

## Валидация содержания теста по информатике в рамках ЕНТ: экспертная оценка и психометрический анализ

Ш.Б. Алтыбаева<sup>1\*</sup>, И.У. Сагиндиков<sup>2</sup>, Г.К. Жабаева<sup>3</sup>, Ш.С. Шакралиева<sup>4</sup>

<sup>1,2,3,4</sup>Национальный центр тестирования

г. Астана, Республика Казахстан

\*e-mail: shugyla.altybayeva@gmail.com



**Аннотация.** Актуальность исследования значима, поскольку Единое национальное тестирование в Казахстане является экзаменом с высокой ставкой, результаты которого оказывают существенное влияние на образовательные траектории выпускников. Это обуславливает необходимость обеспечения валидности и надежности тестовых материалов, особенно по динамично развивающимся предметам, таким как информатика. Дополнительное значение имеет независимая верификация содержания теста в связи с переходом информатики в категорию профильных предметов с 2022 года. Исследование проводилось с целью проведения комплексного анализа содержательной валидности теста по информатике ЕНТ через экспертизу соответствия спецификации требованиям Государственно-го общеобязательного стандарта образования (ГОСО) и психометрическую оценку качества заданий. В исследовании использовались смешанные методы, включавшие контент-анализ 54 учебных целей, независимую экспертную оценку с участием шести экспертов, качественный анализ методом тематического кодирования и количественную оценку по модели Раша на выборке из 116 учащихся. Результаты исследования показали, что в структуре учебных целей преобладали задания на понимание (25%) и применение (30%), при этом наблюдался недостаток заданий высоких когнитивных уровней, таких как анализ (15%) и оценка (10%). Экспертный анализ подтвердил репрезентативность содержания с коэффициентом согласия Каппа 0,80, однако выявил необходимость включения современных тем, таких как 3D-моделирование и информационная безопасность. Психометрический анализ продемонстрировал высокую надежность теста с  $\alpha = 0,94$  и соответствие большинства заданий модели Раша, при этом были идентифицированы два проблемных задания с Outfit MNSQ выше 1,5. По результатам исследования разработаны рекомендации по модернизации теста: уточнены формулировки учебных целей, добавлены новые тематические блоки и скорректированы проблемные задания. Комплексный подход к валидации доказал свою эффективность для повышения качества измерительных материалов. Результаты могут быть применены в практике разработки и обновления тестовых спецификаций.



**Ключевые слова:** валидность, тест по информатике, Единое национальное тестирование, спецификация теста, таксономия Блума, модель Раша, психометрический анализ.



**Қалай дәйексөз алуға болады / Как цитировать / How to cite:** Алтыбаева, Ш.Б., Сагиндиков, И.У., Жабаева, Г.К., Шакралиева, Ш.С. Валидация содержания теста по информатике в рамках ЕНТ: экспертная оценка и психометрический анализ [Текст] // Научно-педагогический журнал «Білім-Образование». – Астана: НАО имени И. Алтынсарина, 2025. – №3. – С. 68-82.

## Введение

Результаты вступительных экзаменов играют решающую роль в определении образовательного и профессионального будущего учащихся, а сами системы отбора в вузы нередко становятся предметом пристального внимания со стороны как общества, так и научного сообщества. В Казахстане отбор абитуриентов в высшие учебные заведения и распределение образовательных грантов осуществляется преимущественно на основе Единого национального тестирования (далее – ЕНТ). Высокие ставки, связанные с результатами ЕНТ, требуют, чтобы каждый компонент теста был тщательно разработан и валиден, особенно по таким предметам, как информатика, которые характеризуются стремительным развитием и технологической сложностью.

В отличие от других стран, где при поступлении в университеты учитываются дополнительные критерии — такие как школьные оценки, мотивационные письма и внеучебные достижения [1], в Казахстане результаты ЕНТ часто являются единственным и определяющим критерием отбора [2]. Это придаёт исключительное значение качеству и валидности самих тестовых заданий.

Среди ключевых характеристик качества теста валидность занимает центральное место. Согласно современной теории измерений, валидность определяется как степень, в которой интерпретации результатов теста и принимаемые на их основе решения являются обоснованными и соответствуют заявленным целям [3]. Одним

из её компонентов является содержательная валидность, отражающая степень соответствия содержания теста той учебной дисциплине, которую он призван оценивать. Как подчеркивала А. Анастази, содержательная валидность особенно важна для образовательных тестов, так как она позволяет судить, насколько задания репрезентируют весь объём содержания, предусмотренного учебной программой [4].

При этом, наряду с внутренними процедурами валидации, проводимыми разработчиками теста, существует устойчивый запрос со стороны академического и профессионального сообщества на открытые, публичные исследования, верифицирующие содержание и валидность измерительных материалов. Подобные исследования повышают прозрачность «высокоставкового тестирования» и укрепляют доверие к его результатам. Актуальность данного исследования, целью которого является восполнение этого пробела, усиливается ещё и тем, что в 2022 году информатика была включена в перечень профильных предметов ЕНТ. Это обусловило необходимость более глубокой и системной экспертизы содержания теста, включая его специфику и структуру заданий.

Таким образом, целью настоящей статьи является экспертный анализ содержательной валидности теста по информатике в формате ЕНТ с фокусом на соответствие структуры и содержания заданий заявленным целям обучения и требованиям ГОСО. В рамках исследования также проводится выявление сильных и про-

блемных элементов спецификации теста и первичная психометрическая проверка качества заданий с использованием модели Раша для сбора предварительных эмпирических данных об их соответствии модели измерения.

**Теоретические основы:** Содержательная валидность и спецификация теста. Содержательная валидность играет ключевую роль в обеспечении достоверности результатов стандартизированного тестирования. Согласно подходу Мессика [3], она отражает степень, в которой содержание теста соответствует заявленным учебным целям. В контексте Казахстана это означает соответствие тестовых заданий требованиям ГОСО. При проектировании ЕНТ по информатике особое внимание уделяется спецификации теста, выступающей в качестве основного инструмента обеспечения содержательной валидности. Разработка тестовых заданий представляет собой строго формализованный процесс, регламентируемый, техническим заданием (далее – ТЗ) и утверждённой спецификацией теста. Последняя содержит точные инструкции относительно содержания, форматов заданий, проверяемых когнитивных умений, критерии оценки и временных параметров. Эти документы обеспечивают сопоставимость, прозрачность и стандартизацию оценивания.

Тест по информатике ЕНТ включает 40 заданий различных форматов. Из них (1) 25 заданий с выбором одного ответа и (2) 5 заданий с множественным выбором (от 1 до 3 правильных ответов) проверяют базовые знания, понимание материала и навыки применения, также проверяют умение анализировать информацию; (3) 5 контекстных заданий оценивают способность анализировать тексты, схемы и графики; (4) 5 заданий на установление соответствия оценивают системное мышление. Структура теста соответствует иерархии Блума, отражая уровневую сложность и разнообразие форматов, что повышает его валидность. Процесс утверждения спецификации сопровождается независимой экспертизой профильными специалистами и согласованием с Национальной академией

образования им. И. Алтынсарина (далее – НАО), что обеспечивает соответствие современным требованиям и учебным программам. При разработке теста учитываются дидактические принципы и изменения в предметной области. С учётом быстрого развития информатики спецификация регулярно обновляется для поддержания актуальности и содержательной валидности.

В целом, соблюдение всех перечисленных процедур и документов позволяет обеспечить высокий уровень содержательной валидности теста ЕНТ по информатике, делая его надёжным инструментом для оценки учебных достижений выпускников.

## Материалы и методы

Настоящее исследование выполнено в рамках подхода кейс-исследования с вложенным дизайном с использованием конвергентного параллельного дизайна, предполагающего сочетание качественных и количественных методов анализа для комплексной оценки содержательной валидности теста по информатике в формате ЕНТ.

Исследование проводилось в три этапа:

1. Документальный анализ, включавший изучение спецификации теста и ТЗ, в котором цели обучения сформулированы в соответствии с ГОСО.
2. Таксономическая классификация 54 учебных целей из ТЗ по уровням когнитивной сложности согласно таксономии Блума.
3. Экспертная оценка, организованная в формате круглого стола с участием шести специалистов, обладающих опытом в области преподавания, разработки учебных программ и оценки образовательных достижений.

Дополнительно был проведён эмпирический анализ теста с использованием модели Раша, реализованной в программе Winsteps (версия 5.8.3.0). В выборку ( $N=116$ )

вошли учащиеся из пяти регионов Казахстана; её демографическое распределение соответствовало характеристикам генеральной совокупности (около 5000 человек, участвовавших в апробации тестовых заданий).

Анализ включал оценку:

- Сложности заданий в логитах (одно логит-значение  $\approx$  изменение вероятности правильного ответа на 63%);
- Соответствия модели: INFIT и OUTFIT MNSQ (0,5–1,5), ZSTD (-2 до +2);
- Надёжности (реальной и модельной) оценивания;
- Коэффициентов разделения по уровню сложности заданий и подготовки участников;
- Точечно-бисериальных корреляций и сравнения ожидаемых и наблюдаемых процентов правильных ответов.

Модель Раша оценивает точность заданий и предсказуемость ответов в зависимости от уровня подготовки. Такой подход обеспечивает количественное подтверждение согласованности эмпирических данных с модельными ожиданиями и позволяет выявить задания, нуждающиеся в содержательной доработке.

Качественный анализ включал тематическую обработку экспертных высказываний по методологии Braun & Clarke с использованием открытого кодирования и категоризации [5]. Надёжность обеспечивалась методом независимого двойного кодирования и согласования расхождений. Также применялась триангуляция данных, включающая сопоставление результатов документального анализа, классификации целей и экспертных заключений.

Этические аспекты были соблюдены в соответствии с Этическим кодексом исследователей образования Казахстана: все участники были информированы о целях исследования, предоставили добровольное согласие, анонимность и конфиденциальность гарантированы.

## **Результаты и обсуждение**

Целью настоящего исследования являлась содержательная верификация спецификации теста по информатике в контексте анализа ТЗ и требованиям ГОСО. Работа была организована на базе Национального центра тестирования (далее – Центр) в формате экспертного круглого стола. В исследовании приняли участие шесть экспертов, обладающих значительным опытом в преподавании, разработке учебных материалов и экспертизе оценочных средств.

Процедура включала два этапа:

1. Индивидуальный анализ спецификации теста, где каждый эксперт давал независимую оценку.
2. Групповое обсуждение, модератором которого выступал эксперт Центр. Обсуждение было записано и подвергнуто тематическому качественному анализу.

Релевантность и репрезентативность содержания. Спецификация теста охватывает ключевые содержательные разделы курса информатики: компьютерные системы; информационные процессы; компьютерное мышление; аппаратное и программное обеспечение; информационные процессы и системы; создание и преобразование информационных объектов. Эти разделы отражают основные положения и служат основой для построения заданий, соответствующих требованиям предметной области.

**Таблица 1. Профиль эксперта**

Экспертный код	Позиция	Стаж
Эксперт 1	Учитель	28 лет
Эксперт 2	Учитель	17 лет
Эксперт 3	Учитель	10 лет
Эксперт 4	Учитель	5 лет
Эксперт 5	Эксперт вуза	12 лет
Эксперт 6	Эксперт Центра	8 лет

Профиль экспертов представлен в таблице 1. Участники подтвердили, что содержание спецификации является репрезентативным по отношению к учебной программе и ожиданиям вузов для ИКТ-специальностей, в рамках которых тест по информатике является профильным.

Количественный анализ целей обучения. Был проведён контент-анализ 54 учебных целей, включённых в ТЗ, с классификацией по таксономии Блума. Полученные данные показали акцент на уровнях «Понимание» (25 %) и «Применение» (30 %), при умеренном присутствии целей уровня «Анализ» (15 %) и незначительной доле уровня «Оценка» (10 %). Цели уровней «Создание» (15 %) и «Запоминание» (5 %) представлены в меньшей степени. Контент-анализ показал, что в структуре

учебных целей преобладают задания на понимание и применение, что отражает специфику предмета. В рамках исследования были уточнены формулировки целей, акцентированы ИКТ-компетенции (интерпретация, аргументация, принятие решений) и добавлены цели, связанные с анализом и оценкой. Это стало основанием для рекомендаций по расширению когнитивного охвата теста. Классификация целей осуществлялась на основе глаголов, соответствующих уровням таксономии Блума, при соблюдении требований ГОСО [6, 7].

Качественный анализ высказываний экспертов. Речевые суждения участников были расшифрованы и проанализированы по методу тематического кодирования. На основе этого были выделены четыре ключевые темы.

**Таблица 2. Темы, выявленные в ходе анализа экспертных высказываний**

Тема	Описание	Частота (из 6)	Пример высказывания
Релевантность содержания	Содержание соответствует ГОСО и школьной программе	5	Содержание охватывает ключевые темы курса
Недостаточная представленность	Не отражены современные темы (3D, ИИ и др.)	3	Следует добавить актуальные разделы, соответствующие трендам
Когнитивная сложность	Дисбаланс между уровнями: мало анализа и оценки	4	Недостаточно заданий на сравнение и обоснование решений

Предложения по улучшению	Дополнить темы и улучшить формулировки целей	4	Цели в блоке «компьютерное мышление» требуют переформулировки
Добавление целей обучения	Отсутствие перечня целей обучения в спецификации теста	6	Следует включить учебные цели в спецификацию для большей прозрачности

Таким образом, тематический анализ подтвердил содержательную релевантность спецификации, но также указал на конкретные области для улучшения, особенно в части когнитивного разнообразия и обновления тематики.

Оценка согласованности экспертовских суждений. Для количественной оценки согласованности был применён коэффициент Каппа Коэна, рассчитанный на основе классификации 54 целей по категориям «Полное соответствие», «Частичное соответствие», «Несоответствие». Средний коэффициент составил 0.80 (диапазон: 0.65–0.81), что указывает на высокий уровень согласия между экспертами [8].

Перспективы и обновление спецификации. По результатам исследования были внесены изменения в спецификацию, вступающие в силу с 2026 года. Добавлены три новых содержательных блока: Современные тенденции развития информационных технологий, IT Startup (ай-ти стартап), 3D моделирования, Информационная безопасность.

Эти темы направлены на развитие умений анализа, аргументации и оценки, соответствующих высоким уровням таксономии Блума. Эксперты положительно оценили эти изменения:

«Эти темы позволяют выйти за рамки базовых умений, развивая способность интерпретировать, рассуждать и принимать решения — ключевые компетенции в ИКТ-среде».

Один из участников подчеркнул:

«Актуализация содержания теста за счёт новых тематических акцентов позволяет

усилить его потенциал в оценке умений анализа и оценки, которые ранее в teste по информатике не хватало».

Другой эксперт отметил расширение диапазона проверяемых когнитивных умений:

«Обновлённая спецификация значительно расширяет спектр когнитивных операций, особенно за счёт новых тем, требующих не только применения, но и аргументации, интерпретации и сравнения».

Также было подчёркнуто значение приведения содержания в соответствие с современными технологическими реалиями:

«Фокус на темах, связанных с цифровыми технологиями и актуальными ИТ-направлениями, отвечает современным требованиям и позволяет тесту быть не только измерительным инструментом, но и механизмом развития предметных и метапредметных компетенций».

Результаты исследования подтверждают высокую содержательную валидность действующей спецификации теста по информатике. Выявлены направления для улучшения – расширение когнитивного спектра, актуализация тем и уточнение формулировок. Обновление спецификации повысило её соответствие ГОСО, программе 11 класса и профилю ИКТ, а также усилило потенциал теста для оценки готовности к обучению в вузе.

Эмпирическая проверка конструктной валидности и внутренней согласованности с использованием модели Раша. Помимо экспертовых оценок содержа-

тельной валидности, валидация теста требует эмпирического подтверждения его внутренних психометрических свойств и конструктной валидности. Одним из эффективных инструментов для оценки соответствия заданий заявленному латентному конструкту (в данном случае – уровню подготовки по информатике, как он определен в спецификации) и их внутренней согласованности является модель Раша. Она позволяет количественно оценить техническое качество заданий и то, насколько согласованно они работают в рамках единой шкалы измерения. Согласно Смиту [9], задания, хорошо соответствующие модели Раша, должны располагать испытуемых в идентичном порядке сложности, то есть измерять один и тот же конструкт. Нарушение этой упорядо-

ченности указывает на потенциальное смещение содержания, многомерность конструкта или на наличие заданий, выходящих за рамки целевого конструкта, что может косвенно затрагивать и вопросы содержательной валидности.

В рамках настоящего исследования была проведена первичная эмпирическая калибровка тестовых заданий по информатике с использованием программного обеспечения Winsteps. В представленных сводных психометрических характеристиках теста (Таблица 3) средняя стандартная ошибка оценки сложности заданий составляет 0,15 логит, что указывает на приемлемую точность измерения и достоверность полученных оценок в пределах данной выборки.

**Таблица 3. Сводные психометрические характеристики теста  
(Сводка по 40 измеренным (неэкстремальным) элементам)**

	Суммарный балл	Количество измерений	Измерение	Стандартная ошибка	Показатель соответствия INFIT		Показатель внешнего соответствия OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Среднее значение	92.4	116.0	0.00	0.22	0.98	-0.08	1.21	0.25
Стандартная ошибка среднего	6.2	0.0	0.15	0.01	0.03	0.24	0.23	0.30
Стандартное отклонение популяции	38.8	0.0	0.95	0.05	0.18	1.49		1.86
Стандартное отклонение выборки	39.3	0.0	0.97	0.05	0.18	1.51	1.45	1.89
Максимальное значение	207.0	116.0	1.97	0.41	1.42	3.43	9.90	5.18
Максимальное значение	38.0	116.0	-2.44	0.14	0.64	-2.30	0.45	-2.25

Дополнительные характеристики							
Реальная среднеквадратическая ошибка	0,23	Истинное стандартное отклонение	0,93	Коэффициент разделения	4,10	Коэффициент надежности элемента	0,94
Модельная среднеквадратическая ошибка измерения	0,22	Истинное стандартное отклонение	0,93	Коэффициент разделения	4,19	Коэффициент надежности элемента	0,95
Стандартная ошибка среднего значения элемента			0,15				

Данные, представленные в таблице 3, свидетельствуют о том, что значения нормализованных остатков (ZSTD) находятся в допустимом диапазоне от -2 до +2, что свидетельствует об отсутствии значимых аномалий в распределении ответов [10].

Показатели INFIT MNSQ (0,98) и OUTFIT MNSQ (1,21) уложились в приемлемый интервал (0,5–1,5), подтверждая соответствие эмпирических данных модели Раша и высокую степень согласованности между предсказанными и фактическими ответами участников [11].

Дополнительно были рассчитаны **коэффициенты надежности**: реальный – 0,94, модельный – 0,95. Эти значения указывают на высокую стабильность измерения. **Коэффициенты разделения** (4,10 и 4,19) подтверждают способность теста дифференцировать задания по уровню трудности и участников – по уровню подготовки.

Также наблюдалась отрицательная корреляция (-0,44) между сырым баллом и мерой сложности задания, что, вероятно, связано с групповыми особенностями структуры теста. Данные выводы были ис-

пользованы для дополнительной проверки отдельных элементов теста на предмет необходимости содержательной доработки.

В ходе исследования, проведённого с использованием модели Раша, были оценены ключевые параметры тестовых заданий: уровень сложности (в логитах), показатели соответствия модели, точечно-бисериальная корреляция, а также наблюдаемые и ожидаемые проценты правильных ответов (Таблица 4).

Самым сложным оказалось задание СО-ОТВ-32 (1,97 логит, 51.7% верных ответов), самым лёгким – МС-12 (-2,44 логит, 94.0% верных ответов).

В представленном анализе для разграничения параметров оценивания по формам заданий использованы следующие обозначения: СООТВ. – задания на установление соответствия; КОНТ. – задания, построенные на основе контекста; МС – задания с множественным выбором из предложенного множества вариантов, включая 25 заданий с выбором одного правильного ответа из четырёх.

Таблица 4. Сводные психометрические характеристики по типам заданий

Тип задания	Примеры заданий	Диапазон сложности (логит)	% верных ответов	Корреляция (набл.)	Проблемные задания (Outfit MNSQ > 1.5)	Выводы
MC (множественный выбор)	MC-1 - MC-26	от -2.44 до +1.68	62.9% – 94.0%	0.27 – 0.62	—	Стабильные показатели, задания согласованы с моделью. Хорошо различают уровень подготовки.
КОНТ (контекстные)	КОНТ-26 — КОНТ-30	от -1.05 до +0.54	67.2% – 84.5%	0.37 – 0.58	—	Подтверждена надёжность. Подходят для использования в teste.
COOTB (соответствия) — высокая нагрузка Outfit	COOTB-31, COOTB-32	-1.04, +1.97	51.7% – 85.3%	0.37, 0.37	COOTB-31: 1.90 COOTB-32: 1.92	Наблюдается несогласование модели Раша, возможны структурные проблемы. Требуется ревизия.
COOTB (содержательно сложные)	CO-OTB-33, COOTB-35	+1.11, +1.39	48.3% – 56.9%	0.42, 0.47	—	Низкий % выполнения при приемлемой корреляции. Желательна содержательная доработка.

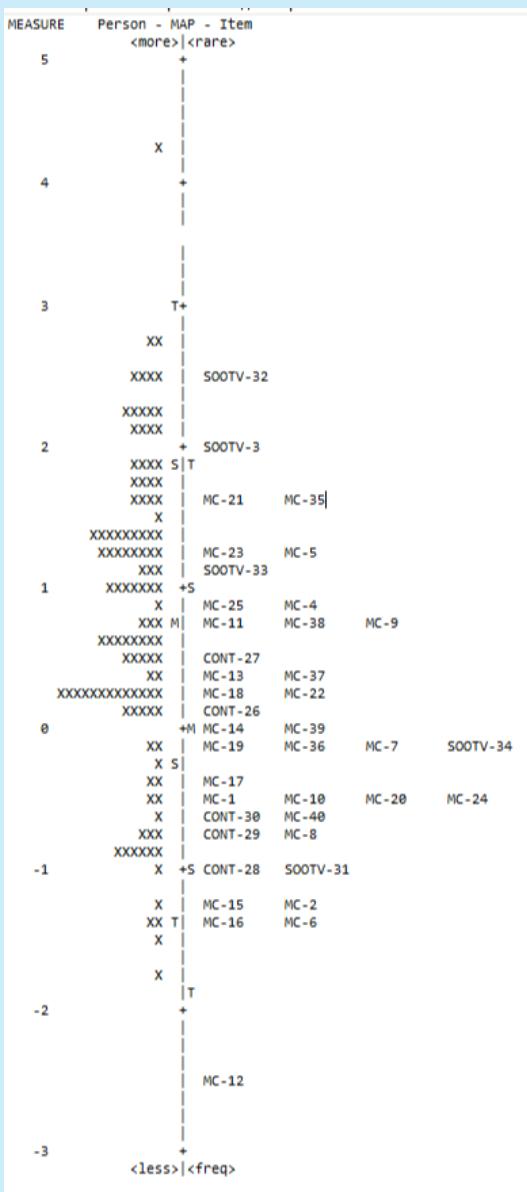
Согласно 2-таблице, большинство заданий соответствуют модели Раша и эффективно различают уровень подготовки (-2.44 до +1.97 логита). Отдельные задания (COOTB-31, COOTB-32) нуждаются в доработке.

Карта переменных (Рисунок 1) визуализирует распределение способностей участников и уровней сложности заданий по шкале логитов –единиц измерения в модели Раша.

- Участники (ось Y):** Способности от -1.5 до +4.5 логитов. Пик в 1.5–2.0 л. (средний/выше среднего). Среднее ~0.8–1.0 л.
- Задания (ось X):** Сложность от -3 до +5 логитов. Лучшее соответствие участникам в зоне 0–2 л.

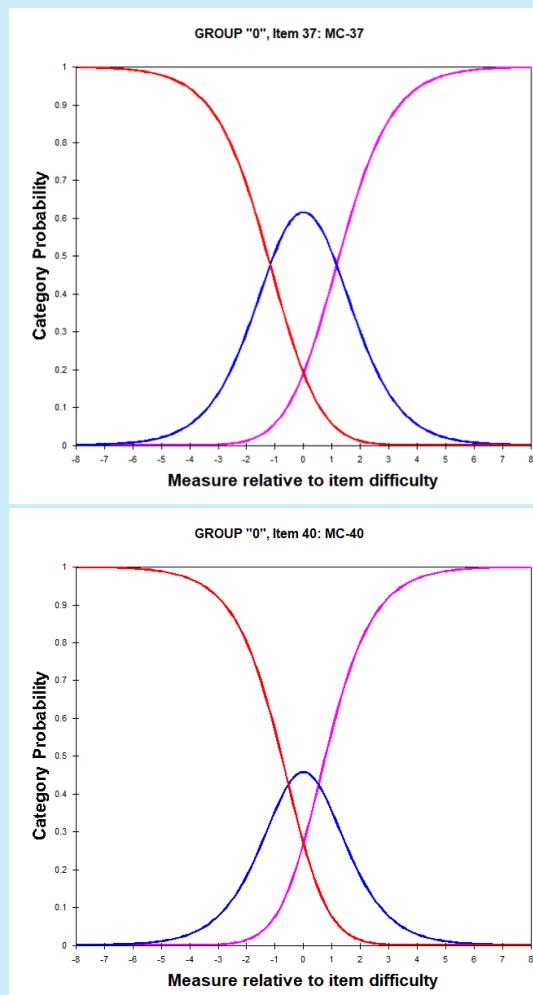
- Оптимальная зона:** Область между пунктами с максимальной точностью измерения.
- Проблема:** Дефицит заданий в диапазоне 2.5–4.0 л. для высокоподготовленных участников → снижает точность их оценки.

В связи с этим в спецификацию теста 2026 года были включены задания, ориентированные на более высокие уровни когнитивной сложности, такие как применение, анализ и оценивание. Эти меры направлены на расширение диапазона сложности заданий, частичную оценку элементов аналитического мышления, а также на повышение валидности теста для всех категорий участников.



**Рисунок 1. Карта переменных: Испытуемые – Задание**

В рамках исследования проведён анализ характеристических кривых заданий (ICC/CCC) по модели Раша. Ось X отражает уровень подготовки (логиты), ось Y — вероятность выбора ответа или категории.



**Рисунок 2. CCC-график MC-37**  
**Рисунок 3. CCC-график MC-40**

Дихотомические задания (один правильный ответ) продемонстрировали типичные S-образные кривые, что свидетельствует о соответствии модели и хорошей дифференцирующей способности.

Политомические задания (несколько правильных ответов) показали как корректно работающие задания (MC-37, MC-40), так и отклоняющиеся от модели (SOOTV-33, MC-38).

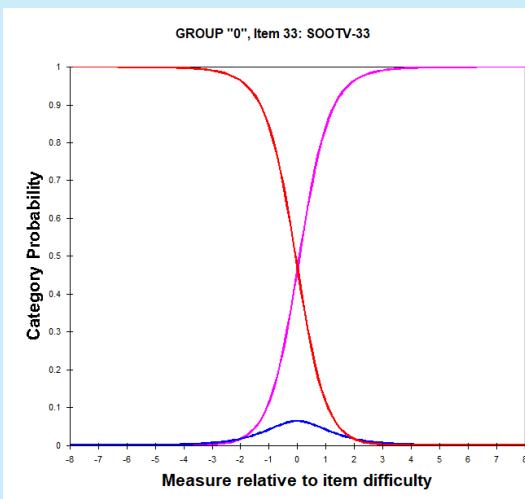


Рисунок 4. CCC-график MC-33

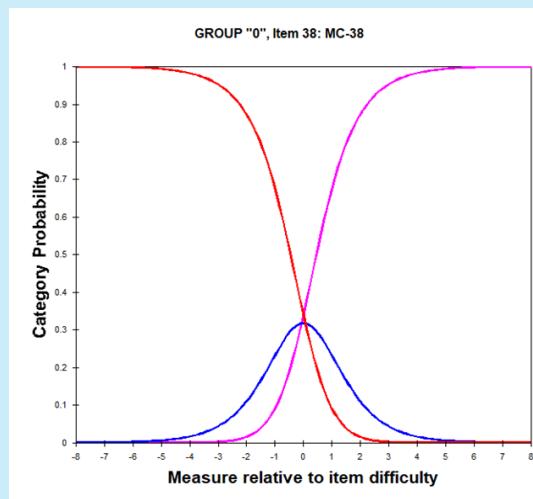


Рисунок 5. CCC-график MC-38

В заданиях СООТВ-33 и МС-38 промежуточные категории (синие кривые) демонстрируют низкую максимальную вероятность выбора (менее 0,1 и около 0,3 соответственно) и узкий диапазон функционирования, что указывает на их слабую работу. Низкий процент успешности по СООТВ-33 (48,3%) свидетельствует о высокой сложности или затруднениях при выборе высоких категорий. Это снижает дифференцирующую способность задания. Аналогичные проблемы выявлены и в задании МС-38. Учитывая выявленные недостатки, данные задания были вынесены на дополнительную экспертизу оценку и подвергнуты содержательной корректировке.

## Обсуждение

Эмпирический анализ (модель Раша) подтвердил высокую конструктивную валидность теста и внутреннюю согласованность заданий в измерении латентного конструктора (подготовка по информатике). Большинство заданий соответствуют модели, обеспечивая надежную дифференциацию. Хорошее соответствие модели косвенно поддерживает содержательную валидность, указывая на согласованное

измерение конструктора, определенного экспертизно верифицированной спецификацией. Однако задания с высокими Outfit MNSQ (>1,5), такие как СООТВ-31 и СООТВ-32 (возможная двусмысленность, проверка нескольких навыков), были выявлены как проблемные для измерения. Эти задания доработаны по результатам дополнительной экспертизы. По результатам контент-анализа и экспертной оценки (качественных методов) были уточнены и дополнены формулировки целей обучения, акцентирующие развитие ключевых компетенций в области ИКТ, таких как интерпретация информации, аргументация, принятие решений и решение проблем.

Для расширения когнитивного охвата теста и повышения его диагностического потенциала в спецификацию теста 2026 года включены новые тематические направления, в том числе «Современные тенденции развития информационных технологий», «IT Startup», «3D моделирования», «Информационная безопасность».

На основании полученных данных также было принято решение включить цели обучения непосредственно в спецификацию теста, что позволит более чётко опре-

делить и обосновать уровень требований, предъявляемых к участникам, а также повысить прозрачность и согласованность между учебными результатами, содержанием теста и его оценочной функцией.

## **Заключение**

Проведённое исследование подтвердило высокую содержательную валидность теста по информатике в рамках ЕНТ (на основе экспертного анализа и контент-анализа соответствия ГОСО), а также его конструктивную валидность и внутреннюю согласованность (на основе эмпирического моделирования по Рашу). Комплексный подход – сочетание документального анализа, экспертной оценки (фокус на содержательной валидности) и эмпирического моделирования (фокус на конструктивной валидности и внутренней согласованности) – позволил выявить как сильные стороны теста, так и области, требующие доработки. В результате были уточнены формулировки целей обучения, расширен когнитивный охват, а также внесены предложения по улучшению спецификации на 2026 год, включая добавление новых тематик и уровней мышления.

Это исследование вносит вклад в развитие научно обоснованных подходов к валидации образовательных тестов, применяя комплексную методологию, сочетающую оценку содержательной валидности через экспертизу и конструктивной валидности через модель Раша, в условиях национального контекста. Его результаты предоставляют разработчикам тестов эмпирические данные, позволяющие выявить как сильные, так и проблемные задания, уточнить цели обучения и обеспечить согласование тестового содержания с образовательными стандартами.

Практическая ценность заключается и в том, что выявленные закономерности могут быть использованы при обновлении спецификаций и разработке новых измерительных материалов. Выводы играют важную роль в образовательной политике, способствуя обоснованию требова-

ний к участникам, обеспечению прозрачности и согласованию целей обучения и оценивания. Применённая методология может служить основой для аналогичных исследований в других предметных областях и уровнях образования.

## **Список использованных источников**

1. Andrich, D., Mercer, A. International perspectives on selection methods of entry into higher education. – Canberra: National Board of Employment, Education and Training [and] Higher Education Council, 1997.
2. Приказ Министра образования и науки Республики Казахстан № 204 «Об утверждении Правил проведения единого национального тестирования» от 2 мая 2017 года. – URL: <https://adilet.zan.kz/rus/docs/V1700015173>
3. Messick, S. Validity // Linn, R. L. (Ed.). Educational Measurement (3rd ed.). – New York: American Council on Education / Macmillan, 1989. – P. 13–103.
4. Анастази, А., Урбина, С., Алексеев, А. А. Психологическое тестирование. – 7-е междунар. изд. – СПб.: Питер, 2007. – (Серия «Мастера психологии»).
5. Thematic analysis. – URL: [https://en.wikipedia.org/wiki/Thematic\\_analysis?utm\\_source](https://en.wikipedia.org/wiki/Thematic_analysis?utm_source)
6. Bloom's Taxonomy for Learning Experience Design. – URL: [https://lxdlearningexperiencedesign.com/frameworks/blooms-taxonomy-for-learning-experience-design/?utm\\_source](https://lxdlearningexperiencedesign.com/frameworks/blooms-taxonomy-for-learning-experience-design/?utm_source)
7. Приказ Министра просвещения Республики Казахстан № 348 «Об утверждении государственных общеобязательных стандартов дошкольного воспитания и обучения, начального, основного среднего и общего среднего, технического и профессионального, послесреднего образования» от 3 августа 2022 года. – URL: <https://adilet.zan.kz/rus/docs/V2200029031>
8. Vieira, A. J., Garrett, J. M. Understanding interobserver agreement: the Kappa statistic // Family Medicine. – 2005. – Vol. 3(5). – URL: <https://www.ncbi.nlm.nih.gov/pubmed/15883903>
9. Smith, E. V., Smith, R. M. Evidence for reliability of measures and validity of measure interpretation: A Rasch measurement perspective // Introduction to Rasch Measurement: Theory, Models and Applications. – Maple Grove: JAM Press, 2004.
10. Wright, B. D., Masters, G. N. Rating Scale Analysis. – Chicago: MESA Press, 1982. – URL: <https://files.eric.ed.gov/fulltext/ED263199.pdf>
11. Bond, T. G., Fox, C. M. Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). – New York: Routledge, 2015. – <https://doi.org/10.1111/j.1745-3984.2003.tb01103.x>

## References

- Andrich, D., Mercer, A. International perspectives on selection methods of entry into higher education. – Canberra: National Board of Employment, Education and Training [and] Higher Education Council, 1997.
- Prikaz Ministra prosvešenia Respubliki Kazahstan № 348 «Ob utverjenii gosudarstvennyh obšeobázatelnyh standartov doškólnogo vospitania i obuchenia, nachálnogo, osnovnogo srednego i obšego srednego, tehnicheskogo i profesionálnogo, poslesrednego obrazovaniya» ot 3 avgusta 2022 goda [Order of the Minister of Education and Science of the Republic of Kazakhstan No. 204 of May 2, 2017 "On Approval of the Rules for Conducting the Unified National Testing"]. – URL: <https://adilet.zan.kz/rus/docs/V2200029031>
- Messick, S. Validity // Linn, R. L. (Ed.). Educational Measurement (3rd ed.). – New York: American Council on Education / Macmillan, 1989. – P. 13–103.
- Anastazi, A., Urbina, S., Alekseev, A. A. Psichologicheskoe testirovanie [Psychological assessment]. – 7-e mejdunar. izd. – SPb.: Piter, 2007. – (Seria «Mastera psihologii»).
- Thematic analysis. – URL: [https://en.wikipedia.org/wiki/Thematic\\_analysis?utm\\_source](https://en.wikipedia.org/wiki/Thematic_analysis?utm_source)
- Bloom's Taxonomy for Learning Experience Design. – URL: <https://lxlearningexperiencedesign.com/frameworks/blooms-taxonomy-for-learning-experience-design/>
- Prikaz Ministra prosvešenia Respubliki Kazahstan № 348 «Ob utverjenii gosudarstvennyh obšeobázatelnyh standartov doškólnogo vospitania i obuchenia, nachálnogo, osnovnogo srednego i obšego srednego, tehnicheskogo i profesionálnogo, poslesrednego obrazovaniya» ot 3 avgusta 2022 goda [Order of the Minister of Enlightenment of the Republic of Kazakhstan No. 348 of August 3, 2022 "On Approval of the State Compulsory Standards of Preschool Education and Training, Primary, Basic Secondary and General Secondary, Technical and Vocational, and Post-Secondary Education"]. – URL: <https://adilet.zan.kz/rus/docs/V2200029031>
- Vieira, A. J., Garrett, J. M. Understanding interobserver agreement: the Kappa statistic // Family Medicine. – 2005. – Vol. 3(5). – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15883903/>
- Smith, E. V., Smith, R. M. Evidence for reliability of measures and validity of measure interpretation: A Rasch measurement perspective // Introduction to Rasch Measurement: Theory, Models and Applications. – Maple Grove: JAM Press, 2004.
- Wright, B. D., Masters, G. N. Rating Scale Analysis. – Chicago: MESA Press, 1982.
- Bond, T. G., Fox, C. M. Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). – New York: Routledge, 2015. – URL: <https://doi.org/10.1111/j.1745-3984.2003.tb01103.x>

## ҰБТ аясындағы информатика пәні бойынша тест мазмұнының валидтілігін бағалау: сараптамалық бағалау және психометриялық талдау

Ш.Б. Алтыбаева<sup>\*1</sup>, И.У. Сагиндыков<sup>2</sup>, Г.К. Жабаева<sup>3</sup>, Ш.С. Шакралиева<sup>4</sup>

<sup>1,2,3,4</sup>Ұлттық тестілеу орталығы

Астана қ., Қазақстан Республикасы



**Аннотация.** Зерттеудің өзектілігі, Қазақстандағы Бірыңғай ұлттық тестілеуінің жоғары деңгейдегі емтихан екенін, оның нәтижелері түлектердің білім алу траекториясына айтарлықтай әсер ететінін дәлелдейді. Бұл тест материалдарының валидтілігі мен сенімділігін қамтамасыз ету қажеттілігін туындалады, әсіресе динамикалық дамып келе жатқан пәндер, мысалы, информатика бойынша. 2022 жылдан бастап информатика бейіндік пәндер санатына өтіне байланысты, тест мазмұнының тәуелсіз тексерілуінің маңызы арта түсті. Зерттеу Мемлекеттік жалпыға міндетті білім беру стандартының (МЖМБС) талаптарына спецификацияның сәйкестігін сараптау және тапсырмалардың сапасын психометриялық бағалау арқылы ҰБТ-ның информатика пәні бойынша тесттің мазмұндық валидтілігін көшенді талдау мақсатында жүргізілді. Зерттеу 54 оку мактаптарының мазмұнын талдауды, алты сарапшының қатысуымен Тәуелсіз сараптамалық бағалауды, тақырыптық кодтау әдісімен сапалы талдауды және 116 оқушының ұлтісіндеғі Раш моделі бойынша сандық бағалауды қамтитын аралас әдістерді қолданды. Зерттеу нәтижелері бойынша оку мактаптарының құрылымында түсіну (25%) және қолдану (30%) тапсырмалары басым екенін көрсетті, талдау (15%) және бағалау (10%)

сияқты жоғары когнитивті деңгейдегі тапсырмалардың жетіспеушілігі байқалды. Сараптамалық талдау мазмұнның репрезентативтілігін 0,80 Каппа келісім көзфициентімен растады, алайда 3D модельдеу және ақпараттық қауіпсіздік сияқты заманауи тақырыптарды қосу қажеттілігін анықтады. Психометриялық талдау  $\alpha = 0,94$  тестінің жоғары сенімділігін және Rash моделінің көтеген тапсырмаларының сәйкестігін көрсетті, Outfit MNSQ 1,5-тен жоғары екі проблемалық тапсырма анықталды. Зерттеу нәтижелері бойынша тестті модернизациялауға арналған ұсыныстар әзірленді: оку мақсаттарының тұжырымдары нақтыланды, жаңа тақырыптық блоктар енгізілді және проблемалық тапсырмалар түзетілді. Валидтілікті көшенді тексеру әдісі өлшеу материалдарының сапасын арттыру үшін тиімді екенін дәлелдеді. Зерттеу нәтижелері тест спецификацияларын әзірлеу және жаңарту тәжірибесінде қолданылуы мүмкін.



**Кілтті сөздер:** валидтілік, информатика тесті, Ұлттық бірыңғай тестілеу, тест спецификациясы, Блум таксономиясы, Раш моделі, психометриялық талдау.

## Validation of Informatics Test Content in the Framework of the UNT: Expert Assessment and Psychometric Analysis

\*Sh. Altybaeva<sup>1</sup>, I Sagindikov<sup>2</sup>, G. Zhabayeva<sup>3</sup>, Sh. Shakralieva<sup>4</sup>

<sup>1,2,3,4</sup>National Testing Center  
Astana, Republic of Kazakhstan



**Abstract.** The relevance of the study is significant because the Unified National Testing (UNT) in Kazakhstan is a high-stakes exam, the results of which have a substantial impact on the educational trajectories of graduates. This necessitates ensuring the validity and reliability of the test materials, particularly for dynamically developing subjects such as computer science. Additionally, the independent verification of the test content gains importance due to the transition of computer science into the category of specialized subjects starting in 2022. The study aimed to conduct a comprehensive analysis of the content validity of the computer science test in the UNT through an expert evaluation of its alignment with the requirements of the State Compulsory Education Standard (SCES) and a psychometric assessment of the quality of the test items. The research employed mixed methods, including content analysis of 54 learning objectives, an independent expert evaluation with the participation of six experts, qualitative analysis using thematic coding, and quantitative assessment using the Rasch model based on a sample of 116 students. The results of the study showed that the structure of the learning objectives was dominated by tasks focusing on understanding (25%) and application (30%), while there was a lack of tasks at higher cognitive levels, such as analysis (15%) and evaluation (10%). Expert analysis confirmed the representativeness of the content with a Kappa agreement coefficient of 0.80, but identified the need to include modern topics such as 3D modeling and information security. Psychometric analysis demonstrated high reliability of the test with  $\alpha = 0.94$  and the alignment of most items with the Rasch model, while two problematic items were identified with an Outfit MNSQ value above 1.5. Based on the findings, recommendations for the modernization of the test were developed: the formulations of the learning objectives were refined, new thematic blocks were added, and problematic items were corrected. The comprehensive approach to validation proved effective in improving the quality of measurement materials. The results can be applied in the practice of developing and updating test specifications.



**Keywords:** validity, computer science test, Unified National Testing, test specification, Bloom's taxonomy, Rasch model, psychometric analysis.

Материал поступил в редакцию 31.07.2025