-12	124	274	0.73	0.13	1.14	1.71	0.27	0.41
13	MC-13	187	274	-0.41	0.14	0.98	-0.31	0.38	0.36
14	MC-14	243	274	-1.85	0.20	0.97	-0.16	0.29	0.24
15	MC-15	208	274	-0.85	0.15	0.81	-1.77	0.53	0.33
16	MC-16	171	274	-0.10	0.14	1.01	0.16	0.38	0.38
17	MC-17	247	274	-2.02	0.21	0.90	-0.65	0.36	0.23
18	MC-18	93	274	1.31	0.14	1.27	2.02	0.13	0.41
19	MC-19	185	274	-0.37	0.14	0.76	-4.61	0.60	0.37
20	MC-20	156	274	0.17	0.13	1.03	0.68	0.38	0.40
21	MC-21	221	274	-1.16	0.16	0.90	-1.15	0.42	0.30

22	MC-22	133	274	0.57	0.13	1.18	1.58	0.23	0.41
23	MC-23	121	274	0.79	0.13	0.90	-1.94	0.51	0.41
24	MC-24	237	274	-1.63	0.18	0.92	-0.61	0.35	0.26
25	MC-25	219	274	-1.11	0.16	1.01	0.09	0.30	0.31
26	CONT-26	202	274	-0.71	0.15	0.99	-0.11	0.34	0.34
27	CONT-27	209	274	-0.87	0.15	1.00	0.07	0.33	0.33
28	CONT-28	156	274	0.17	0.13	1.10	2.06	0.31	0.40
29	CONT-29	147	274	0.33	0.13	0.91	-1.84	0.49	0.40
30	CONT-30	107	274	1.04	0.14	0.98	-0.29	0.43	0.41
31	SOOTV-31	133	274	1.70	0.10	1.33	3.34	0.32	0.54
32	SOOTV-32	366	274	-0.43	0.10	1.06	0.75	0.42	0.47
33	SOOTV-33	354	274	-0.11	0.09	0.74	-1.98	0.67	0.52
34	SOOTV-34	224	274	0.89	0.09	1.04	0.59	0.53	0.56
35	SOOTV-35	319	274	0.19	0.08	0.93	-1.03	0.60	0.57
36	MC-36	143	274	1.70	0.10	1.28	1.05	0.34	0.53
37	MC-37	93	274	2.02	0.11	1.14	1.19	0.45	0.52
38	MC-38	385	274	-0.73	0.11	0.82	-1.50	0.60	0.45
39	MC-39	422	274	-0.62	0.09	0.73	-1.50	0.61	0.46
40	MC-40	362	274	-0.19	0.09	1.10	1.31	0.45	0.51

Примечание: составлено авторами

В таблице 4 представлены основные психометрические характеристики 40 заданий теста, полученные в рамках модели Раша.

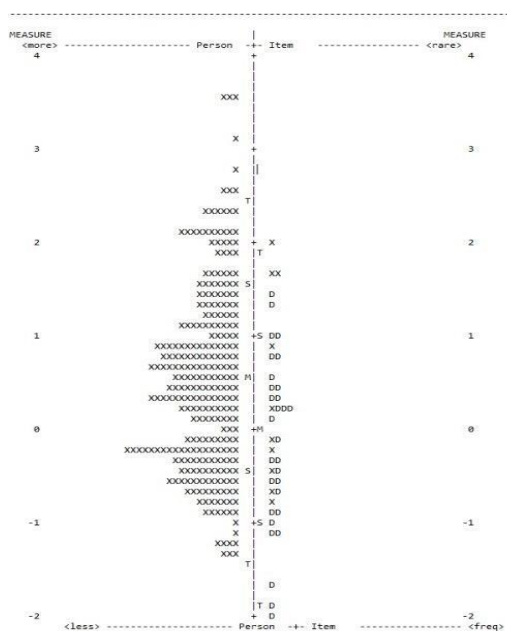


Рисунок 1 - Карта распределения респондентов и заданий (Wright map) в модели Раша (n = 274, 40 заданий)
Примечание: составлено авторами

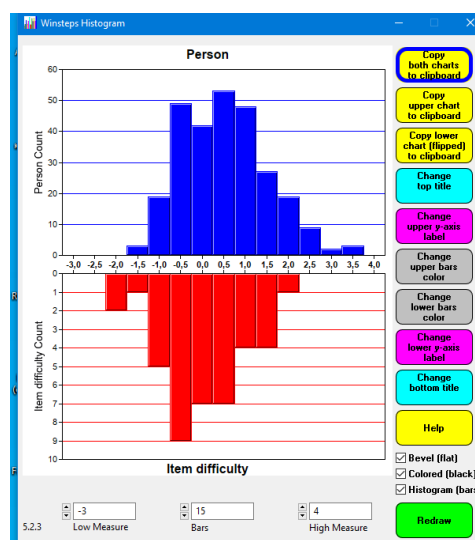


Рисунок 2 - Гистограммы распределения уровней респондентов и трудности заданий в модели Раша
Примечание: составлено авторами

Большинство заданий демонстрируют удовлетворительное соответствие модели Раша: значения статистики Infit MNSQ преимущественно находятся в пределах допустимого интервала (0.7–1.5). Анализ заданий с отклонениями показал:

- MC-11 (Infit MNSQ = 1,26; PTMEA Corr. = 0,15) имеет низкую дискриминативность и нестабильные ответы, что может быть вызвано неоднозначностью формулировки или содержательной ошибкой. Рекомендуется исключить или полностью переработать данное задание.
- MC-18 (Infit MNSQ = 1,27; PTMEA Corr. = 0,13) демонстрирует аналогичные проблемы; требуется экспертиза содержания.
- SOOTV-31 (Infit MNSQ = 1,33; ZSTD = 3,34) — значительное отклонение, указывающее на то, что задание измеряет не тот же конструкт, что и остальные. Подлежит

обязательному исключению. Остальные задания с показателями в пределах 0,7–1,3 функционируют адекватно и могут быть сохранены.

Карта распределения респондентов и заданий иллюстрирует взаимное расположение уровней подготовленности респондентов и параметров трудности заданий в единой логит-шкале модели Раша. Распределение заданий в целом соотносится с распределением респондентов, что указывает на приемлемое согласование структуры теста с характеристиками выборки. Основная часть заданий и респондентов сосредоточена в средней зоне шкалы, где обеспечивается наибольшая информативность измерения. В крайних областях шкалы наблюдается меньшее количество заданий, что может ограничивать точность измерения для отдельных групп респондентов [8-10].

На рисунке 2 представлены гистограммы распределения оценок респондентов и параметров трудности заданий, полученные в рамках модели Раша. Распределение уровней респондентов имеет выраженную концентрацию в средней части шкалы, что указывает на преобладание респондентов со средним уровнем измеряемого признака [11-13]. Распределение трудности заданий также сосредоточено преимущественно в центральном диапазоне логит-шкалы, что свидетельствует о в целом согласованном уровне сложности теста по отношению к характеристикам выборки.

Заключение

Проведённое исследование верификации и психометрической валидации ЕНТ по биологии подтвердило ключевую роль комплексного подхода, объединяющего психометрический анализ и содержательный бенчмаркинг. Результаты в **целом** подтвердили выдвинутые гипотезы, но также выявили системные вызовы и методологические ограничения.

Интерпретация результатов гипотез

Гипотеза 1 (психометрическая): анализ по модели Раша показал удовлетворительные показатели соответствия для большинства заданий (Infit MNSQ в пределах 0.7–1.5). Однако выявлены задания с отклонениями (например, MC-11, SOOTV-31), что указывает на необходимость их доработки — как и предполагалось в гипотезе.

Гипотеза 2 (содержательная): сравнение с AQA A-Level выявило структурный перекокс в сторону заданий уровня AO1 (59% в ЕНТ против 30–35% в AQA), что свидетельствует о недостаточной представленности заданий на анализ и применение.

Гипотеза 3 (интегративная): задания, соответствующие AO2/AO3, действительно показали более высокую дифференцирующую способность, однако их доля в тесте остаётся недостаточной для полноценной диагностики сложных компетенций.

Практические рекомендации и перспективы. С учётом выявленных ограничений предлагаются следующие приоритетные направления развития ЕНТ по биологии:

Пересмотр когнитивного баланса. Учитывая выявленное доминирование заданий репродуктивного уровня (59,1%), целесообразно постепенно увеличивать долю заданий, направленных на применение и анализ (AO2/AO3), до 40–50%. При этом каждый новый тип заданий должен проходить предварительную психометрическую апробацию для подтверждения его соответствия модели Раша и достаточной дискриминативности.

Ограничения исследования: Во-первых, анализ проведён на материалах одного варианта ЕНТ, поэтому выводы о психометрических характеристиках не могут быть автоматически распространены на все варианты теста и для обобщения результатов необходимы повторные исследования на репрезентативной выборке вариантов. Для обобщения результатов необходимо повторение исследования на репрезентативной выборке вариантов. Во-вторых, часть психометрического анализа выполнена на реальных данных, однако некоторые расчёты могли быть подвержены влиянию случайных факторов, связанных с размером выборки. В-третьих, сравнительный бенчмаркинг ограничен одной международной программой (AQA); включение в анализ других систем (например, IB, Advanced Placement) позволило бы получить более полную картину. Несмотря на

указанные ограничения, полученные результаты дают надёжную основу для совершенствования ЕНТ по биологии.

СПИСОК ЛИТЕРАТУРЫ

1. Abdrasilov B., Niyazov T., Baizhaov N., Altybaeva Sh., Iskakova A., & Umerbaeva G. Unified National Testing: History, Experience, Perspective. Astana: National Test Center, - 2024 - 25 p.
2. Алтыбаева Ш, Сагиндиков И., Жабаева Г, & Шақралиева Ш. Валидация содержания теста по информатике в рамках ЕНТ: экспертная оценка и психометрический анализ. Научно-педагогический журнал «Білім-Образование», - 2025. – С. 68–82.
3. Национальный центр тестирования МОН РК. Вопросы тестирования – 2020. - № 1(12). [Электронный ресурс] — URL: <https://testcenter-zhurnal.kz/index.php/sap/article/view/>
4. Портал электронного правительства Республики Казахстан. (б.д.). ЕНТ в Казахстане: подготовка и порядок проведения. Получено 6 января 2026.
5. Информационно-правовая система нормативных правовых актов Республики Казахстан «Әділет». О внесении изменений в приказ Министра просвещения Республики Казахстан от 3 августа 2022 года № 348. (Приказ от 28 декабря 2024 года № 500)
6. Bloom B. S. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company. -1956. – 45 p.
7. AQA AS and A-level Biology specification - 2020. №1 (5). –P. 7401-7402.
8. Didarbekova N., Altybaeva S., Abdrasilov B., & Karkenova A. Advancing digital item development in Asian assessment systems: Insights from Kazakhstan's Unified National Testing. Frontiers in Education, 10, 1654674. - 2025. – 45 p.
9. Landis J. R., & Koch G. G. The measurement of observer agreement for categorical data. Biometrics, - 1977. №33 (1). –P.159–174.
10. OECD Reviews of Evaluation and Assessment in Education. OECD Publishing. [Электронный ресурс] — URL: <https://doi.org/10.1787/22230955> OECD. 2019.
11. Winsteps.com. (б.д.). Rasch analysis. Получено 5 января 2026, [Электронный ресурс] — URL: <https://www.winsteps.com/winman/raschanalysis.htm>
12. Study from Home. (б.д.). What Topics Are Covered in AQA A Level Biology Получено 6 января 2026. [Электронный ресурс] — URL: <https://www.studyfromhome.co.uk/how-it-works/what-topics-are-covered-in-aqa-a-level-biology-7402>
13. Winsteps.com. (б.д.).*Winsteps and Facets: Rasch Analysis + Rasch Measurement Software + 1PL IRT*. 05. 01. 2026 [Электронный ресурс] — URL: <https://www.winsteps.com/>

БИОЛОГИЯДАН ҰЛТТЫҚ БІРЫҢҒАЙ ТЕСТТІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ ЖӘНЕ ПСИХОМЕТРИЯЛЫҚ ВАЛИДАЦИЯСЫ: ХАЛЫҚАРАЛЫҚ СТАНДАРТТАРҒА ҚАРАЙ (AQA, A-LEVEL МЫСАЛ РЕТІНДЕ)

Аңдатпа

Мақалада Қазақстандағы Ұлттық бірыңғай тестілеудің (ҰБТ) биология пәні бойынша тест сапасын бағалауға бағытталған кешенді зерттеу нәтижелері ұсынылған. Зерттеудің мақсаты Раш моделі бойынша психометриялық талдауды және танылған халықаралық эталонмен (AQA A-Level Biology) сыртқы мазмұнды салыстыруды қамтитын кешенді тәсілді пайдалана отырып, 2025 жылғы эмпирикалық деректер негізінде ЕО биология тестінің психометриялық қасиеттері мен мазмұндық жарамдылығын бағалау болып табылады. 2025 жылғы мәліметтер негізінде Раш моделін қолдану арқылы психометриялық талдау, сондай-ақ британдық AQA A-Level Biology бағдарламасымен мазмұндық бенчмаркинг жүргізілді. Салыстыру Блум таксономиясының когнитивті деңгейлерінің бөлінуі, бағалау мақсаттары (Assessment Objectives) және математикалық дағдылардың интеграциялану дәрежесі бойынша жүргізілді. Нәтижелер ҰБТ тапсырмаларының басым бөлігінің қанағаттанарлық психометриялық сипаттамаларын растады. Дегенмен, құрылымдық теңгерімсіздік анықталды: тесттің 59,1%-ын репродуктивті деңгейдегі тапсырмалар (АО1) құрайды, ал білімді қолдану мен талдау дағдылары (АО2/АО3) жеткіліксіз ұсынылған. Сондай-ақ заманауи бағыттардың, соның ішінде генетикалық технологиялардың фрагментарлы түрде қамтылуы атап өтілді. Алынған мәліметтер негізінде ҰБТ форматын жаңғырту бойынша практикалық ұсыныстар тұжырымдалды. Авторлар жоғары когнитивті деңгейдегі тапсырмалардың үлесін арттыруды, міндетті математикалық компонентті енгізуді және сарапшылық-академиялық валидация механизмдерін күшейтуді ұсынады. Ұсынылған шаралар емтиханның диагностикалық құндылығын және оның халықаралық білім беру стандарттарына сәйкестігін арттыруға бағытталған, бұл жоғары оқу орындарына абитуриенттерді сапалы іріктеу үшін өте маңызды.

Негізгі сөздер: ҰБТ, биология, психометриялық валидация, Раш моделі, Блум таксономиясы, АQA, A-Level.

COMPARATIVE ANALYSIS AND PSYCHOMETRIC VALIDATION OF THE UNT BIOLOGY TEST: TOWARDS INTERNATIONAL STANDARDS (THE CASE OF AQA A-LEVEL)

Abstract

The article presents the results of a comprehensive study evaluating the quality of the Unified National Testing (UNT) in Biology in Kazakhstan. The aim of the study is to evaluate the psychometric qualities and substantive validity of the UNT biology test based on empirical data from 2025, using an integrated approach that includes psychometric analysis based on the Rasch model and external meaningful benchmarking with an internationally recognized standard (AQA A-Level Biology). Based on 2025 data, a psychometric analysis was conducted using the Rasch model, alongside a content benchmarking against the British AQA A-Level Biology curriculum. The comparison focused on the distribution of cognitive levels (Bloom's Taxonomy), Assessment Objectives (AOs), and the degree of integration of mathematical skills. The results confirmed satisfactory psychometric characteristics for most UNT items. However, a structural imbalance was identified: 59,1% of the test consists of lower-order thinking skills (AO1 – recall), while application and analysis skills (AO2/AO3) are underrepresented. Furthermore, the study noted fragmented coverage of contemporary topics, such as genetic technologies. Based on the findings, practical recommendations for modernizing the UNT format are formulated. The authors propose increasing the proportion of higher-order thinking tasks, implementing a mandatory mathematical component, and strengthening mechanisms for expert-academic validation. These measures aim to enhance the exam's diagnostic value and ensure its alignment with international educational standards, which is critical for the qualitative selection of university applicants.

Keywords: UNT, Biology, psychometric validation, Rasch model, Bloom's Taxonomy, AQA, A-Level.

REFERENCES

- 1 Abdrasilov B., Niyazov T., Baizhaov N., Altybaeva Sh., Iskakova A., & Umerbaeva G. *Unified National Testing: History, Experience, Perspective*. Astana: National Test Center, - 2024. 25 p. [in English].
- 2 Altybaeva SH, Sagindikov I, ZHabaeva G, & SHakralieva SH. Validaciya sodержaniya testa po informatike v ramkah ENT: ekspertnaya ocenka i psihometricheskij analiz. [Validation of the computer science test content within the UNT: Expert evaluation and psychometric analysis. *Nauchno-pedagogicheskij zhurnal*] «Bilim-Obrazovanie», - 2025. C. 68–82 [in Russian].
- 3 Nacional'nyj centr testirovaniya MON RK. Voprosy testirovaniya [Testing Issues] – 2020. - № 1(12). [Elektronnyj resurs] — Available at:— URL : <https://testcenter-zhurnal.kz/index.php/sap/article/view/> [in Russian].
- 4 Portal elektronnoho pravitel'stva Respubliki Kazahstan. (b.d.). ENT v Kazahstane: podgotovka i poryadok provedeniya. [UNT in Kazakhstan: Preparation and procedure]Polucheno 6 yanvarya 2026. [in Russian].
- 5 Informacionno-pravovaya sistema normativnyh pravovyh aktov Respubliki Kazahstan «Әдилет». О внесении изменений в приказ Министра просвещения Республики Казахстан от 3 августа 2022 года № 348. [Information and Legal System of Normative Legal Acts of the Republic of Kazakhstan "Adilet"] (Приказ от 28 декабря 2024 года № 500). [in Russian].
- 6 Bloom B. S. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company. - 1956. 45 p. [in English].
- 7 AQA AS and A-level Biology specification - 2020. №1 (5). P. 7401-7402.[in English].
- 8 Didarbekova N., Altybaeva S., Abdrasilov B., & Karkenova A. Advancing digital item development in Asian assessment systems: Insights from Kazakhstan's Unified National Testing. *Frontiers in Education*, 10, 1654674. Available at:— URL: <https://doi.org/10.3389/educ.1654674> - 2025. 45 p. [in English].
- 9 Landis J. R., & Koch G. G. The measurement of observer agreement for categorical data. *Biometrics*, -1977. №33 (1), P.159–174 [in English].
- 10 OECD Reviews of Evaluation and Assessment in Education. OECD Publishing. Available at:— URL: <https://doi.org/10.1787/22230955> OECD. - 2019 [in English].
- 11 Winsteps.com. (n.d.). *Rasch analysis*. Retrieved 05.01. 2026, from Available at:— URL: <https://www.winsteps.com/winman/raschanalysis.htm> [in English].
- 12 Study from Home. (n.d.). *What Topics Are Covered in AQA A Level Biology (7402)?* Retrieved January 6, 2026, from Available at:— URL: <https://www.studyfromhome.co.uk/how-it-works/what-topics-are-covered-in-aqa-a-level-biology-7402> [in English]
- 13 Winsteps.com. (n.d.). *Winsteps and Facets: Rasch Analysis + Rasch Measurement Software + 1PL IRT*. Retrieved January 5, - 2026, from Available at:— URL: <https://www.winsteps.com/> [in English].

Information about authors:

Arailym Kuanyshbekova – **corresponding author**, Lead Expert of the National Learning Center, Astana, Republic of Kazakhstan

E-mail: quanyshbekova98@mail.ru

ORCID: <https://orcid.org/0009-0000-3780-6866>

Almat Kozhabayev - Chief Expert of the National Testing Center, Astana, Republic of Kazakhstan

E-mail: almat-kozhabaev@mail.ru

ORCID: <https://orcid.org/0009-0008-2637-6168>

Раушан Зияханова - Chief Expert of the National Testing Center, Astana, Republic of Kazakhstan

E-mail: ziahanova.raushan@yandex.ru

ORCID: <https://orcid.org/0009-0001-9392-8363>

Aigul Turganbekova - Biology Teacher at the "Specialized Lyceum No. 82 'Daryn'" of the Akimat of Astana, Republic of Kazakhstan

E-mail: aigul_turganbekova@mail.ru

ORCID: <https://orcid.org/0009-0002-2415-1754>

Информация об авторах:

Арайлым Қуанышбекова – **основной автор**, ведущий эксперт Национального Центра Тестирования, г. Астана, Республика Казахстан.

E-mail: quanyshbekova98@mail.ru

ORCID: <https://orcid.org/0009-0000-3780-6866>

Алмат Қожабаев - главный эксперт Национального Центра Тестирования, г. Астана, Республика Казахстан.

E-mail: almat-kozhabaev@mail.ru

ORCID: <https://orcid.org/0009-0008-2637-6168>

Раушан Зияханова - главный эксперт Национального центра тестирования, г.Астана, Республика Казахстан.

E-mail: ziahanova.raushan@yandex.ru

ORCID: <https://orcid.org/0009-0001-9392-8363>

Айгул Турганбекова - Учитель биологии КГУ «Специализированный лицей № 82 «Дарын» акимата города Астаны, Республика Казахстан.

E-mail: aigul_turganbekova@mail.ru

ORCID: <https://orcid.org/0009-0002-2415-1754>

Авторлар туралы ақпарат:

Арайлым Қуанышбекова – **негізгі автор**, Ұлттық тестілеу орталығының жетекші сарапшысы, Астана қ., Қазақстан Республикасы.

E-mail: quanyshbekova98@mail.ru

ORCID: <https://orcid.org/0009-0000-3780-6866>

Алмат Қожабаев - Ұлттық тестілеу орталығының бас сарапшысы, Астана қ., Қазақстан Республикасы.

E-mail: almat-kozhabaev@mail.ru

ORCID: <https://orcid.org/0009-0008-2637-6168>

Раушан Зияханова - Ұлттық тестілеу орталығының бас сарапшысы, Астана қ., Қазақстан Республикасы.

E-mail: ziahanova.raushan@yandex.ru

ORCID: <https://orcid.org/0009-0001-9392-8363>

Айгул Турганбекова – Астана қаласы әкімдігінің «№ 82 «Дарын» мамандандырылған лицейі» КММ-нің биология пәні мұғалімі, Астана қ., Қазақстан Республикасы.

E-mail: aigul_turganbekova@mail.ru

ORCID: <https://orcid.org/0009-0002-2415-1754>