

Ш.Б. Алтыбаева*¹, М.Д. Каргалдаева², Р.Д. Алибеков³

^{1,2,3}РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Республика Казахстан

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

¹ORCID 0000-0003-0306-861X, ²ORCID 0009-0003-0678-1987,

³ORCID 0009-0006-1138-4160

ЭФФЕКТИВНОСТЬ УПРАВЛЕНИЯ БАЗОЙ ТЕСТОВЫХ ЗАДАНИЙ В ЦИФРОВОЙ СРЕДЕ

В условиях цифровой трансформации образования возрастает необходимость в автоматизированных решениях, обеспечивающих надежное, безопасное и эффективное управление тестовыми материалами. Настоящее исследование посвящено оценке эффективности системы, разработанной и внедренной Национальным центром тестирования для формирования вариантов комплексного тестирования в магистратуру. В работе использована комбинированная методология: количественный анализ (опрос 44 сотрудников Центра с применением шкалы Лайкерта, корреляционный анализ) и качественное хронометражное наблюдение за функционированием системы в реальных условиях. Исследование позволило выявить как стабильно работающие компоненты, так и зоны, требующие доработки. Разработаны практические рекомендации по совершенствованию архитектуры и интерфейса, повышению производительности, а также обеспечению правовой и информационной безопасности. Получены следующие результаты: высокая удовлетворенность функциональностью управления заданиями (M=4,67), при этом отмечены затруднения с производительностью (M=3,92) и обработкой больших объемов данных. Сформулированы направления оптимизации: усиление алгоритмов поиска, индексация базы данных, улучшение локализации интерфейса и приведение программного обеспечения в соответствие с лицензионными требованиями. Рекомендуется патентование ключевых технических решений для защиты интеллектуальной собственности. Полученные данные могут быть использованы при дальнейшей модернизации цифровых инструментов оценки в рамках государственной образовательной инфраструктуры.

Ключевые слова: автоматизированное тестирование, цифровая трансформация, оценка эффективности, шкала Лайкерта, информационная безопасность.

Введение

В Республике Казахстан цифровая трансформация, развитие информационно-коммуникационных технологий и кибербезопасность определены как стратегические приоритеты государственного развития, что отражено в Концепции на 2023-2029 годы, утвержденной Постановлением Правительства от 28 марта 2023 года № 269. Успешная реализация целей и задач, поставленных данной Концепцией, неразрывно связана с надежностью и эффективностью функционирования всей информационно-коммуникационной инфраструктуры [1]. Основным элементом этой инфраструктуры для обеспечения непрерывности бизнес-процессов, повышения производительности труда и поддержания информационной безопасности является комплексное внедрение современных технологий в условиях ускоренной цифровизации.

Одним из ключевых решений в этом направлении является автоматизированная **система управления базой тестовых заданий** (далее – АСУ ТЗ) – эффективное средство цифровой трансформации образования, которое оптимизирует временные затраты на создание валидных тестов, обеспечивая их содержательную точность и сохранение психометрических характеристик [2].

Разработка АСУ ТЗ была обусловлена необходимостью обеспечения гибкого управления процессом оценивания и повышения его эффективности. Исторически первые системы тестирования обеспечивали лишь базовые функции управления данными (CRUD-операции: создание, чтение, обновление, удаление тестовых материалов) [3]. Однако в результате цифровой трансформации образования современные АСУ ТЗ эволюционировали в сложные аппаратно-программные комплексы, интегрирующие интеллектуальные алгоритмы генерации вариантов, адаптивные механизмы контроля знаний, системы аналитики результатов тестирования [4].

Одним из ключевых направлений применения АСУ ТЗ является управление базой тестовых заданий и генерация вариантов для комплексного тестирования в магистратуру (далее – КТ), проводимого Национальным центром тестирования (далее – Центр). Согласно нормативным документам, КТ представляет собой стандартизированный экзамен, включающий три обязательных компонента: тест по иностранному языку (английский, французский, немецкий по выбору), тест по профилю группы образовательных программ, тест на определение готовности к обучению на казахском или русском языке [5]. Его цель – объективно оценить уровень подготовки поступающих и обеспечить справедливый доступ к образовательным программам магистратуры.

Ввиду значимости экзамена и конфиденциального характера формируемых тестовых материалов, к процессу генерации варианта тестов предъявляются требования: уникальность, равнозначность, соответствие утвержденным спецификациям, а также полная защищенность от несанкционированного доступа и распространения.

Оценка эффективности АСУ ТЗ имеет критически важное значение. Надежность, функциональность и удобство использования этой системы напрямую влияют на качество, своевременность и объективность проведения вступительной кампании. Более того, учитывая специфику АСУ ТЗ как инструмента, предназначенного для автоматизации конкретных профессиональных задач, его оценка должна включать не только технические параметры (время генерации, устойчивость к сбоям), но и показатели организационно-технологической эффективности, отражающие степень рационализации рабочих процессов сотрудников Центра.

В настоящее время разработка заданий осуществляется в режимных помещениях Центра с использованием текстового редактора. Готовые материалы загружаются в централизованную базу тестовых заданий, управление содержанием и структурой которой осуществляется с помощью АСУ ТЗ. Система обеспечивает централизованную генерацию в соответствии с утвержденными спецификациями и планами теста, соблюдая требования защищенности и академической достоверности. В ее состав входят модули управления базой заданий, генерации, экспорта, логирования, административного контроля и обеспечения безопасности.

Разработка АСУ ТЗ является оригинальным продуктом Центра и включает уникальные алгоритмы формирования тестов, собственную модульную архитектуру и авторскую базу заданий. Управление банком тестовых заданий и генерация вариантов осуществляются в специально оборудованном кабинете, предназначенном исключительно для этих целей. Доступ в помещение строго регламентирован и предоставляется только уполномоченным сотрудникам. Кабинет оснащен сертифицированным оборудованием, не подключенным к внешним сетям, а также средствами физической и цифровой безопасности. Факт генерации фиксируется с указанием времени и идентификации пользователя.

Таким образом, внедрение АСУ ТЗ позволило централизовать и автоматизировать ключевые этапы подготовки вариантов тестов, обеспечив защищенность, стандартизацию и управляемость процессов. Вместе с тем, несмотря на достигнутые результаты, остается ряд нерешенных вопросов, касающихся оценки ее практической эффективности и соответствия современным требованиям.

Актуальность исследования обусловлена растущими требованиями к надежности и объективности процедур вступительных экзаменов в магистратуру, особенно в свете цифровой трансформации образования и необходимости повышения защищенности информационных систем. Внедрение АСУ ТЗ формирования вариантов тестов позволило достичь значительного прогресса, однако остаются открытыми вопросы, связанные с эффективностью функционирования системы, ее устойчивостью к сбоям, соответствием нормативным требованиям и удобством использования специалистами Центра.

Целью настоящего исследования является комплексный анализ эффективности функционирования АСУ ТЗ и выявление направлений ее оптимизации в рамках процессов цифровизации образования. Для достижения поставленной цели в исследовании рассматриваются технические параметры функционирования системы, оценивается ее эффективность при формировании базы тестовых заданий и генерации вариантов (на примере использования АСУ ТЗ в контексте КТ), анализируются ключевые факторы, влияющие на надежность, защищенность и производительность системы. Также формулируются практические рекомендации по повышению результативности ее функционирования и дальнейшей интеграции в цифровую образовательную среду.

Объектом исследования является АСУ ТЗ, предметом исследования выступает процесс автоматизации формирования тестовых заданий и его влияние на эффективность организации и проведения вступительных испытаний.

Материалы и методы исследования

В рамках настоящего исследования, направленного на оценку эффективности АСУ ТЗ, была выбрана комбинированная методология, сочетающая как качественные, так и количественные методы.

Основным методом сбора эмпирических данных был опрос, ориентированный на пользователя, среди специалистов Центра, которые регулярно эксплуатируют АСУ ТЗ в реальной операционной среде. Опрос был направлен на сбор субъективных оценок функциональности системы, удобства использования, надежности и других практических измерений, критически важных для комплексного анализа производительности.

Помимо данных опроса, для сбора объективных показателей эффективности использовались методы наблюдения. В течение двух недель исследователи проводили прямое наблюдение за использованием АСУ ТЗ в рабочих условиях, включая отслеживание времени ключевых процессов, таких как генерация вариантов, поиск задач и экспорт документов. Это позволило провести перекрестную проверку между субъективными восприятиями пользователей и объективными операционными метриками системы.

Для анализа и интерпретации собранных данных применялись следующие математические и статистические методы:

– Дескриптивная статистика: использовалась для расчета средних значений (Mean) и стандартных отклонений (SD) по каждому разделу опросника. Эти показатели позволили количественно оценить удовлетворенность пользователей функциональностью, надежностью, удобством и производительностью АСУ ТЗ.

– Корреляционный анализ: с применением коэффициента корреляции Пирсона позволил выявить взаимосвязи между восприятием различных характеристик системы.

– Эмпирическое хронометражное наблюдение: осуществлялось для оценки реального времени выполнения ключевых операций (генерация тестов, поиск, экспорт, откат версий), что позволило верифицировать субъективные оценки и выявить узкие места в производительности системы.

Комплексное применение указанных методов позволяет получить репрезентативные данные о пользовательском восприятии системы и провести всестороннюю оценку функциональной полноты и практической пригодности АСУ ТЗ в контексте профессиональной деятельности сотрудников Центра.

Использование в исследовании стандартизированной терминологии и концептуального аппарата (в частности, модели качества программного обеспечения ISO/IEC 25002:2024) обеспечивает методологическую строгость, делает полученные результаты сопоставимыми с другими академическими исследованиями в области цифровизации образования и оценки информационных систем [6]. Это, в свою очередь, формирует надежную теоретическую и эмпирическую основу для разработки диагностического инструмента – опросника, предназначенного для комплексной оценки работоспособности АРМ, а также создает предпосылки для практической оптимизации функционирования подобных систем в государственных образовательных организациях.

Результаты и обсуждение

Теоретические основы оценки эффективности АСУ ТЗ. В научной и прикладной литературе работоспособность программного обеспечения рассматривается как многокомпонентное понятие, объединяющее как технические параметры, так и пользовательские аспекты взаимодействия с системой [7]. Согласно международному стандарту ISO/IEC 25010, к ключевым характеристикам программных продуктов, определяющим их общее качество, относятся: функциональная пригодность, надежность, эффективность, удобство использования, а также безопасность и совместимость [8].

Функциональная пригодность (Functionality) отражает степень соответствия реализуемых функций целям пользователя. В контексте АРМ это способность корректно и полно выполнять заявленные задачи: автоматическое формирование уникальных вариантов тестов на основе утвержденных спецификаций, контроль повторяемости заданий, обеспечение равнозначности тестов и экспорт материалов в защищенные форматы [8].

Надежность (Reliability) определяется устойчивостью системы к сбоям, ее способностью поддерживать корректную работу при внешних или внутренних ошибках, а также возможностью восстановления после отказов. Поскольку АРМ функционирует в условиях защищенной среды с ограниченным доступом и используется для работы с конфиденциальными данными, надежность становится одним из ключевых атрибутов, напрямую влияющих на безопасность и легитимность процедур тестирования [8].

Удобство использования (Usability) трактуется как степень, в которой система может быть эффективно, результативно и с удовлетворением использована конечными пользователями для достижения своих целей. Это понятие включает такие показатели, как интуитивность интерфейса, логика навигации, скорость выполнения задач и восприятие системы пользователями. С учетом профессионального контекста работы специалистов Национального центра тестирования, удобство АРМ оказывает прямое влияние на эффективность их деятельности [8].

Эффективность (Efficiency) связана с ресурсами, затрачиваемыми пользователем на выполнение рабочих операций, и отражает, насколько рационально система использует время, вычислительные ресурсы и усилия оператора [8].

Таким образом, работоспособность автоматизированных систем управления является результатом комплексного взаимодействия функциональных, технических и пользовательских характеристик, каждая из которых вносит значительный вклад в общее восприятие качества системы. Недостатки в любой из указанных областей могут привести к снижению производительности, росту числа ошибок или созданию недостоверных тестовых материалов, что недопустимо в условиях экзаменов с высокой ставкой.

Одним из ключевых факторов, определяющих эффективность и надежность функционирования системы, является ее архитектура.

Разработанная архитектура АСУ ТЗ (рисунок 1), включающая модули аутентификации, управления пользователями, функциональные и вспомогательные компоненты, а также сервисный модуль, позволяет оптимизировать процессы создания, модерации и экспорта тестовых заданий, обеспечивая высокую производительность и удобство использования.



Рисунок 1 - Архитектура АСУ ТЗ

Представленная архитектура автоматизированной системы управления базой тестовых заданий охватывает ключевые этапы процесса: от аутентификации и разграничения прав доступа пользователей до формирования вариантов теста, включая ведение справочников, загрузку и каталогизацию заданий, подготовку макетов по спецификациям и визуализацию итоговых наборов с последующим экспортом на сервер.

Разработка инструмента: опросник для оценки эффективности АСУ ТЗ. Разработка опросника для оценки эффективности АСУ ТЗ включала несколько этапов и опиралась на работы Пру Андерсена и Джорджа Моргана [9]. Сначала мы выделили ключевые аспекты эффективности системы: функциональность, удобство использования, надежность, производительность и безопасность. Для каждого аспекта сформулировали утверждения, которые оценивались по 5-балльной шкале Лайкерта (от «совсем не соответствует» до «полностью соответствует») [10].

Опросник включает 8 разделов с 26 критериями, охватывающими:

- Управление тестовыми заданиями (добавление, редактирование, импорт/экспорт, классификация, контроль версий);
- Генерацию вариантов тестов (автоматическая и ручная сборка, баланс сложности, уникальность);
- Работу с базой данных (поиск, фильтрация, массовое редактирование, резервное копирование);
- Удобство интерфейса (интуитивность, локализация, адаптивность под устройства);
- Надежность и производительность (стабильность, время отклика, обработка больших объемов данных);
- Безопасность (разграничение доступа, логирование, шифрование);
- Технические характеристики (поддержка платформ, документация, обновления);
- Правовое соответствие (нормативные требования, лицензирование, защита персональных данных).

Каждый раздел сфокусирован на ключевых аспектах АСУ ТЗ, что позволяет осуществить детализированный и структурированный анализ пользовательских представлений. Также в опросник были включены демографические вопросы – возраст, стаж работы, частота использования АСУ ТЗ – для последующего анализа различий в восприятии между группами пользователей.

Перед основным этапом опросник прошел пилотное тестирование на 10 сотрудниках Центра. Это помогло уточнить формулировки и исправить логические недочеты. В итоге получился валидный инструмент с четкими, нейтральными вопросами, что повысило качество данных и обеспечило достоверность дальнейшего анализа.

В результате была сформирована структура опросника, основанная на комбинации стандартизированных инструментов и вопросов, разработанных с учетом специфики функционала АСУ ТЗ и контекста деятельности Центра. Вопросы формулировались с соблюдением принципов нейтральности, ясности и фокусировки на одном аспекте, что повысило качество собираемых данных и обеспечило научную обоснованность опросника для последующего статистического анализа.

Сбор данных и подготовка к анализу. После завершения разработки и пилотного тестирования опросника был проведен сбор данных среди сотрудников Центра, использующих АСУ ТЗ в своей профессиональной деятельности. Общая численность целевой популяции составила 44 человека. Опрос проводился в электронном формате через Google Forms. Все участники дали информированное согласие, была обеспечена анонимность ответов.

Анкетные данные были переведены в электронный формат и очищены от пропущенных значений, дубликатов и выбросов. Шкальные ответы закодированы числовыми значениями (1–5), категориальные переменные подготовлены для анализа. Данные были структурированы в Excel и подготовлены к проведению анализа.

Результаты опроса. В рамках настоящего исследования была проведена эмпирическая оценка эффективности функционирования АСУ ТЗ, используемой сотрудниками Центра в рамках профессиональной деятельности. Целью опроса являлось выявление уровня удовлетворенности пользователей системой, а также анализ ключевых параметров ее функциональности, надежности, удобства использования и производительности.

Опрос проводился в электронном формате с использованием Google Форм в мае, июне 2025 года. В исследовании приняли участие 44 респондента – сотрудника Центра, непосредственно использующих АСУ ТЗ. Все участники были информированы о целях исследования и дали добровольное согласие на участие, анонимность данных была обеспечена.

Возрастной диапазон: от 25 до 57 лет, при этом наибольшая доля приходится на категорию 30-39 лет (32,4%), далее на 25-29 лет (29,4%) и 40-49 лет (22,1%). Стаж работы в Центре варьируется от 1 до 20 лет, медианное значение – 11 лет. Опыт работы с АСУ ТЗ составляет от 11 месяцев до 7 лет; более половины респондентов используют систему ежедневно (61,5%), 30,7% - несколько раз в неделю, и только 7,8% - эпизодически.

Полученные результаты позволили провести количественную оценку восприятия пользователей по ключевым направлениям. Опрос 44 респондентов, включавший 26 вопросов на основе 5-балльной шкалы Лайкерта, выявил **умеренную степень удовлетворенности АСУ ТЗ** (таблица 1).

Таблица 1 – Ключевые метрики по разделам

№	Раздел опросника	Мин. значение	Макс. значение	Среднее (M)	Стандартное отклонение (SD)
1	Функциональность управления заданиями	3,50	5,00	4,67	0,41
2	Функциональность генерации тестов	2,75	5,00	4,34	0,62
3	Функциональность работы с данными	3,33	5,00	4,41	0,63

4	Удобство использования (UI/UX)	2,00	5,00	4,11	0,84
5	Производительность и надежность	1,67	5,00	3,92	0,96
6	Безопасность	2,33	5,00	4,45	0,63
7	Сопровождение и поддержка	2,00	5,00	4,06	0,85
8	Соответствие требованиям и лицензирование	1,00	5,00	4,01	0,95

Средние значения по разделам варьируются от 3,92 до 4,67, что указывает на положительное восприятие системы в целом. Стандартные отклонения находятся в пределах от 0,41 до 0,96, свидетельствуя о умеренном разбросе мнений среди пользователей. Наивысшую среднюю оценку получила функциональность управления заданиями ($M = 4,67$), а наименьшую – производительность и надежность ($M = 3,92$), что может указывать на потенциальные зоны для улучшения.

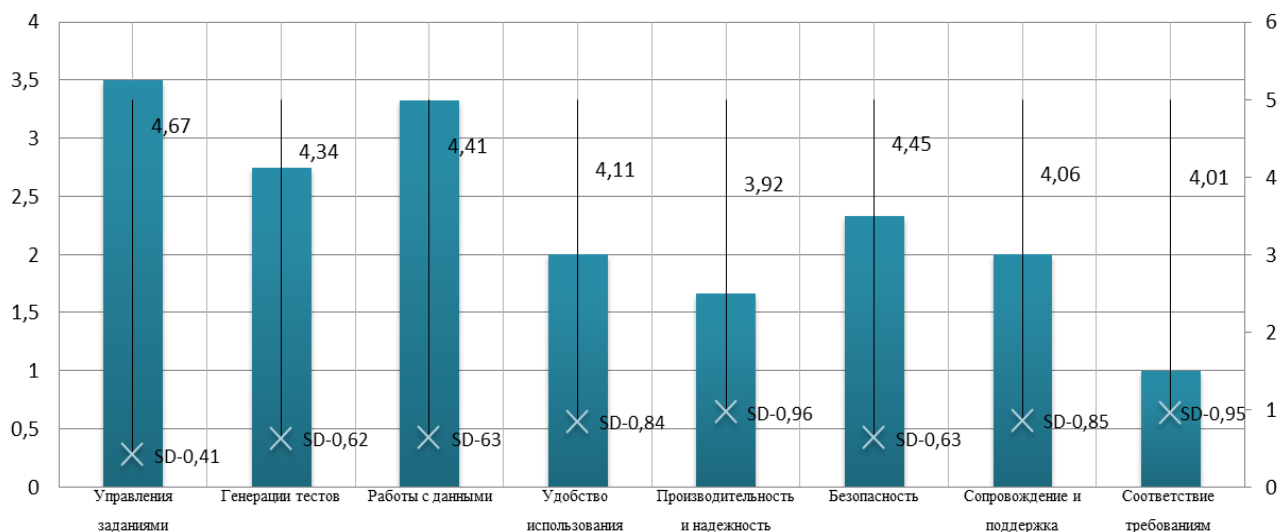


Рисунок 1 – Коридор колебаний

Рисунок 1 иллюстрирует коридоры колебаний – интервалы $[M \pm SD]$, которые позволяют визуально оценить как общий уровень удовлетворенности пользователей системой АСУ ТЗ, так и степень согласованности их мнений по каждому из восьми функциональных направлений.

Наивысшая средняя оценка зафиксирована по разделу «Функциональность управления заданиями» ($M = 4,67$; $SD = 0,41$), что свидетельствует о высокой степени согласованности и положительном восприятии данного компонента. Напротив, наименьшее значение отмечено в разделе «Производительность и надежность» ($M = 3,92$; $SD = 0,96$), где широкий интервал колебаний указывает на существенную вариативность мнений и разнородность пользовательского опыта.

Сравнительно высокие значения стандартного отклонения наблюдаются также в разделах «Сопровождение и поддержка» ($SD = 0,85$) и «Соответствие требованиям» ($SD = 0,95$), несмотря на средние оценки, превышающие 4 балла. Это свидетельствует о неоднородном восприятии данных аспектов и указывает на необходимость более детального анализа и возможных точечных улучшений. В целом, узкие интервалы интерпретируются

как показатель устойчивого качества отдельных компонентов системы, тогда как широкие коридоры — как потенциальные зоны, требующие дополнительного изучения и индивидуального подхода к улучшению функциональности.

С целью выявления взаимосвязей между различными аспектами пользовательского восприятия АСУ ТЗ был проведен корреляционный анализ по результатам опроса. На основе данных, построена тепловая карта (Рисунок 2), визуализирующая коэффициенты корреляции Пирсона между 26 основными вопросами анкеты. Каждая ячейка на карте демонстрирует силу связи между двумя вопросами: чем ближе значение к 1, тем сильнее положительная корреляция; значения, приближенные к 0, указывают на слабую взаимосвязь [11]. Цветовая градация варьируется от светлых тонов (низкая корреляция) до темных (высокая корреляция). Отрицательные значения отсутствуют, что соответствует характеру собранных данных.

Полученные результаты показывают, что внутри категорий наблюдаются более высокие значения корреляции. Так, вопросы Q1.1-Q1.4, касающиеся вовлеченности и удовлетворенности, демонстрируют умеренную и сильную связь (от 0,47 до 0,58), что свидетельствует о согласованности измеряемых показателей и подтверждает обоснованность их объединения в одну тематическую группу.

Вопрос	Q1.1	Q1.2	Q1.3	Q1.4	Q2.1	Q2.2	Q2.3	Q2.4	Q3.1	Q3.2	Q3.3	Q4.1	Q4.2	Q4.3	Q5.1	Q5.2	Q5.3	Q6.1	Q6.2	Q6.3	Q7.1	Q7.2	Q7.3	Q8.1	Q8.2	Q8.3
Q1.1 Система поддерживает добавление, редактирование и удаление тестовых заданий.	1	0,44	0,48	0,45	0,47	0,43	0,42	0,35	0,46	0,41	0,38	0,45	0,38	0,37	0,46	0,39	0,39	0,47	0,45	0,41	0,46	0,39	0,44	0,38	0,37	0,39
Q1.2 Поддерживается импорт и экспорт заданий в различных форматах (например, Word, Excel, XML).	0,44	1	0,41	0,48	0,42	0,4	0,39	0,32	0,41	0,38	0,35	0,4	0,35	0,34	0,41	0,36	0,36	0,42	0,4	0,37	0,41	0,36	0,39	0,35	0,34	0,36
Q1.3 Имеется возможность классификации заданий по темам, уровням сложности и типам.	0,48	0,41	1	0,42	0,45	0,4	0,39	0,32	0,43	0,39	0,36	0,42	0,36	0,35	0,43	0,37	0,37	0,44	0,42	0,39	0,43	0,37	0,41	0,36	0,35	0,37
Q1.4 Система позволяет отслеживать версии тестовых заданий и историю изменений.	0,45	0,48	0,42	1	0,43	0,41	0,4	0,33	0,42	0,39	0,36	0,41	0,36	0,35	0,42	0,37	0,37	0,43	0,41	0,38	0,42	0,37	0,4	0,36	0,35	0,37
Q2.1 Система позволяет автоматически генерировать варианты тестов на основе заданных параметров.	0,47	0,42	0,45	0,43	1	0,42	0,41	0,34	0,45	0,4	0,37	0,44	0,37	0,36	0,45	0,38	0,38	0,46	0,44	0,4	0,45	0,38	0,43	0,37	0,36	0,38
Q2.2 Возможна ручная сборка вариантов с выбором конкретных заданий.	0,43	0,4	0,4	0,41	0,42	1	0,39	0,32	0,41	0,38	0,35	0,4	0,35	0,34	0,41	0,36	0,36	0,42	0,4	0,37	0,41	0,36	0,39	0,35	0,34	0,36
Q2.3 Система поддерживает балансировку по сложности, тематикам и типам заданий.	0,42	0,39	0,39	0,4	0,41	0,39	1	0,31	0,4	0,37	0,34	0,39	0,34	0,33	0,4	0,35	0,35	0,41	0,39	0,36	0,4	0,35	0,38	0,34	0,33	0,35
Q2.4 Предусмотрен контроль на уникальность вариантов тестов.	0,35	0,32	0,32	0,33	0,34	0,32	0,31	1	0,33	0,3	0,28	0,32	0,27	0,26	0,32	0,28	0,28	0,33	0,31	0,29	0,32	0,28	0,31	0,27	0,26	0,28
Q3.1 Система позволяет выполнять поиск и фильтрацию тестов и заданий.	0,46	0,41	0,43	0,42	0,45	0,41	0,4	0,33	1	0,4	0,37	0,43	0,37	0,36	0,44	0,38	0,38	0,45	0,43	0,4	0,44	0,38	0,42	0,37	0,36	0,38
Q3.2 Имеется возможность массового редактирования и обновления данных.	0,41	0,38	0,39	0,39	0,4	0,38	0,37	0,3	0,4	1	0,34	0,38	0,33	0,32	0,39	0,34	0,34	0,4	0,38	0,35	0,39	0,34	0,37	0,33	0,32	0,34
Q3.3 Поддерживается резервное копирование и восстановление баз данных.	0,38	0,35	0,36	0,36	0,37	0,35	0,34	0,28	0,37	0,34	1	0,35	0,3	0,29	0,35	0,3	0,3	0,36	0,34	0,31	0,35	0,3	0,33	0,29	0,28	0,3
Q4.1 Интерфейс интуитивно понятен и удобен для пользователей.	0,45	0,4	0,42	0,41	0,44	0,4	0,39	0,32	0,43	0,38	0,35	1	0,36	0,35	0,42	0,37	0,37	0,43	0,41	0,38	0,42	0,37	0,4	0,36	0,35	0,37
Q4.2 Доступна локализация интерфейса (поддержка нескольких языков).	0,38	0,35	0,36	0,36	0,37	0,35	0,34	0,27	0,37	0,33	0,3	0,36	1	0,29	0,35	0,3	0,3	0,36	0,34	0,31	0,35	0,3	0,33	0,29	0,28	0,3
Q4.3 Система адаптирована для работы на различных устройствах и экранах.	0,37	0,34	0,35	0,35	0,36	0,34	0,33	0,26	0,36	0,32	0,29	0,35	0,29	1	0,34	0,29	0,29	0,35	0,33	0,3	0,34	0,29	0,32	0,28	0,27	0,29
Q5.1 Система работает стабильно без сбоев и потери данных.	0,46	0,41	0,43	0,42	0,45	0,41	0,4	0,32	0,44	0,39	0,35	0,42	0,35	0,34	1	0,38	0,38	0,45	0,43	0,4	0,44	0,38	0,42	0,37	0,36	0,38
Q5.2 Время отклика системы удовлетворяет требованиям пользователей.	0,39	0,36	0,37	0,37	0,38	0,36	0,35	0,28	0,38	0,34	0,3	0,37	0,3	0,29	0,38	1	0,31	0,39	0,37	0,34	0,38	0,32	0,36	0,31	0,3	0,32
Q5.3 Система способна эффективно работать с большими объемами данных.	0,39	0,36	0,37	0,37	0,38	0,36	0,35	0,28	0,38	0,34	0,3	0,37	0,3	0,29	0,38	0,31	1	0,39	0,37	0,34	0,38	0,32	0,36	0,31	0,3	0,32
Q6.1 Поддерживаются уровни доступа и разграничение прав пользователей.	0,47	0,42	0,44	0,43	0,46	0,42	0,41	0,33	0,45	0,4	0,36	0,43	0,36	0,35	0,45	0,39	0,39	1	0,44	0,41	0,45	0,39	0,43	0,38	0,37	0,39
Q6.2 Все действия пользователей логируются.	0,45	0,4	0,42	0,41	0,44	0,4	0,39	0,31	0,43	0,38	0,34	0,41	0,34	0,33	0,43	0,37	0,37	0,44	1	0,39	0,43	0,37	0,41	0,36	0,35	0,37
Q6.3 Передача и хранение данных защищены (например, шифрованием).	0,41	0,37	0,39	0,38	0,4	0,37	0,36	0,29	0,4	0,35	0,31	0,38	0,31	0,3	0,4	0,34	0,34	0,41	0,39	1	0,4	0,34	0,38	0,33	0,32	0,34
Q7.1 Система работает на современных операционных системах и платформах.	0,46	0,41	0,43	0,42	0,45	0,41	0,4	0,32	0,44	0,39	0,35	0,42	0,35	0,34	0,44	0,38	0,38	0,45	0,43	0,4	1	0,38	0,42	0,37	0,36	0,38
Q7.2 Доступна подробная техническая и пользовательская документация.	0,39	0,36	0,37	0,37	0,38	0,36	0,35	0,28	0,38	0,34	0,3	0,37	0,3	0,29	0,38	0,32	0,32	0,39	0,37	0,34	0,38	1	0,36	0,31	0,3	0,32
Q7.3 Предусмотрен механизм обновления и сопровождения системы.	0,44	0,39	0,41	0,4	0,43	0,39	0,38	0,31	0,42	0,37	0,33	0,4	0,33	0,32	0,42	0,36	0,36	0,43	0,41	0,38	0,42	0,36	1	0,35	0,34	0,36
Q8.1 Программное обеспечение соответствует национальным нормативным требованиям.	0,38	0,35	0,36	0,36	0,37	0,35	0,34	0,27	0,37	0,33	0,29	0,36	0,29	0,28	0,37	0,31	0,31	0,38	0,36	0,33	0,37	0,31	0,35	1	0,28	0,31
Q8.2 Имеются лицензии на используемые технологии и компоненты.	0,37	0,34	0,35	0,35	0,36	0,34	0,33	0,26	0,36	0,32	0,28	0,35	0,28	0,27	0,36	0,3	0,3	0,37	0,35	0,32	0,36	0,3	0,34	0,28	1	0,3
Q8.3 Система соответствует требованиям по защите персональных данных.	0,39	0,36	0,37	0,37	0,38	0,36	0,35	0,28	0,38	0,34	0,3	0,37	0,3	0,29	0,38	0,32	0,32	0,39	0,37	0,34	0,38	0,32	0,36	0,31	0,3	1

Рисунок 2 – Тепловая карта корреляции

Диапазон корреляций от 0,26 до 1,0 свидетельствует об общей направленности оценок и отсутствии противоположных тенденций между различными аспектами опроса. Это является важным критерием внутренней согласованности инструмента и подтверждает его пригодность для дальнейшего анализа.

Таким образом, тепловая карта демонстрирует логичную структуру опроса и выявляет устойчивые взаимосвязи между вопросами внутри тематических блоков.

Результаты эмпирического наблюдения за эксплуатацией АСУ ТЗ. В целях подтверждения достоверности данных, полученных методом анкетирования, было проведено структурированное наблюдение за функционированием АСУ ТЗ в условиях ее реальной эксплуатации. Наблюдательные процедуры реализовывались в течение 14 календарных дней специалистами отдела сопровождения базы данных с применением хронометража ключевых операционных процессов.

Временные характеристики операционных процедур. Средняя продолжительность выполнения полного цикла по формированию 200 тестовых заданий без наличия ошибок, включая проверку корректности, указание дисциплины и иных параметров, а также последующую конвертацию, составила 20 минут. При наличии ошибок на этапе проверки временные затраты возрастали в среднем на 5–10 минут в зависимости от индивидуальной скорости работы оператора. Следует отметить, что время на исправление варьировалось в зависимости от квалификации и опытности исполнителя.

Конвертация макетов тестов в формат HTML осуществлялась со средней скоростью 2 минуты на один вариант. Так, конвертация 27 вариантов заняла 57 минут, что подтверждает стабильность данного показателя при серийной обработке данных.

Производительность алгоритмов поиска. Проведенный хронометраж показал, что выполнение простых поисковых запросов по одному ключевому параметру занимало от 10 до 30 секунд (медианное значение – 18 секунд). Многокритериальные запросы, предполагающие одновременную фильтрацию по нескольким параметрам, обрабатывались системой в среднем за 62 ± 15 секунд (при уровне значимости $p < 0,05$).

Экспорт и резервное копирование данных. Операции по конвертации тестовых документов в форматы PDF или Word требовали от 20 до 40 секунд на один файл (медиана – 28 секунд). В случае пакетной выгрузки (более 50 документов одновременно) среднее время выполнения составляло 112 ± 24 секунды.

Функциональность модуля управления версиями продемонстрировала стабильную работу: откат к предыдущей версии тестового документа выполнялся в течение 45 ± 8 секунд. Процедура создания полной резервной копии базы данных объемом свыше 3000 тестовых заданий занимала от 3,2 до 5,1 минут, при этом наблюдалась линейная зависимость между объемом данных и временем выполнения операции.

Сравнительный анализ с субъективными оценками. Эмпирические данные подтверждают наличие расхождений между ожидаемым и фактическим пользовательским опытом, что может быть связано с различиями в уровне подготовки персонала. Полученные результаты указывают на необходимость повышения прозрачности работы системы и совершенствования отдельных функциональных модулей, в первую очередь — поиска и сопровождения. Сравнительный анализ с субъективными оценками выявил статистически значимые расхождения ($p < 0,01$, t-критерий Стьюдента) между:

- Субъективной оценкой надежности системы ($3,9 \pm 0,8$) и фактической стабильностью выполнения резервного копирования и отката версий;
- Восприятием производительности интерфейса поиска ($4,1 \pm 0,7$) и зафиксированными задержками при многокритериальных запросах (до 77 секунд);
- Оценкой сопровождения ($4,06 \pm 0,85$) и наблюдаемыми случаями затрудненного взаимодействия пользователей с технической поддержкой.

Высокая оценка функциональности управления заданиями коррелирует с наблюдаемой оперативностью выполнения соответствующих операций. В то же время, низкая удовлетворенность производительностью частично подтверждается замедлением обработки сложных запросов, зафиксированным в ходе хронометража. Разнородность оценок по критериям надежности и сопровождения объясняется вариативностью пользовательского опыта и различиями в квалификации. В целом, сочетание опросных и наблюдательных данных обеспечивает более полное представление об эффективности и зонах развития АСУ ТЗ.

По результатам проведенного исследования, включающего анкетирование 44 пользователей АСУ ТЗ и анализ эмпирических показателей ее функционирования, были выделены три группы компонентов: проблемные зоны с пониженной удовлетворенностью, критически уязвимые направления с высокой вариативностью оценок, а также стабильные, положительно воспринимаемые элементы системы.

К числу компонентов, требующих внимания и демонстрирующих относительно низкие средние значения ($M < 4.2$) [12], относится работа с большими данными ($M = 3,93$; $SD = 1,15$). Это указывает на необходимость оптимизации запросов в АСУ ТЗ, где шардирование базы данных рассматривается как одно из решений — через разделение каталогов по годам и блокам. Данный подход требует дальнейшего усиления и системной реализации. Также нуждаются в доработке механизмы контроля уникальности тестов ($M = 4,30$; $SD = 0,98$) — рекомендовано усиление логики проверки на дублирование и внедрение визуализации совпадений. Аналогично, модуль обновлений ($M = 4,18$; $SD = 0,99$) должен быть дополнен автоматическим обновлением и возможностью отката (rollback) в случае некорректной инсталляции.

Критически уязвимыми участками, сочетающими низкие оценки и высокий разброс мнений ($SD > 1,2$), оказались скорость отклика ($M = 3,91$; $SD = 1,29$), лицензирование используемых технологий ($M = 3,82$; $SD = 1,21$) и локализация интерфейса ($M = 4,11$; $SD = 1,19$). Для их улучшения предлагается внедрить кэширование, провести аудит лицензионных рисков и расширить поддержку языков с последующим проведением юзабилити-тестирования.

В то же время, стабильными компонентами системы были признаны управление заданиями ($M = 4,67$; $SD = 0,41$) и блок безопасности ($M = 4,18$; $SD = 0,95$). Первая из них может быть использована в качестве референсной модели при совершенствовании других модулей, в то время как вторая требует расширения логирования и мониторинга на уязвимые участки, включая генерацию тестов и экспорт.

Таким образом, проведенный анализ позволяет выработать приоритетные направления для совершенствования АСУ ТЗ с опорой как на пользовательское восприятие, так и на объективные характеристики эксплуатации.

Заключение

Программа включает в себя ряд модулей: управление банком заданий, автоматическая генерация тестов, контроль повторяемости и равномерности, экспорт документов, ведение логов и обеспечение безопасности. Генерация тестов осуществляется в специально оборудованных помещениях с ограниченным доступом, а итоговые материалы сохраняются в зашифрованной форме. Такой подход позволяет обеспечить академическую достоверность, предотвратить утечки информации и соблюсти высокий уровень информационной безопасности.

На основании проведенного исследования можно сделать вывод, что АСУ ТЗ, внедренная в работу Центра, в целом демонстрирует высокую степень эффективности и удовлетворяет ключевым требованиям пользователей в контексте функциональности, надежности и удобства использования. Тем не менее, собранные данные выявили отдельные

компоненты, требующие внимания: это недостаточная скорость выполнения отдельных операций, ограниченные возможности пользовательской настройки и необходимость расширения функционала системы в части генерации отчетной документации.

С учетом полученных результатов исследования, особое внимание следует уделить обеспечению правовой и технологической защиты интеллектуальных решений. Поскольку один из компонентов – система автоматической генерации тестовой документации – потенциально уязвим с точки зрения воспроизводимости, рекомендуется рассмотреть вопрос о патентовании ключевых технических решений. Это обеспечит как защиту интеллектуальной собственности, так и устойчивость к внешним рискам.

Актуальность дальнейшего развития АСУ ТЗ обусловлена устойчивым трендом на цифровизацию процедур оценки в системе образования, особенно в условиях проведения экзаменов с высокой ставкой, где требования к надежности, прозрачности и воспроизводимости решений предельно высоки. В этом контексте АСУ ТЗ может рассматриваться как базовая технологическая платформа, способная обеспечить объективность и стандартизацию процедуры тестирования.

Таким образом, цифровые решения, такие как АСУ ТЗ, являются неотъемлемой частью процесса модернизации отечественного образования. Их внедрение требует системного подхода, включающего техническую оптимизацию, профессиональную подготовку пользователей, нормативно-правовое сопровождение и регулярную обратную связь как основу для непрерывного совершенствования.

Список литературы

1. Постановление Правительства Республики Казахстан от 28 марта 2023 года № 269 «Концепция цифровой трансформации, развития отрасли информационно-коммуникационных технологий и кибербезопасности на 2023–2029 годы». – <https://adilet.zan.kz/rus/docs/P2300000269>
2. Gierl M. J., Lai H. Automated Item Generation: Practical Implications for Educational Testing // *Journal of Educational Measurement*. – 2018. – Vol. 55, No. 3. – P. 297-313. – DOI: 10.1111/jedm.12186
3. Baker F. B. The Basics of Item Response Theory. Chapter 2: Historical Development. – 2001. – <https://eric.ed.gov/?id=ED458219>
4. Chen L., Chen P. Machine Learning in Test Construction // *Journal of Educational Data Mining*. – 2023. – Vol. 15, No. 1. – P. 45-68. – <https://doi.org/10.5281/zenodo.7564321>
5. Приказ Министра образования и науки Республики Казахстан от 8 мая 2019 года № 190 «Об утверждении Правил проведения комплексного тестирования». – <https://adilet.zan.kz/rus/docs/V2100026093>
6. ISO/IEC 25002:2024. Systems and software engineering – Quality requirements and evaluation (SQuaRE). – 2024. – <https://www.iso.org/standard/78175.html>
7. Gakhov A. Цифровизация обучения: вызовы, проблемы и пути решения. – 2023. – <https://tecedu.academy/en/blog/cifrovizaciia-obuceniia-vyzovy-problemy-i-puti-reseniia>
8. ISO 25010. System and Software Quality Models. – <https://www.iso25000.com/index.php/en/iso-25000-standards/iso-25010>
9. Андерсен П., Морган Д. Разработка тестов и анкет для национальной оценки учебных достижений. Книга 2. – М. : Логос, 2011. – 204 с.
10. Шкала Ликерта. – https://ru.wikipedia.org/wiki/Шкала_Ликерта
11. Pearson correlation coefficient. – https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
12. Sauro J., Lewis J. Quantifying the User Experience: Practical Statistics for User Research (2nd ed.). – Elsevier, 2016. – 384 p.
13. Закон Республики Казахстан от 10 июня 1996 года № 6-І «Об авторском праве и смежных правах». – https://adilet.zan.kz/rus/docs/Z960000006_

Ш.Б. Алтыбаева, М.Д. Каргалдаева, Р.Д. Алибеков

ЦИФРЛЫҚ ОРТАДА ТЕСТ ТАПСЫРМАЛАРЫ БАЗАСЫН БАСҚАРУДЫҢ ТИІМДІЛІГІ

Білім беру саласының цифрлық трансформация кезеңінде тест тапсырмаларын сенімді, қауіпсіз және тиімді басқаруды қамтамасыз ететін автоматтандырылған шешімдерге қажеттілік туындайды. Бұл зерттеу Ұлттық тестілеу орталығында магистратураға кешенді тестілеу нұсқаларын қалыптастыру мақсатында әзірлеп, іске асырған жүйенің тиімділігін бағалауға арналған. Зерттеу барысында аралас әдіснама қолданылды: сандық талдау (Лайкерт шкаласын пайдалана отырып Орталықтың 44 қызметкеріне сауалнама жүргізу, корреляциялық талдау) және жүйенің нақты жағдайда жұмыс істеуін сапалық хронометраждық бақылау. Зерттеу нәтижесінде тұрақты жұмыс істейтін компоненттермен қатар жетілдіруді қажет ететін бағыттары да анықталды. Жүйе архитектурасын және интерфейсін жетілдіру, өнімділікті арттыру, сондай-ақ құқықтық және ақпараттық қауіпсіздікті қамтамасыз ету бойынша практикалық ұсынымдар әзірленді. Зерттеу нәтижелері жүйенің тапсырмаларды басқару функционалына жоғары қанағаттанушылықты көрсетті ($M=4,67$), алайда өнімділік ($M=3,92$) пен үлкен көлемдегі деректерді өңдеу мәселелері де байқалды. Оптимизациялау бағыттары ретінде іздеу алгоритмдерін күшейту, деректерді индекстеу, интерфейсті жетілдіру және бағдарламалық қамтамасыз етуді лицензиялық талаптарға сәйкестендіру ұсынылды. Зияткерлік меншікті қорғау мақсатында негізгі техникалық шешімдерді патенттеу ұсынылады. Алынған деректер мемлекеттік білім беру инфрақұрылымы жағдайында бағалау саласындағы цифрлық шешімдерді одан әрі жаңғырту үшін пайдалануға болады.

Түйін сөздер: автоматтандырылған тестілеу, цифрлық трансформация, тиімділікті бағалау, Лайкерт шкаласы, ақпараттық қауіпсіздік.

Sh.B. Altybayeva, M.D. Kargaldayeva, R.D. Alibekov

EFFECTIVENESS OF TEST ITEM BANK MANAGEMENT IN A DIGITAL ENVIRONMENT

In the context of the digital transformation of education, there is an increasing need for automated solutions that ensure reliable, secure, and efficient management of test materials. This study focuses on evaluating the effectiveness of the system developed and implemented by the National Testing Center for generating versions of the comprehensive master's degree entrance test. A combined methodology was used in this research: quantitative analysis (a survey of 44 Center employees using the Likert scale, correlation analysis) and qualitative time-tracking observation of the system's operation in real-world conditions. The study identified both consistently functioning components and areas requiring improvement. Practical recommendations were developed for enhancing the system's architecture and interface, improving performance, and ensuring legal and information security. The findings revealed a high level of user satisfaction with the task management functionality ($M=4.67$), while issues were noted with system performance ($M=3.92$) and the processing of large data volumes. Optimization directions were formulated, including improved search algorithms, database indexing, better interface localization, and software compliance with licensing requirements. It is recommended to patent key technical solutions to protect intellectual property. The results can be used for further modernization of digital assessment tools within the national educational infrastructure.

Keywords: automated testing, digital transformation, effectiveness evaluation, Likert scale, information security.

References

1. Postanovlenie Pravitel'stva Respubliki Kazakhstan ot 28 marta 2023 goda № 269 "Kontseptsiya tsifrovoy transformatsii, razvitiya otrasli informatsionno-kommunikatsionnykh tekhnologiy i kiberbezopasnosti na 2023–2029 gody" [Resolution of the Government of the Republic of Kazakhstan dated March 28, 2023, No. 269 "Concept of digital transformation, development of the information and

communication technologies industry and cybersecurity for 2023–2029"]. (2023). <https://adilet.zan.kz/rus/docs/P2300000269>

2. Gierl, M. J., & Lai, H. (2018). Automated Item Generation: Practical Implications for Educational Testing. *Journal of Educational Measurement*, 55(3), 297-313. <https://doi.org/10.1111/jedm.12186>

3. Baker, F. B. (2001). *The Basics of Item Response Theory. Chapter 2: Historical Development*. <https://eric.ed.gov/?id=ED458219>

4. Chen, L., & Chen, P. (2023). Machine Learning in Test Construction. *Journal of Educational Data Mining*, 15(1), 45-68. <https://doi.org/10.5281/zenodo.7564321>

5. Prikaz Ministra obrazovaniya i nauki Respubliki Kazakhstan ot 8 maya 2019 goda № 190 "Ob utverzhdenii Pravil provedeniya kompleksnogo testirovaniya" [Order of the Minister of Education and Science of the Republic of Kazakhstan dated May 8, 2019, No. 190 "On approval of the Rules for conducting complex testing"]. (2019). <https://adilet.zan.kz/rus/docs/V2100026093>

6. ISO/IEC 25002:2024. (2024). *Systems and software engineering – Quality requirements and evaluation (SQuaRE)*. <https://www.iso.org/standard/78175.html>

7. Gakhov, A. (2023). Tsifrovizatsiya obucheniya: vyzovy, problemy i puti resheniya [Digitalization of learning: challenges, problems and solutions]. <https://tecedu.academy/en/blog/cifrovizatsiia-obuceniia-vyzovy-problemy-i-puti-reseniia>

8. ISO 25010. (n.d.). *System and Software Quality Models*. <https://www.iso25000.com/index.php/en/iso-25000-standards/iso-25010>

9. Andersen, P., & Morgan, D. (2011). *Razrabotka testov i anket dlya natsional'noy otsenki uchebnykh dostizheniy. Kniga 2* [Development of tests and questionnaires for national assessment of educational achievements. Book 2]. Logos.

10. Shkala Likerta [Likert scale]. (n.d.). Wikipedia. https://ru.wikipedia.org/wiki/Шкала_Ликерта

11. Pearson correlation coefficient. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

12. Sauro, J., & Lewis, J. (2016). *Quantifying the User Experience: Practical Statistics for User Research (2nd ed.)*. Elsevier.

13. Zakon Respubliki Kazakhstan ot 10 iyunya 1996 goda № 6-I "Ob avtorskom prave i smezhnykh pravakh" [Law of the Republic of Kazakhstan dated June 10, 1996, No. 6-I "On copyright and related rights"]. (1996). https://adilet.zan.kz/rus/docs/Z960000006_

Авторлар туралы мәлімет:

Алтыбаева Шугыла Болатовна (автор-корреспондент) - экономика ғылымдарының магистрі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: shugyla.altbayeva@gmail.com.

Каргалдаева Молдир Дуйсенбаевна - PhD, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: moldirk2018@gmail.com.

Алибеков Рахымжан Джаныбекович - техникалық ғылымдар магистрі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: uto@testcenter.kz

Сведения об авторах:

Алтыбаева Шугыла Болатовна (автор-корреспондент) - магистр экономических наук, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: shugyla.altbayeva@gmail.com.

Каргалдаева Молдир Дуйсенбаевна - PhD, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: moldirk2018@gmail.com.

Алибеков Рахымжан Джаныбекович - магистр технических наук, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: uto@testcenter.kz

Information about authors:

Altybaeva Shugyla Bolatovna (corresponding author) - master of economics, RST on the REM «National Testing Center» Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: shugyla.altbayeva@gmail.com.

Kargaldayeva Moldir Duissenbayevna - PhD, RST on the REM «National Testing Center» Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: moldirk2018@gmail.com.

Alibekov Rakhymzhan - master of Technical Sciences, RST on the REM «National Testing Center» Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: uto@testcenter.kz