

Абдрасилов Б.С.,^{1*}  Бейсенова Ж.Ж.,¹  Шинетова Л.Е.,¹ 
Ниязов Т.Ж.,¹  Жабаева Г.К.¹ 

¹Национальный центр тестирования, г.Астана, Казахстан

АКАДЕМИЧЕСКОЕ ИССЛЕДОВАНИЕ: ОЦЕНКА КРИТИЧЕСКОГО МЫШЛЕНИЯ ПРЕТЕНДЕНТОВ В МАГИСТРАТУРУ

Аннотация

Представленная статья посвящена исследованию оценки критического мышления претендентов на поступление в магистратуру в Республике Казахстан. Авторы подчеркивают значимость критического мышления для успешной научной и профессиональной деятельности в контексте современного рынка труда. Основное внимание уделено разработке и апробации тестов, предназначенных для оценки готовности к обучению на уровне магистратуры, включающих задания, направленные на оценку критического и аналитического мышления. В исследовании применен количественный описательный метод с участием разработчиков тестов, экспертов и членов Республиканской апелляционной комиссии. Проведенный психометрический анализ подтвердил надежность и валидность предложенных тестовых заданий, что свидетельствует об их эффективности в процессе отбора магистрантов. Особую ценность представляет собой систематизация опыта оценки критического и аналитического мышления в Республике Казахстан. Практическая значимость работы заключается в ее потенциальной пользе для разработчиков и экспертов тестовых материалов, а также для кандидатов, поступающих на программы послевузовского образования. Результаты исследования демонстрируют, что разработанный тест является действенным инструментом для оценки готовности абитуриентов к обучению и исследовательской деятельности, способствуя повышению качества подготовки обучающихся в соответствии с современными требованиями.

Ключевые слова: тест на определение готовности к обучению, психометрический анализ, качество измерительного инструмента, модель Раша.

Б.С.Абдрасилов,^{1*}  Ж.Ж.Бейсенова,¹  Л.Е.Шинетова,¹ 
Т.Ж.Ниязов,¹  Г.К.Жабаева¹ 





¹Ұлттық тестілеу орталығы, Астана қ., Қазақстан

АКАДЕМИЯЛЫҚ ЗЕРТТЕУ: МАГИСТРАТУРАҒА ҮМІТКЕРЛЕРДІҢ СЫНИ ОЙЛАУЫН БАҒАЛАУ

Аңдатпа

Мақалада Қазақстан Республикасындағы магистратураға түсуге үміткерлердің сыни тұрғыдан ойлауын бағалауға арналған зерттеу ұсынылған. Авторлар сыни тұрғыдан ойлаудың табысты ғылыми және кәсіби қызмет үшін маңыздылығына тоқталып, оның қазіргі еңбек нарығындағы өзектілігін атап көрсетеді. Сыни және аналитикалық ойлау тапсырмаларын қамтитын магистратураға дайындықты бағалауға бағытталған тестілерді әзірлеуге және сынақтан өткізуге баса назар аударылады. Зерттеуде тест тапсырмаларын әзірлеуші, сарапшылар және Республикалық апелляциялық комиссия мүшелері қатысқан сандық сипаттама әдісі қолданылды. Психометриялық талдау ұсынылған тесттің сенімділігі мен негізділігін айқындап, магистранттар контингентін қалыптастыру үшін тиімділігін көрсетеді. Зерттеудің ерекше құндылығы Қазақстан Республикасындағы сыни және аналитикалық ойлауды бағалау тәжірибесін жүйелеу болып табылады. Жұмыстың практикалық маңыздылығы, нәтижелер тест әзірлеушілер мен сарапшыларға ғана емес, сонымен қатар жоғары оқу орнынан кейінгі білім беру бағдарламаларына үміткерлерге де пайдалы болады. Зерттеу нәтижесінде әзірленген тест үміткерлердің оқуға және ғылыми-зерттеу жұмыстарына дайындығын бағалаудың тиімді құралы болып табылатыны анықталды, бұл қазіргі заманғы талаптарға сәйкес білім алушылардың дайындық сапасын жақсартуға ықпал етеді.

Түйін сөздер: оқуға дайындықты анықтау тесті, психометриялық талдау, өлшеу құралының сапасы, Раш моделі.

Abdrasilov B.,*¹  Beisenova Zh.,¹  Shinetova L.,¹  Niyazov T.,¹  Zhabayeva G.¹ 
¹National Testing Center, Astana, Kazakhstan

ACADEMIC RESEARCH: ASSESSMENT OF CRITICAL THINKING OF APPLICANTS FOR MASTER'S DEGREE

Abstract

The article presents a study on the assessment of critical thinking among applicants for admission to a master's degree in the Republic of Kazakhstan. The authors emphasize the importance of critical thinking for successful scientific and professional activities, emphasizing its relevance in the modern labor market. The main focus is on the development and testing of tests aimed at assessing readiness for graduate studies, which include tasks on critical and analytical thinking. The study used a quantitative descriptive method in which test developers, experts and members of the Republican Appeals Commission participated. The psychometric analysis confirmed the reliability and validity of the proposed test tasks, which indicates their effectiveness for the selection of undergraduates. The special value of the research lies in the systematization of the experience of evaluating critical and analytical thinking in the Republic of Kazakhstan. The practical significance of the work lies in the fact that the results can be useful not only to test developers and experts, but also to candidates entering postgraduate education programs. As a result of the research, it was found that the developed test is an effective tool for assessing candidates' readiness for training and research activities, which contributes to improving the quality of students' training in accordance with modern requirements.

Keywords: test to determine readiness for learning, psychometric analysis, quality of measuring instrument, Rasch model.

Введение. Республика Казахстан проводит последовательную государственную политику, ориентированную на модернизацию системы образования и наращивание высококвалифицированного кадрового потенциала, что находит отражение в стратегическом документе – Концепции развития высшего образования и науки на 2023-2029 годы. В рамках данной Концепции приоритетное значение придается развитию послевузовского образования, как ключевому элементу подготовки кадров высшей квалификации, способных обеспечить научно-технологический прогресс страны [1].

Процесс становления независимого Казахстана сопровождался целенаправленной интеграцией в глобальное образовательное пространство, важнейшим аспектом которой стала реструктуризация системы высшего образования. В настоящее время в Казахстане функционирует трехуровневая модель подготовки кадров (бакалавриат, магистратура, докторантура PhD), внедрение которой явилось следствием присоединения страны к Болонскому процессу в 2010 году [2]. Последовавшие реформы были ориентированы на обеспечение высокого качества подготовки специалистов на всех ступенях высшего образования в соответствии с международными стандартами.

Статистические данные Бюро национальной статистики за начало 2024-2025 учебного года свидетельствуют о поступательном увеличении контингента обучающихся на послевузовском уровне в Республике Казахстан, включая рост числа магистрантов (39 172), резидентов (8 882) и докторантов (7 633) [3]. Данная динамика коррелирует с положениями Концепции развития высшего образования и науки на 2023-2029 годы, предусматривающей расширение государственной грантовой поддержки программ магистратуры и докторантуры. Вместе с тем, усиление формирования контингента обучающихся по программам магистратуры актуализирует необходимость разработки научно обоснованных методов оценки готовности претендентов к обучению на данном уровне, а также их потенциала к успешной научно-исследовательской и профессиональной деятельности. Эффективный отбор претендентов, ориентированный на присуждение степени магистра, требует разработки критериев качественного инструментария и соответствующей структуры оценочных процедур, что является ключевым фактором в обеспечении высокого качества контингента магистрантов.

В целях повышения качества отбора, в соответствии с правилами приема в магистратуру, поступающие сдают вступительный экзамен в форме комплексного тестирования [4].

В структуре комплексного тестирования, применяемого для отбора абитуриентов на программы магистратуры, одним из ключевых критериев выступает результат теста на

определение готовности к обучению. Включение заданий, направленных на оценку критического и аналитического мышления, говорит о признании значимости данных когнитивных компетенций в контексте современного образовательного процесса.

Актуальным вызовом для системы высшего образования Республики Казахстан выступает обеспечение адекватности образовательных программ требованиям современного рынка труда, который демонстрирует отчетливую потребность в специалистах с высоким уровнем критического мышления, аналитических способностей и компетенций в решении сложных задач. Это обуславливает необходимость пересмотра образовательных стратегий с акцентом на развитие критического мышления, исследовательских и аналитических навыков, а также внедрение валидных инструментов их оценки.

В этой связи, объектом настоящего исследования является процесс оценки теста на определение готовности к обучению (далее – ТГО) при поступлении в магистратуру на основе критического и аналитического мышления, а также профессиональных навыков. Предметом исследования выступает методологическая разработка тестовых заданий ТГО, способных эффективно верифицировать способность претендентов к критическому рассуждению, решению базовых математических задач, интерпретации данных и ассимиляции информации. Основной целью данного исследования является формулирование научно-практических рекомендаций для разработчиков ТГО, предназначенных для использования в процедуре поступления на программы магистратуры. Значимость разработки валидного ТГО детерминирована потребностью в совершенствовании процедуры комплексного тестирования (далее – КТ) при отборе претендентов в магистратуру и повышении качественных характеристик принимаемого контингента в высших учебных заведениях Казахстана.

Актуальность темы развития критического мышления в образовательном контексте и его роли в формировании квалифицированных специалистов привлекает все большее внимание в Казахстане. Современные требования рынка труда и высокие стандарты образовательных программ позиционируют критическое мышление как ключевую компетенцию для успешного обучения на уровне магистратуры. В рамках данного контекста, настоящее исследование направлено на разработку и апробацию тестовых заданий, оценивающих уровень критического мышления претендентов на программы послевузовского образования.

Основные положения. В современном мире, который быстро меняется, важно уметь быстро адаптироваться к изменениям. Сегодняшнюю среду часто описывают термином «VUCA», что является аббревиатурой, обозначающей нестабильность, неопределенность, сложность и двусмысленность [5]. Это явление реально и ускоряется. Люди, обладающие хорошими навыками решения проблем, критическим мышлением и умением принимать решения, смогут успешно справляться с вызовами и возможностями, с которыми они сталкиваются. Согласно информации, предоставленной Всемирным экономическим форумом 2023 года, критическое мышление занимает первое место среди самых востребованных навыков XXI века [6].

Прежде чем мы рассмотрим, почему критическое мышление играет столь важную роль в магистратуре, необходимо понять, что оно собой представляет. Критическое мышление – это процесс, включающий в себя способность анализировать аргументы, предсказывать потенциальные последствия, разрабатывать стратегии, находить альтернативы, подходящие способы решения ситуаций и принимать обоснованные решения. Этот процесс включает рефлексивное мышление, процедуры, практики и характеры [7]. Также стоит отметить, в статье, посвященной анализу опыта комплексного тестирования для поступающих на программы послевузовского образования в Республике Казахстан, впервые предпринята попытка изучить и систематизировать опыт оценки навыков критического и аналитического мышления в стране [8].

Навыки критического мышления являются одной из самых актуальных и важных тем в исследованиях в области образования и привлекают внимание многих исследователей. Например, в мета-анализе Huber C.R. и Kuncel N.R. рассматривалось, способствует ли обучение

в колледже развитию критического мышления [9]. Lai E.R. также проводил исследования в этой области, акцентируя внимание на сотрудничестве и взаимодействии в учебном процессе [10].

Исследования показывают, что критическое мышление не только способствует академическим успехам, но и играет ключевую роль в профессиональном развитии и личностном росте. Фасоне подчеркивает, что критическое мышление включает в себя как когнитивные, так и эмоциональные компоненты, что делает его комплексным навыком, требующим развития через целенаправленное обучение и практику [11].

Эннис Р.Х. также отмечает, что фактически, критическое мышление является важным компонентом эффективного принятия решений и решения проблем в различных профессиональных контекстах [12].

Разработка тестов, направленных на оценку критического мышления, требует строгого соблюдения методологических принципов для обеспечения валидности и надежности измерительного инструмента. В соответствии с положениями В.С.Аванесова [13], процесс конструирования тестовых заданий включает последовательные этапы: определение целей тестирования, генерация тестового контента, проведение пилотного тестирования, статистический анализ полученных данных и последующая модификация тестовых элементов. Кроме того, современные подходы к педагогическим измерениям, представленные в работах А.Н.Майорова [14], содержат общеметодические рекомендации по разработке как отдельных тестовых заданий, так и целостной структуры измерительного инструмента.

В работах Р.Линна, Т.Халадайны и Н.Гронлунда изложены основные принципы и методики разработки тестовых заданий и создания дистракторов, применяемые в образовательных целях. В этих исследованиях обсуждаются процессы оценки в образовательной среде, включая методы создания заданий с множественным выбором и качественных дистракторов, а также вопросы валидности и надежности тестов [15] [16] [17].

Надежность является важнейшим показателем качества теста, определяющим точность и стабильность измерений, свободных от случайных ошибок. По сути, надежность отражает устойчивость и согласованность результатов, получаемых при повторном тестировании одних и тех же лиц в разное время, при использовании альтернативных вариантов теста или при изменении условий проведения [18].

В идеале, надежный тест должен демонстрировать одинаковые результаты при многократном применении к одним и тем же учащимся, однако на практике это сложно достижимо. Например, результаты могут варьироваться между разными группами из-за различий в их подготовке. При повторном тестировании той же группы результаты могут снизиться из-за забывания материала или повыситься благодаря опыту предыдущего тестирования и дополнительной подготовке. Кроме того, на результаты могут влиять условия проведения, а также физическое и психологическое состояние испытуемых. К ключевым характеристикам надежности относятся согласованность, стабильность и чувствительность. Надежность тестов определяется различными коэффициентами, среди которых выделяется именно коэффициент Кронбаха (α). Коэффициент Кронбаха является распространенным индикатором надежности тестов, оценивающим внутреннюю согласованность элементов, измеряющих один и тот же конструкт; значения 0,70 и выше свидетельствуют о приемлемой надежности, 0,80 и выше – о хорошей, а 0,90 и выше – об отличной [19].

Валидность означает пригодность тестовых результатов для той цели, ради чего проводилось тестирование. В отличие от надежности, валидность не поддается прямой статистической оценке и подтверждается эмпирически. Доказательства валидности теста основываются на следующих аспектах:

- Четкость и точность формулировок заданий. Задания должны быть однозначно понятны тестируемым.

- Соответствие заданий учебной программе и оцениваемым навыкам. Каждое задание должно быть четко соотнесено с определенной областью знаний или умением.

- Минимизация влияния посторонних знаний и навыков. Выполнение теста не должно зависеть от знаний или умений, не являющихся целью оценки.

- Полнота охвата содержания. Итоговый тест должен включать задания по всем ключевым разделам и темам учебной программы.

- Сбалансированность уровня сложности заданий. Структура теста должна включать задания различной степени трудности.

Валидность теста представляет собой многомерную концепцию, включающую содержание и дизайн теста, его формулировку и применение, а также анализ и интерпретацию полученных результатов. Такие тесты играют важную роль в образовательных исследованиях, поскольку качество их заданий напрямую влияет на точность оценки знаний и навыков участников.

Материалы и методы. Настоящее исследование основывается на анализе структуры ТГО, предназначенного для оценки уровня знаний претендентов. ТГО включает 30 тестовых заданий закрытого типа, с максимальным баллом 30, и состоит из двух частей [20]:

1) Критическое мышление (15 заданий): Оценка способности к критическому рассуждению, решению элементарных математических задач, анализу и интерпретации графической, табличной и текстовой информации, а также консолидации данных различных форматов. Секция включает задания на сравнение значений (3), решение уравнений, неравенств, геометрических и текстовых задач (8), и анализ таблиц, диаграмм, схем и текстов (4).

2) Аналитическое мышление (15 заданий): Проверка способности к пониманию и анализу академических и научно-популярных текстов, определению основной идеи, установлению причинно-следственных связей и выявлению закономерностей. Секция содержит задания на заполнение пропусков (3), анализ текста с последующими вопросами (9) и выявление общей точки зрения в двух текстах (3). Текстовый материал охватывает различные научные дисциплины.

Данное исследование применяло количественный описательный метод для изучения процесса разработки тестовых заданий ТГО. В исследовании приняли участие разработчики тестовых заданий, эксперты и члены Республиканской апелляционной комиссии по ТГО, обладающие опытом работы в формировании базы тестовых заданий от 3 до 5 лет.

Для получения экспертных оценок относительно корректности разработки ТГО на базе Национального центра тестирования (далее – Центр) был организован круглый стол, в котором приняли участие 25 представителей профессорско-преподавательского состава вузов, выступавших в роли разработчиков тестовых заданий, экспертов и членов Республиканской апелляционной комиссии.

Обсуждение охватывало три основных аспекта:

- Понимание критического мышления, включенного в ТГО в разделе «Критическое мышление».

- Реализация критического мышления в тестах.

- Сбор мнений и рекомендаций по структуре и содержанию ТГО в разделе «Критическое мышление».

Наряду с качественным анализом был проведен психометрический анализ результатов апробации разработанных тестовых заданий с использованием методов классического (Classical Test Theory, CTT) и современного (Item Response Theory, IRT) анализа тестов. Психометрический анализ был направлен на подтверждение адекватности разработанных заданий в измерении заявленных навыков и их соответствия установленным критериям качества.

Результаты и обсуждение. Результаты проведенного исследования включают экспертную оценку разработки ТГО и психометрический анализ апробированного варианта теста.

Понимание критического мышления: Эксперты продемонстрировали глубокое понимание концепции критического мышления, включающей анализ аргументов, прогнозирование последствий, разработку стратегий, поиск альтернатив и принятие обоснованных решений.

Реализация критического мышления в тестах: Участники в целом положительно оценили отражение различных аспектов критического мышления в текущих заданиях ТГО для поступающих в магистратуру, однако отметили необходимость доработки ряда формулировок для повышения точности оценки данных навыков. В качестве мер по улучшению тестовых заданий было предложено включение более комплексных сценариев и задач, требующих углубленного анализа и оценки.

Рекомендации по структуре и содержанию тестов: Экспертами был предложен ряд рекомендаций по оптимизации структуры и содержания тестов. В частности, обсуждались изменения в спецификации и содержании тестов. Было отмечено, что текущее содержание охватывает школьную математику и статистику, исключая тригонометрию, исчисление и другие разделы математики более высокого уровня. Среди конкретных предложений выделялось усиление компонента задач на анализ данных и увеличение числа заданий, оценивающих навыки количественного анализа и интерпретации информации. Особое внимание уделялось вопросам формулировки заданий и подбору дистракторов, с целью исключения неоднозначности и обеспечения адекватной оценки способностей тестируемых. В ходе круглого стола были определены несколько ключевых аспектов, касающихся разработки ТГО по разделу «Критическое мышление».

По результатам экспертных оценок, касающихся разработки раздела «Критическое мышление» ТГО, а также учитывая предложения по изменению структуры и содержания теста, были внесены соответствующие коррективы в спецификацию ТГО. Данные изменения и результаты экспертных оценок были рассмотрены на заседании Научно-технического совета (далее – НТС) Центра 23 мая 2024 года, где по решению НТС было принято решение о внесении изменений в спецификацию теста и проведении дальнейшей разработки в соответствии с предложенными результатами.

В рамках исследования, с учетом выявленных экспертами ключевых моментов по разработке ТГО, а также учитывая предложения по изменению структуры и содержания, были сформированы задания и подготовлен вариант теста.

Анализ данных. В исследовании, направленном на психометрическую оценку теста на определение готовности к обучению, на добровольной основе приняли участие 1593 обучающихся выпускных курсов высших учебных заведений из различных регионов Республики Казахстан по разным группам образовательных программ (далее - ГОП), что обеспечивает достаточную выборку для статистического анализа и повышения надежности полученных результатов. Перед началом исследования все участники предоставили информированное согласие в соответствии с этическими нормами и получили подробные инструкции по проведению компьютерного тестирования.

Анализ данных проводился на основе ответов всех 1593 обучающихся, чей возрастной диапазон на момент сбора данных составлял 21-24 года, что соответствует типичному возрасту обучающихся выпускных курсов вузов.

Разработанный вариант теста, спецификация которого была подготовлена на основе экспертной оценки, был предложен всей выборке обучающихся (n=1593) для проведения психометрического анализа.

Классический анализ тестов (Classical Test Theory, СТТ) с помощью программы ТАР (Test Analysis Program)

Таблица 1 содержит ключевые параметры и их значения, описывающие результаты анализа теста. Эти данные позволяют оценить общие характеристики теста и качество измерительных инструментов.

Таблица 1. Сводные характеристики теста

Параметры	Показатели
Общее количество участников апробации	1593
Максимальный набранный балл	29
Минимальный набранный балл	1
Среднее значение	16,9
Стандартное отклонение	6,46
Средний показатель коэффициента трудности	0,57
Максимальный показатель коэффициента трудности	0,84
Минимальный показатель коэффициента трудности	0,18
Средний показатель коэффициента дискриминативности	0,52
Средняя точечно-бисериальная корреляция	0,39
Коэффициент надежности Альфа-Кронбаха	0,85
Ошибка измерения	2,33

Минимальный балл составляет 1, максимальный 29 (из 30 возможных), среднее равно 16,9 при стандартном отклонении 6,46. Таким образом, размах полученных баллов является умеренным. Это значит, что существует определенная вариативность в результатах, но эта вариативность не является чрезмерной. Умеренное стандартное отклонение позволяет лучше дифференцировать испытуемых. Если бы разброс был слишком мал, тест не смог бы эффективно различать испытуемых с разными уровнями знаний или способностей, так как большинство получили бы схожие баллы. С другой стороны, слишком большой разброс может указывать на неоднородность группы испытуемых или на то, что тест измеряет слишком много разных факторов. Стандартное отклонение в 6,46 говорит о том, что тест способен относительно неплохо различать испытуемых с разными уровнями подготовки, но при этом нет чрезмерной популяризации результатов.

Показатель трудности теста, равный 0,57 указывает на то, что тест имеет среднюю сложность, что означает, что тест достаточно дифференцирует участников и позволяет выявить как тех, кто хорошо усвоил материал, так и тех, кто испытывал трудности. Показатель дискриминативности 0,52 указывает на то, что тест имеет достаточно высокую дискриминативность и обладает хорошей способностью различать участников с высокими и низкими результатами.

Коэффициент надежности Альфа-Кронбаха составляет 0,85, что указывает на высокую степень стабильности и согласованности результатов. Это говорит о том, что данный тест является качественным и эффективным инструментом для оценки знаний участников. Значение выше 0,8 считается приемлемым для большинства ответственных тестов.

Ошибка измерения составляет 2,33 балла. Для теста с высокой степенью надежности, такая ошибка измерения не является критичной.

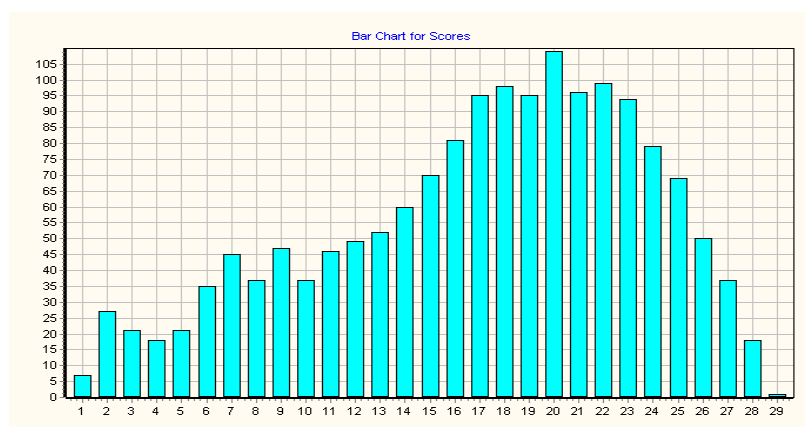


Рисунок 1. Распределение баллов по итогам тестирования

На рисунке 1 видно, что распределение баллов имеет форму, близкую к колоколообразной, что предполагает тенденцию к нормальному распределению. Однако оно несколько скошено влево (отрицательная асимметрия). Это значит, что большинство обучающихся продемонстрировали результаты выше среднего значения, в то время как количество участников с очень низкими баллами было относительно невелико.

Приведенный ниже график (рисунок 2) иллюстрирует результаты анализа заданий теста, сравнивая их сложность (коэффициент решаемости) и дискриминационную способность (точечно-бисериальная корреляция).

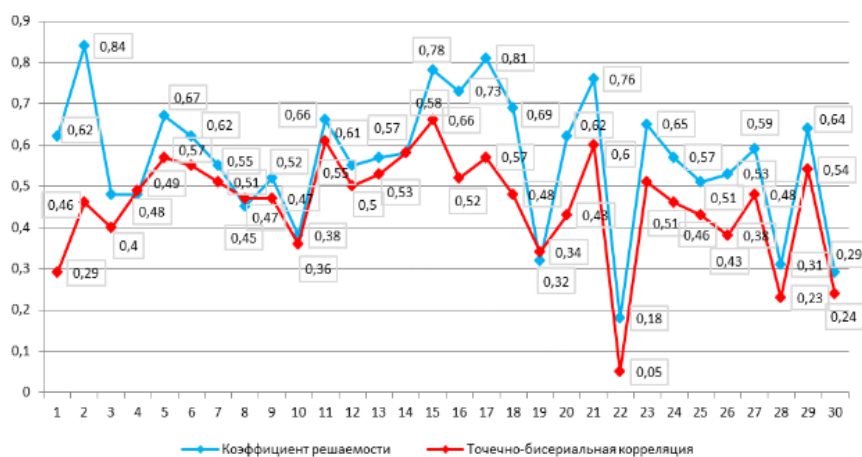


Рисунок 2. График соответствия коэффициента решаемости и точечно-бисериальной корреляции

Рисунок 2 представляет детальный анализ заданий, критически важный для оценки психометрических свойств теста. Коэффициент решаемости, отображенный синей линией, иллюстрирует долю участников теста, которые правильно ответили на каждое задание. Этот показатель выявляет значительный разброс в сложности заданий. Например, задание 2 демонстрирует высокую решаемость, указывая на его относительную легкость, в то время как задания 22, 28 и 30 показывают более низкую решаемость, предполагая большую сложность. Такая изменчивость ожидается в хорошо составленных тестах, стремящихся оценить спектр знаний и навыков.

Точечно-бисериальная корреляция (красная линия) оценивает способность задания различать участников теста с высокими и низкими показателями. График показывает, что большинство заданий демонстрируют умеренную положительную дискриминацию, подразумевая, что они вносят вклад в способность теста эффективно ранжировать участников теста.

Хотя тест демонстрирует диапазон сложностей заданий и в целом приемлемую дискриминацию, однако некоторые исключения заслуживают внимания. Задание 22 демонстрирует почти нулевую точечно-бисериальную корреляцию, сигнализируя о серьезном недостатке дискриминации. Этот результат особенно тревожен, поскольку он предполагает, что задание не вносит значимого вклада в различие между более или менее знающими людьми и может вносить шум в результаты теста. Задания 28 и 30 также показывают сниженную дискриминацию по сравнению с остальной частью теста. В дальнейшем мы углубленно исследуем специфику функционирования этих заданий, используя возможности специализированного программного обеспечения Winsteps.

Для более глубокого понимания взаимодействия между испытуемыми и заданиями, а также для получения мер, инвариантных относительно выборки, современная теория тестов предлагает более мощный методологический аппарат, реализуемый с использованием модели Раша и программного обеспечения Winsteps.

Анализ дихотомических данных с использованием Winsteps: интерпретация выходных параметров (Item Response Theory, IRT)

Таблица 2 содержит сводные статистические данные, полученные в результате анализа ответов 1593 испытуемых на 30 вопросов с использованием программы Winsteps (версия 5.8.3.0).

Таблица 2. Сводная статистика

Сводная статистика по 1593 респондентам								
1	Всего баллов 2	Количество 3	Измерение 4	Стандартная ошибка 5	Статистики соответствия			
					INFIT		OUTFIT	
					MNSQ 6	ZSTD 7	MNSQ 8	ZSTD 9
Среднее значение	17,0	30,0	0,33	0,45	0,99	0,00	1,04	0,09
Стандартная ошибка среднего	0,2	0,0	0,03	0,00	0,00	0,02	0,01	0,03
Стандартное отклонение популяции	6,5	0,0	1,22	0,09	0,16	0,86	0,46	1,04
Стандартное отклонение выборки	6,5	0,0	1,22	0,09	0,16	0,86	0,46	1,04
Максимальное значение	29,0	30,0	3,75	1,04	1,60	3,44	7,03	3,73
Минимальное значение	1,0	30,0	-3,70	0,39	0,62	-2,95	0,18	-2,47
Реальная среднеквадратическая ошибка измерения		0,48	Истинное стандартное отклонение	1,13	Коэффициент разделения	2,36	Коэффициент надежности оценки респондентов	0,85
Модельная среднеквадратическая ошибка измерения		0,46	Истинное стандартное отклонение	1,13	Коэффициент разделения	2,45	Коэффициент надежности оценки респондентов	0,86
Стандартная ошибка среднего способности респондентов			значения	0,03				
Сводная статистика по 30 элементам								
	Всего баллов	Количество	Измерение	Стандартная ошибка	Статистики соответствия			
					INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Среднее значение	901,5	1593,0	0,0	0,06	0,99	-0,32	1,04	0,14
Стандартная ошибка среднего	46,1	0,0	0,17	0,00	0,02	0,77	0,06	0,78
Стандартное отклонение популяции	248,3	0,0	0,89	0,01	0,13	4,15	0,31	4,18
Стандартное отклонение выборки	252,6	0,0	0,90	0,01	0,13	4,22	0,32	4,25
Максимальное значение	1336,0	1593,0	2,24	0,08	1,26	8,77	2,34	9,90
Минимальное значение	286,0	1593,0	-1,75	0,06	0,73	-7,62	0,56	-7,42
Реальная среднеквадратическая ошибка измерения		0,06	Истинное стандартное отклонение	0,89	Коэффициент разделения	14,22	Коэффициент надежности элемента	1,00
Модельная среднеквадратическая ошибка измерения		0,06	Истинное стандартное отклонение	0,89	Коэффициент разделения	14,54	Коэффициент надежности элемента	1,00
Стандартная ошибка среднего значения элемента				0,17				

Анализ сводной статистики по респондентам выявил, что средний уровень способности участников тестирования (0,33) незначительно превышает средний уровень сложности заданий, централизованный на нуле в рамках модели Раша. Это хорошо, потому что тест в целом оказался подходящим по сложности для нашей группы. Распределение способностей характеризуется умеренным разбросом (стандартное отклонение 1,22 логита), а средняя стандартная ошибка измерения способности составила 0,45 логита, что указывает на достаточную точность индивидуальных оценок, т.е. мы можем быть уверены в том, что тест хорошо различает тех, кто знает материал лучше, и тех, кто знает его хуже.

Ключевым аспектом оценки качества теста является соответствие эмпирических данных теоретической модели Раша. Статистики соответствия INFIT и OUTFIT, отражающие взвешенную и невзвешенную дисперсии остатков, соответственно, продемонстрировали близкие

к идеальному значению 1,0 среднеквадратичные значения MNSQ и значения ZSTD, находящиеся в пределах нормы. Это свидетельствует об общем хорошем соответствии поведения респондентов ожиданиям модели, что повышает уверенность в интерпретации полученных оценок способностей.

Надежность оценки способностей респондентов оказалась высокой, о чем свидетельствует коэффициент надежности 0,85. Данный показатель указывает на то, что тест способен стабильно и точно дифференцировать респондентов с различными уровнями. Коэффициент разделения (2,36) говорит о том, что тест достаточно хорошо разделяет участников примерно на 2-3 статистически значимо отличающиеся группы по уровню их знаний.

Анализ сводной статистики по элементам теста показал, что средняя сложность заданий ожидаемо центрирована на нуле, а стандартная ошибка измерения сложности элементов является низкой (0,06), что говорит о высокой точности оценки трудности каждого задания, т.е. большинство вопросов работают так, как мы и ожидали. Однако, следует обратить внимание на максимальное значение OUTFIT, которое значительно превышает общепринятые критические значения. Это может указывать на наличие отдельных элементов, поведение ответов на которые является менее предсказуемым в рамках модели Раша и требует детального изучения.

Высокий коэффициент надежности элементов (1,00) свидетельствует о стабильности и точности оценки сложности каждого задания. Коэффициент разделения элементов (14,22) указывает на способность теста дифференцировать задания по уровню сложности на большом количестве статистически значимых градаций. Такая высокая способность разделять элементы по сложности означает, что данный тест может точно измерять знания участников на разных уровнях.

Информацию, которая поможет более глубоко понять характеристики каждого задания и выявить потенциальные проблемы, содержит таблица 14.1 в Winsteps, представляющая собой детальную статистику по каждому из 30 заданий. Анализ таблицы позволил выявить несколько заданий, которые не совсем соответствуют модели Раша. Особенно выделяются элементы 22, 28 и 30 с высокими положительными значениями статистик несоответствия, что требует дальнейшего изучения причин такого поведения. Возможно, эти элементы нуждаются в пересмотре или исключении из теста. Элементы 15 и 17 с высокими отрицательными значениями также заслуживают внимания.

Обратимся к более подробному рассмотрению заданий 22, 28 и 30, которые были предварительно выявлены нами как потенциально проблемные еще в ходе классического анализа тестовых заданий.

Представленный график отображает Item Characteristic Curve (ICC) для задания 22 (рисунок 3), демонстрирующем почти нулевую точечно-бисериальную корреляцию. ICC показывает вероятность правильного ответа на задание в зависимости от уровня способности респондента относительно трудности данного задания.

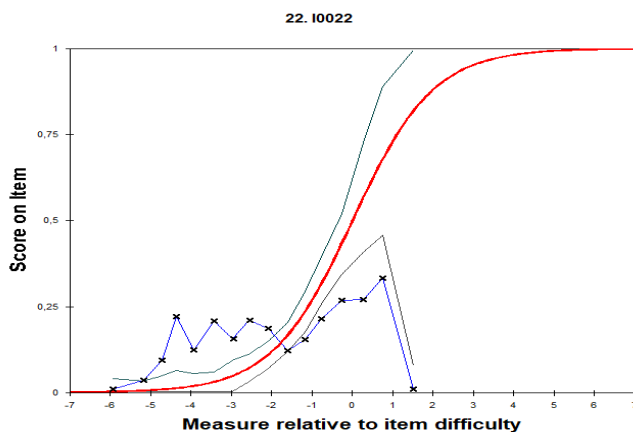


Рисунок 3. График Item Characteristic Curve (ICC) для задания 22

График демонстрирует некоторое несоответствие наблюдаемых данных теоретической модели Rasch. Наиболее выраженным является неожиданное снижение вероятности правильного ответа среди респондентов с уровнем способности, превышающим трудность задания. Это требует дальнейшего изучения данного задания. Необходимо провести качественный анализ содержания задания для выявления возможных проблем с формулировкой или содержанием. Также следует проверить данные на наличие аномальных паттернов ответов.

Полученные результаты могут указывать на необходимость пересмотра или исключения данного задания из тестового инструмента, поскольку его поведение не соответствует ожидаемой логике измерения и может негативно влиять на валидность результатов тестирования.

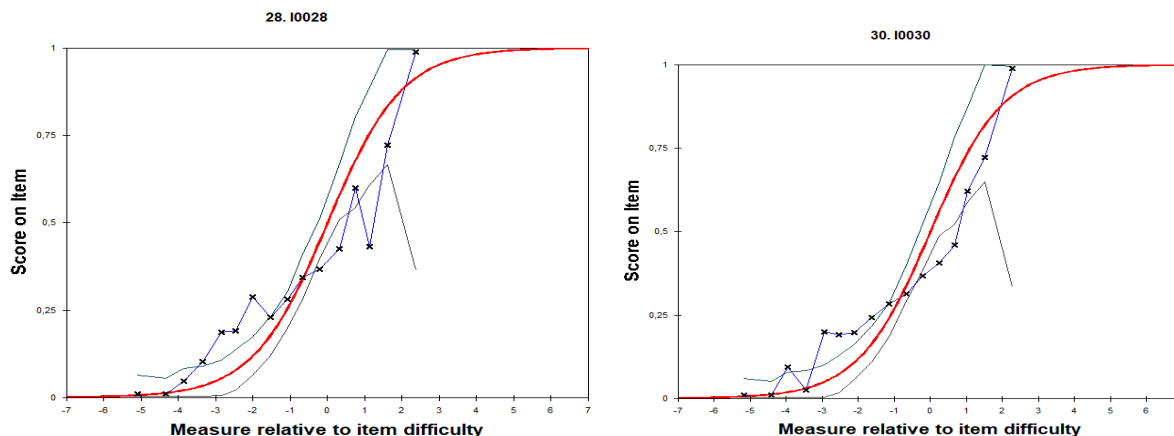


Рисунок 4. График Item Characteristic Curve (ICC) для заданий 28 и 30

Задания 28 и 30 (рисунок 4) характеризуются смещением трудности в сторону легких заданий и проблемами с соответствием модели эмпирическим данным, особенно при высоких уровнях способности. Следует более тщательно изучить формулировку заданий на предмет возможных неоднозначностей или «ловушек», которые могут приводить к неожиданным ошибкам у способных испытуемых. Дальнейшее исследование причин нестабильности эмпирических кривых и анализ содержания заданий могут помочь улучшить качество этих элементов и, как следствие, всего теста.

Обратимся к рисунку 5, где наглядно представлено функционирование категорий ответов заданий посредством дихотомических кривых.

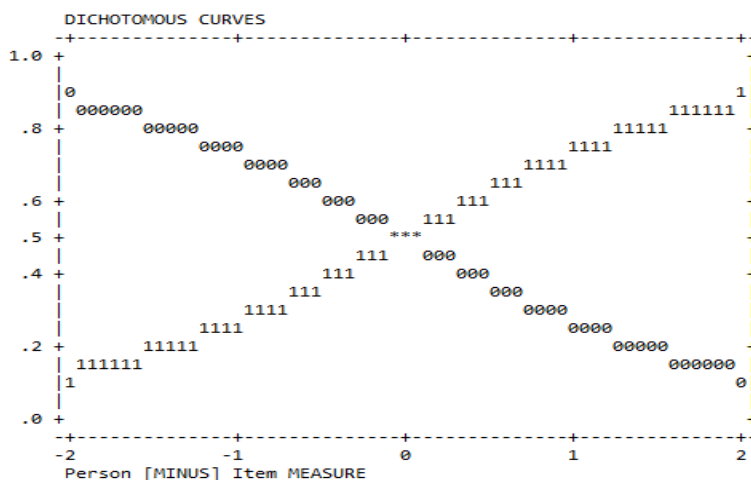


Рисунок 5. Дихотомические кривые

На графике видно, что обе линии имеют характерную S-образную форму. Это означает, что вероятность выбора ответа меняется постепенно, а не резко. Линия, идущая вверх справа, показывает вероятность выбора правильного ответа. Чем лучше подготовлен человек по сравнению со сложностью задания, тем выше вероятность, что он выберет правильный ответ. Линия, идущая вверх слева, показывает вероятность выбора неправильного ответа. Чем хуже подготовлен человек по сравнению со сложностью задания, тем выше вероятность, что он выберет неправильный ответ. Наклон линий в центре графика показывает, насколько хорошо задание различает людей с немного разным уровнем подготовки. Более крутые линии означают, что даже небольшая разница в уровне подготовки сильно влияет на вероятность выбора правильного ответа. В данном случае наклон умеренный, что характерно для используемой модели анализа, что значит, что задания в целом работают ожидаемым образом.

Для более наглядного представления распределения уровней способности респондентов относительно трудности элементов теста обратимся к карте элементов и респондентов, представленной на рисунке 6. Данная карта визуализирует расположение респондентов (#) и элементов (I) на одной и той же шкале измерения.

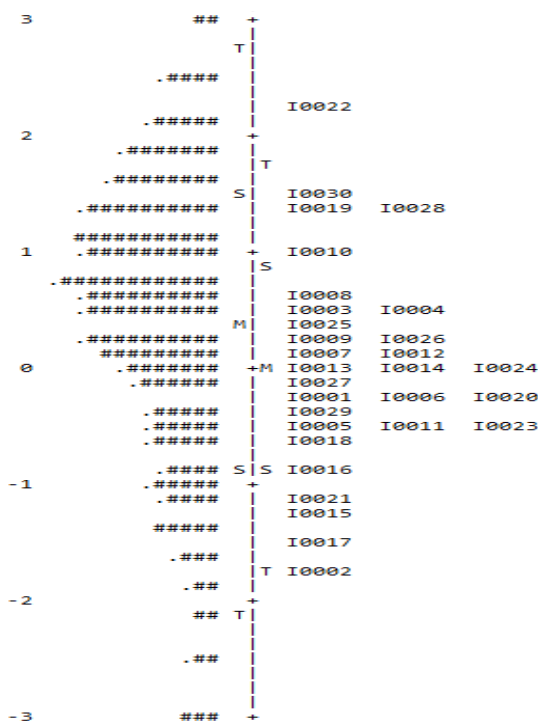


Рисунок 6. Карта переменных

Анализ распределения респондентов показывает, что основная масса выборки сосредоточена в диапазоне уровней способности от -1 до +1 логита, со средним значением, близким к 0. Распределение имеет тенденцию к нормальному, с незначительным смещением в сторону более низких значений. При этом отмечается наличие респондентов с как очень низкими, так и очень высокими уровнями способности, хотя их доля невелика. Распределение элементов охватывает более широкий диапазон трудности – от -2 до +3, что свидетельствует о разнообразии заданий по сложности. Наибольшая концентрация элементов наблюдается в центральной части шкалы (от -1 до +1), что совпадает с областью, где сосредоточено большинство респондентов. Такое соответствие обеспечивает максимальную информативность теста для основной части выборки.

Сопоставление распределения трудности элементов и способности респондентов указывает на общее хорошее соответствие. Большинство заданий ориентированы на уровень подготовки основной массы участников. Тем не менее, анализ карты позволяет выявить области шкалы, где

плотность элементов несколько ниже, например, в верхней части (выше +2 логит). Это может указывать на потенциальную целесообразность разработки дополнительных заданий, направленных на более точную оценку респондентов с очень высоким уровнем способности.

Заключение. Академические исследования в рамках магистерских программ играют ключевую роль в углублении знаний в выбранной области. В этом процессе критическое мышление выступает фундаментальным инструментом, необходимым для проведения экспериментов, анализа данных и формулирования обоснованных выводов при решении сложных научных проблем.

Проведенное исследование подтвердило эффективность разработанного теста для оценки критического мышления как инструмента определения уровня подготовленности претендентов на программы магистратуры. Тщательно разработанная структура теста, включающая разделы «Критическое мышление» и «Аналитическое мышление», позволяет комплексно оценить ключевые когнитивные компетенции, необходимые для успешного обучения и научно-исследовательской деятельности на послевузовском уровне. Следует также отметить, что первая научно обоснованная рекомендация по включению раздела «Аналитическое мышление» в структуру теста для поступающих в магистратуру и докторантуру была представлена академическому сообществу в статье «Оценка аналитического мышления поступающих в магистратуру и докторантуру: опыт Казахстана» [8]. В данной работе изложены результаты анализа разработки и оформления заданий по аналитическому мышлению, включая рекомендации по объему текста, принципам построения, содержанию и формулировке вопросов, а также по материалам общеинформационной части.

Психометрический анализ продемонстрировал достаточную дифференцирующую способность теста и его хорошую надежность, что подтверждается показателями среднеквадратичного отклонения и коэффициента надежности. Это свидетельствует о том, что разработанный тест является надежным инструментом для оценки критического мышления абитуриентов магистратуры. Вместе с тем, анализ отдельных тестовых заданий выявил необходимость дальнейшего изучения некоторых элементов, не полностью соответствующих модели Раша.

Раздел «Критическое мышление» в структуре комплексного тестирования выступает важным и эффективным компонентом, обеспечивающим объективную оценку готовности кандидатов к дальнейшему обучению и исследовательской деятельности.

Результаты исследования позволили сформулировать ряд рекомендаций по разработке и оформлению тестовых заданий, направленных на повышение их качества. Ключевыми аспектами являются:

Охват содержания. Разработка ТГО требует коллегиальной работы разработчиков, экспертов и специалистов по измерениям, включающей определение структуры теста, разработку и экспертную оценку заданий, апробацию и психометрический анализ. Тестовые задания были разработаны в формате задач на решение или определение достаточности данных для решения задач по алгебре и геометрии, а также задач на анализ данных. Содержание теста охватывает области: арифметика, алгебра, геометрия и анализ данных. При этом уровень сложности соответствует программе математики средней школы и основам статистики средней школы. Тест не включает задания по тригонометрии, сложному исчислению или другим разделам высшей математики и не проверяет способность к построению математических доказательств. Среди конкретных предложений было выделено усиление компонента задач на анализ данных и увеличение числа заданий, направленных на оценку навыков количественного анализа и интерпретации информации.

Терминология и символы. Используемые математические символы, терминология и условные обозначения должны соответствовать стандартному уровню средней школы.

Геометрические фигуры. Все фигуры должны быть расположены на плоскости, геометрические фигуры, такие как линии, круги, треугольники и квадраты, не следует рисовать в масштабе. Ответы на задачи, связанные с геометрическими фигурами, должны основываться на

геометрических рассуждениях, а не на визуальном измерении или сравнении величин. Точки, принадлежащие одной линии, должны располагаться в соответствии с указанным порядком. Геометрические объекты должны быть представлены в заданных относительных положениях.

Анализ данных. В заданиях на анализ данных основное внимание уделяется визуализации данных и пониманию представленной информации. Раздел содержит информацию в виде таблиц, графиков, диаграмм или текста. Предоставленная информация должна быть четкой и однозначной.

Типы заданий. Часть «Критическое мышление» ТГО включает задания на сравнение величин и задания с выбором одного правильного ответа, где каждое задание является независимым. Задания на сравнение величин требуют сравнения двух величин на основе арифметических, алгебраических и геометрических знаний. Задания с выбором одного правильного ответа предполагают выбор единственного верного ответа из пяти предложенных [4]. Задания должны быть сконструированы таким образом, чтобы исключить возможность нахождения правильного ответа простым подставлением предложенных вариантов в уравнение или неравенство, решение должно требовать применения логических рассуждений, математических методов и понимания концепций.

Формулировка заданий. Разработка качественных тестовых заданий является ключевым аспектом эффективной оценки знаний и навыков. Предлагаемые принципы построения тестовых заданий, включая принцип логичности содержания, основаны на фундаментальных работах в области педагогических измерений, в частности, на трудах В. Аванесова [13]:

- Задания должны быть изложены ясно, лаконично и допускать двойственного толкования.
- Задания следует формулировать в виде утверждений, истинность и ложность которых определяется ответом участника.
- Содержание заданий должно быть потенциально интересным для целевой аудитории, что может способствовать повышению мотивации и вовлеченности в процесс тестирования.
- Задания должны охватывать ключевые единицы знаний и навыков, необходимые для успешного освоения курса. Например, при оценке «Критическое мышление» задания должны быть направлены на проверку способности анализировать, интерпретировать информацию и делать обоснованные выводы.
- Содержание заданий должно основываться на проверенных и достоверных фактах, описывающих детали, наблюдения, взаимосвязи и характеристики рассматриваемых объектов или явлений. Например, утверждение «математический символ равенства « \Rightarrow » обозначает взаимосвязь между двумя идентичными по значению математическими выражениями» представляет собой достоверный факт.
- Задания следует разрабатывать с учетом последних достижений в соответствующей области науки и техники.
- Задания должны предоставлять возможности для поиска ответов на вопросы, быть самодостаточными и основываться на самом содержании задания, а не на предварительных знаниях участника.
- Задания могут включать тесты, иллюстрации, диаграммы, графики и таблицы, которые служат дополнительным источником информации для анализа и формирования выводов.
- Язык формулировок заданий должен быть простым и доступным для понимания. Ожидается, что ответы будут основываться на аналитических навыках, логике и общих знаниях, а не на глубокой специализации в конкретной узкой области.

Таким образом, тщательное формулирование содержания задания играет определяющую роль в процессе разработки эффективных тестовых инструментов, обеспечивающих объективную и всестороннюю оценку знаний и навыков участников.

Дистракторы. Дистрактором называется неправильный, но правдоподобный вариант ответа в заданиях с выбором одного или нескольких правильных ответов. Качественные дистракторы должны создаваться на основе следующих ключевых принципов:

- Все варианты ответа должны быть грамматически правильными, чтобы избежать непреднамеренных подсказок.
- Дистракторы должны быть правдоподобными и логичными в рамках рассматриваемого вопроса или предмета исследования, т.е. соответствовать контексту.
- Каждый дистрактор должен иметь продуманное содержание и должен быть разработан таким образом, чтобы его могли счесть правильным те, кто недостаточно владеет материалом.
- Варианты ответа должны быть сопоставимы по длине, формату, сложности и грамматической структуре.
- Следует избегать случайных отвлекающих факторов, т.е. не следует включать элементы, не имеющие логической связи с вопросом и способные запутать испытуемого.
- Следует избегать искажающих факторов, т.е. варианты ответа не должны содержать вводящие в заблуждение или двусмысленные элементы.
- Следует избегать повторений сути задачи, т.е. варианты ответа не должны прямо или косвенно повторять информацию из вопроса.
- Порядок дистракторов не должен подсказывать правильный ответ.
- Следует избегать повторяющихся шаблонов в структуре вариантов ответов, чтобы не создавать легко распознаваемых закономерностей.

Эти принципы являются основой для разработки тестов, которые обеспечивают более точную и объективную оценку знаний и навыков испытуемых. Важно, что результаты исследования подтверждают надежность и валидность используемых инструментов тестирования, что делает их ценным инструментом для оценки знаний и навыков абитуриентов магистратуры. Соблюдение предложенных принципов разработки и оформления заданий будет способствовать дальнейшему повышению качества оценочных процедур и, как следствие, улучшению качества подготовки кадров высшей квалификации в Республике Казахстан.

Список использованной литературы:

1. Концепция развития высшего образования и науки на 2023-2029 годы [Электронный ресурс]. URL: <https://adilet.zan.kz/rus/docs/P2300000248> (дата обращения: 19.03.2025).
2. Мы в Болонском процессе: достижения за 10 лет, 27.10.2020 г. [Электронный ресурс]. URL: <https://bilimdinews.kz/?p=124060> (дата обращения: 19.03.2025).
3. Послевузовское образование в Республике Казахстан (На начало 2024-2025 учебного года) [Электронный ресурс]. URL: <https://stat.gov.kz/ru/industries/social-statistics/stat-edu-science-inno/publications/196855/> (дата обращения: 06.02.2025).
4. Об утверждении Правил проведения комплексного тестирования: Приказ Министра образования и науки Республики Казахстан от 8 мая 2019 года № 190. [Электронный ресурс]. URL: <https://adilet.zan.kz/rus/docs/V1900018657> (дата обращения: 20.03.2025).
5. Что на самом деле значит VUCA и как преуспеть в VUCA-мире? 06.12.2022 г. [Электронный ресурс]. URL: <https://lpgenerator.ru/blog/chto-takoe-vuca> (дата обращения: 20.03.2025).
6. The top soft skills to develop by 2027: «Future of Jobs», World Economic Forum 2023 report, 30.04.2023 г. [Электронный ресурс]. URL: <https://www.coorpacademy.com/en/blog/learning-innovation-en/the-top-soft-skills-to-develop-by-2027-future-of-jobs-world-economic-forum-2023-report> (дата обращения: 02.04.2025).
7. Alsaleh N.J. Teaching Critical Thinking Skills: Literature Review// The Turkish Online Journal of Educational Technology. – 2020. - 19(1). – P.21-39.
8. Шинетова Л.Е., Алтыбаева Ш.Б., Бейсенова Ж.Ж., Новицкая Ю.В. Оценка аналитического мышления поступающих в магистратуру и докторантуру: опыт Казахстана // ал-Фараби атындағы Қазақ Ұлттық университетінің ХАБАРШЫСЫ, 2023. - 2(190). – С. 221-232. <https://doi.org/10.26577/EJPh.2023.v190.i2.ph21>.
9. Huber C.R., Kuncel N.R. Does College Teach Critical Thinking?// A Meta-Analysis. Review of Educational Research. – 2016. - 86(2). - P. 431-468.
10. Lai E.R. Collaboration: A Literature Review Research Report // Pearson Publisher. – 2011.
11. Facione P.A. Critical Thinking: What It Is and Why It Counts. Insight Assessment. - 2011 <https://www.insightassessment.com/wp-content/uploads/ia/pdf/whatwh>.
12. Эннис Р.Х. Критическое мышление: размышления и перспектива. Часть I // Исследование: критическое мышление в разных дисциплинах.– 2011. - 26(1). – С. 4-18.
13. Аванесов В.С. Теория и методика педагогических измерений Подготовлено ЦТ и МКО УГТУ-УПИ. – 2005. – 165 с.

14. Майоров А.Н. Теория и практика создания тестов для системы образования // Материалы курса «Основы теории и практики разработки тестов для оценки знаний школьников»: лекции 5-8. Педагогический университет «Первое сентября». – 2010. – 100 с.
15. Linn R. *Educational Measurement*. American Council on Education. – 2000. – 914 p.
16. Haladyna T. *Developing and Validating Multiple-choice Test Items*. Lawrence Erlbaum Associates. – 2004. – P. 296.
17. Gronlund N.E., Waugh C.K. *Assessment of Student Achievement*. Pearson Education // Издательство: Pearson Education. – 2009. – 408 p.
18. Анастаси А. *Психологическое тестирование: Книга 1* (К.М. Гуревич, В.И. Лубовский, Ред.; Пер. с англ.) // Педагогика. – 1982. – 318 p.
19. Карданова Е.Ю. *Моделирование и параметризация тестов: основы теории и приложения* // Москва: Федеральный центр тестирования. – 2008. – 272 с.
20. *Поступающим в магистратуру [Электронный ресурс]*. URL: <https://testcenter.kz/ru/postupayushchim-v-magistraturu-i-doktoranturu/vstupitelnye-ekzameny/informatsiya-ob-itogakh-vstupitelnogo-ekzamena-po-yazykam-na-poluchenie-pvo/index.php> (дата обращения: 06.04.2025).

References:

1. *Kontseptsiya razvitiya vysshego obrazovaniya i nauki na 2023-2029 gody [Elektronnyy resurs]*. URL: <https://adilet.zan.kz/rus/docs/P2300000248> (дата обращения: 19.03.2025).
2. *Мы в Болонском протсессе: dostizheniya za 10 let, 27.10.2020 g. [Elektronnyy resurs]*. URL: <https://bilimdinews.kz/?p=124060> (дата обращения: 19.03.2025).
3. *Poslevuzovskoye obrazovaniye v Respublike Kazakhstan (Na nachalo 2024-2025 uchebnogo goda) [Elektronnyy resurs]*. URL: <https://stat.gov.kz/ru/industries/social-statistics/stat-edu-science-inno/publications/196855/> (дата обращения: 06.02.2025).
4. *Ob utverzhenii Pravil provedeniya kompleksnogo testirovaniya: Prikaz Ministra obrazovaniya i nauki Respubliki Kazakhstan ot 8 maya 2019 goda № 190. [Elektronnyy resurs]*. URL: <https://adilet.zan.kz/rus/docs/V1900018657> (дата обращения: 20.03.2025).
5. *Chto na samom dele znachit VUCA i kak preuspet' v VUCA-mire? 06.12.2022 g. [Elektronnyy resurs]*. URL: <https://lpgenerator.ru/blog/chto-takoe-vuca> (дата обращения: 20.03.2025).
6. *The top soft skills to develop by 2027: «Future of Jobs», World Economic Forum 2023 report, 30.04.2023 g. [Elektronnyy resurs]*. URL: <https://www.coorpacademy.com/en/blog/learning-innovation-en/the-top-soft-skills-to-develop-by-2027-future-of-jobs-world-economic-forum-2023-report> (дата обращения: 02.04.2025).
7. Alsaleh N.J. *Teaching Critical Thinking Skills: Literature Review. The Turkish Online Journal of Educational Technology*. – 2020. – 19(1). – P. 21-39.
8. Shinetova L.Y., Altybayeva Sh.B., Beisenova Zh.Zh., Novitskaya Yu.V. *Otsenka analiticheskogo myshleniya postupayushchikh v magistraturu i doktoranturu: opyt Kazakhstana. // al-Farabi atyndaǵy Qazaq Yltyq universitetiniń KHABARSHYSY. - 2023. - 2(190). – S. 221-232. https://doi.org/10.26577/EJPh.2023.v190.i2.ph21.*
9. Huber C.R., Kuncel N.R. *Does College Teach Critical Thinking?//A Meta-Analysis. Review of Educational Research*, 2016. – 86(2). – P. 431-468.
10. Lai E.R. *Collaboration: A Literature Review Research Report* // Pearson Publisher. – 2011.
11. Facione P.A. *Critical Thinking: What It Is and Why It Counts. Insight Assessment, 2011. https://www.insightassessment.com/wp-content/uploads/ia/pdf/whatwh.*
12. Ennis R.Kh. *Kriticheskoye myshleniye: razmyshleniya i perspektiva. Chast' I. Issledovaniye: kriticheskoye myshleniye v raznykh distsiplinakh. Issledovaniye: kriticheskoye myshleniye v raznykh distsiplinakh. – 2011. – 26(1). – S. 4-18.*
13. Avanesov V.S. *Teoriya i metodika pedagogicheskikh izmereniy// Podgotovleno TsT i MKO UGTU-UPI. – 2005. – 165 s.*
14. Mayorov A.N. *Teoriya i praktika sozdaniya testov dlya sistemy obrazovaniya // Materialy kursa «Osnovy teorii i praktiki razrabotki testov dlya otsenki znaniy shkol'nikov»: lektsii 5-8. Pedagogicheskiy universitet «Pervoye sentyabrya». – 2010. – 100 s.*
15. Linn R. *Educational Measurement* // American Council on Education. – 2000. – 914 p.
16. Haladyna T. *Developing and Validating Multiple-choice Test Items* // Lawrence Erlbaum Associates. – 2004. – P. 296.
17. Gronlund N.E., Waugh C.K. *Assessment of Student Achievement* // Pearson Education. – 2009. – 408 p.
18. Anastaz, A. *Psikhologicheskoye testirovaniye: Kniga 1* (K.M. Gurevich, V.I. ubovskiy, Red.; Per. s angl.). // Pedagogika. – 1982. – 318 p.
19. Kardanova Ye.Yu. *Modelirovaniye i parametrizatsiya testov: osnovy teorii i prilozheniya* // Moskva: Federal'nyy tsentr testirovaniya. – 2008. – 272 s.
20. *Postupayushchim v magistraturu [Elektronnyy resurs]*. URL: <https://www.google.com/search?q=https://testcenter.kz/ru/postupayushchim-v-magistraturu-i-doktoranturu/vstupitelnye-ekzameny/informatsiya-ob-itogakh-vstupitel%27nogo-ekzamena-po-yazykam-na-poluchenie-pvo/index.php> (дата обращения: 06.04.2025).