

Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Комсомольский-на-Амуре государственный университет»

На правах рукописи

Закусило Александр Михайлович

**Проектирование и разработка системы адаптивного
тестирования учащихся**

Направление подготовки 09.04.03
«Прикладная информатика»

АВТОРЕФЕРАТ
МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

2022

Работа выполнена в ФГБОУ ВО
«Комсомольский-на-Амуре государственный университет»

Научный руководитель: кандидат технических наук,
доцент кафедры «Проектирование, управление и разработка информационных систем»
Абарникова Елена Борисовна

Рецензент: кандидат физико-математических наук,
доцент кафедры информационной безопасности,
ФГБОУ ВО «АМГПУ»
Анисимов Антон Николаевич

Защита состоится 23 июня 2022 г. в 9.50 часов на заседании государственной экзаменационной комиссии по направлению подготовки 09.04.03 – «Прикладная информатика» в ФГБОУ ВО «КнАГУ» по адресу: 681000, г. Комсомольск-на-Амуре, пр. Ленина, 27, ауд. 204/5.

Автореферат разослан 21 июня 2022 г.

Секретарь ГЭК

Е.В. Абрамсон

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИОННОЙ РАБОТЫ

Актуальность темы исследования. На сегодняшний день становится очевидно, что для успешного обучения студентов необходимо применение систем контроля знаний, которые позволяют определять недостаток подготовленности по определенным темам для корректировки учебной нагрузки и повышения эффективности учебного процесса.

Основным методом контроля знаний является прямое тестирование. Под прямым тестированием будем понимать метод, при котором структура теста не зависит от фактических ответов испытуемого, а оценка знаний тестируемого происходит после полного завершения теста. Режим прямого тестирования заложен в основу функционирования подавляющего большинства существующих систем оценивания уровня знаний. При всех своих достоинствах прямое тестирование не лишено недостатков, таких как фиксированное количество вопросов теста и низкая вовлеченность тестируемого (необходимость отвечать на вопросы, не соответствующие уровню подготовленности).

Для решения проблем прямого тестирования был разработан метод адаптивного тестирования. Адаптивное тестирование - технология тестирования слушателей, где каждый следующий вопрос подбирается автоматически, исходя из ответов, данных на предыдущие вопросы. Такой процесс тестирования позволяет адаптироваться под уровень знаний каждого испытуемого и пропускать слишком сложные или лёгкие вопросы.

Проблема состоит в том, что на данный момент системы с адаптивным тестированием до сих пор находятся на стадии становления, а эффективность моделей, которые используются в данных системах, не изучена.

Цель исследования:

- сокращение временных потерь учебно-преподавательского состава на разработку тестов;
- сокращение временных потерь тестируемого на прохождение теста;
- увеличение вовлеченности и снижение влияния на тестируемого дополнительных факторов (потеря интереса, отвлечение, утомление, беспокойство);
- увеличение точности результатов тестирования.

Для достижения поставленных целей необходимо решить следующие задачи исследования:

- провести анализ существующих систем адаптивного тестирования и выявить перспективные направления в области их развития;

- исследовать применение математических моделей теории IRT в системах адаптивного тестирования;
- спроектировать и разработать систему адаптивного тестирования на основе наиболее эффективной математической модели.

Объектом исследования являются методы и модели контроля знаний.

Предметом исследования являются технологии адаптивного тестирования.

Методологические основы и методы исследования: для решения поставленных задач использовались методы математического моделирования, системного анализа, статистического анализа и численные методы.

Хронологические рамки исследования определяются периодом с 2020 года по 2022 году.

Информационную базу исследования составили материалы опубликованных исследований отечественных авторов, ориентированных на описание математических моделей для систем адаптивного тестирования.

Научная новизна предлагаемых исследований заключается:

- 1 в нахождении наиболее эффективной математической модели IRT для анализа и подбора вопросов в системах адаптивного тестирования;
- 2 в авторском алгоритме прохождения адаптивного тестирования.

Достоверность результатов исследования подтверждается:

- 1 практической реализацией и успешным применением системы адаптивного тестирования учащихся, созданной на основе разработанного подхода;
- 2 вычислительными экспериментами, подтвердившими эффективность и преимущества созданных численных методов идентификации вероятностных моделей.

Практическая значимость и ценность полученных результатов исследования заключается в том, что разработанное научно-методическое сопровождение и программно-инструментальная среда адаптивного тестирования могут быть применены в учебном процессе и использованы для контроля знаний во всех уровнях образования.

Основные положения, выносимые на защиту:

- 1 авторский алгоритм прохождения адаптивного тестирования;
- 2 программный комплекс системы адаптивного тестирования учащихся.

Апробация результатов. Результаты работы докладывались и обсуждались на V Всероссийской национальной научной конференции молодых учёных «Молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований».

Объём и структура диссертационной работы: диссертация включает в себя введение, три основные главы, заключение, список используемой литературы и одно приложение. Объём диссертации составляет 91 страницу. Текст работы содержит 1 таблицу и 43 рисунка. Для написания работы использовалось 30 источников литературы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении дано обоснование актуальности и характеристика работы, сформулированы цель и задачи, указаны методы исследования, представлены основные положения, показана научная новизна и практическая значимость, описана структура работы.

В первой главе рассмотрена концепция, алгоритмы и основные модели систем адаптивного тестирования. Проведён анализ основных этапов проведения адаптивного теста и процесс формирования тестовой базы. Описан процесса формирования и структурирования базы вопросов для проведения адаптивного тестирования. Приведён базовый алгоритм прохождения адаптивного теста.

Алгоритм варьирующего адаптивного тестирования является замкнутым и выполняется до достижения останавливающего условия. Схема описывающая данный процесс представлена на рисунке 1.



Рисунок 1 - Алгоритм варьирующего многошагового тестирования

На первом этапе определяется априорный или начальный уровень подготовки тестируемого. Это необходимо сделать для выбора оптимальной сложности при начале тестирования. На втором этапе происходит подбор и предъявление задания тестируемому. На третьем этапе происходит обработка ответа от тестируемого и установка апостериорной оценки уровня подготовленности. На четвёртом этапе происходит сравнение апостериорной оценки и других критериев с останавливающим критерием (время

тестирования, минимальная и максимальная сложность теста). Если условие остановки не выполнено, то процесс тестирования возвращается в третий этап и происходит подбор очередного вопроса. Если условие остановки выполнено, то процесс тестирования переходит в пятый этап – шкалирования оценки уровня знаний и предоставление результатов теста.

Каждый этап алгоритма может быть реализован уникальным образом в зависимости от вида стратегии и имеющихся технологических возможностей. Существующие системы, реализующие данный алгоритм, являются закрытыми и это не позволяет их проанализировать.

На рисунке 2 изображены траектории адаптивного тестирования трех учащихся, начинавших свой вход в адаптивные тестирования без выполнения протест теста.

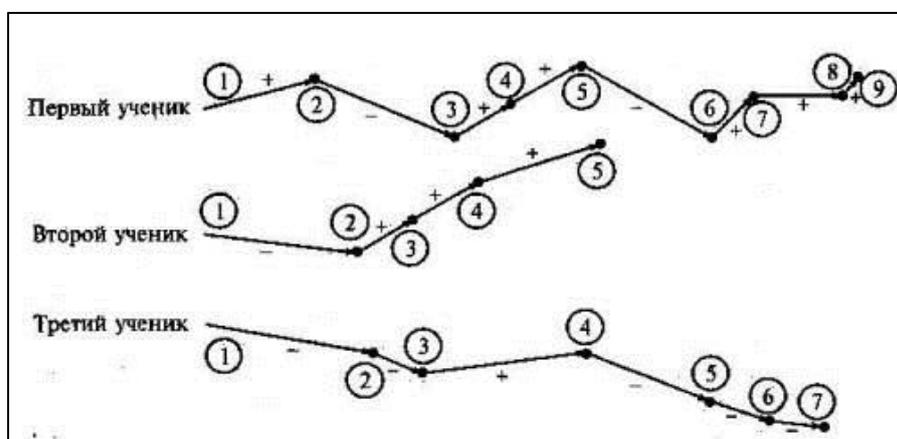


Рисунок 2 - Визуализация индивидуальных траекторий испытуемых

Сложность предоставляемых заданий на рисунке 2 определена высотой ломаной. Продолжительность тестирования и количество вопросов определено длиной ломаной. Для визуального удобства траектории тестирования на графике отображаются как непересекающиеся ломаные линии. При правильном ответе на тестовый вопрос ломаная линия на графике возрастает, при неправильном ответе - убывает.

Надежность в таком тестировании будет зависеть от совокупности факторов. К ним относятся: высокая степень вариативности тестовых заданий, как по уровню сложности, так и по типу; наличие рецензентов для проверки качества заданий; наличие фокус группы для правильного распределения заданий по уровню сложности. На надежность также влияет качество начального контроля (наличие или отсутствие проктор теста) и выбранная математическая модель адаптивного тестирования.

Во второй главе описывается теория IRT, приводятся математические модели, основанные на данной теории. Приведённые математические модели сравниваются по критерию величины получаемых знаний. Наиболее

эффективная математическая модель интегрируется в базовый алгоритм адаптивного тестирования.

На сегодняшний день при разработке систем адаптивного тестирования наибольшее распространение получила теория тестовых заданий (в дальнейшем - IRT), которая описывает связь между уровнем знаний испытуемых и результатами выполнения тестов, что позволяет определить уровень знаний независимо от сложности заданий.

Основное назначение аппарата IRT заключается в том, чтобы обеспечить эффективное прогнозирование результатов тестирования для задания различного уровня сложности. Согласно IRT вероятность правильного ответа на задание испытуемого с высоким уровнем знаний должна быть больше вероятности правильного ответа у испытуемого с низким уровнем знаний. Чем выше уровень знаний, тем выше вероятность правильного ответа на задание данного уровня сложности. Особенность применения IRT заключается в том, что ответы множества испытуемых на множество заданий теста прогнозируются на основе математических моделей, при наличии эмпирически полученной матрицы исходных тестовых баллов. Эмпирический способ получения матрицы исходных тестовых баллов предполагает для распределения вопросов по уровням сложности на основе выполнения их тестовой группой. Традиционно за правильный ответ ставится 1 балл, за неправильный – 0 баллов. Теория IRT является основополагающей в рамках концепции адаптивного тестирования. На основе IRT строятся большинство моделей адаптивного тестирования.

В качестве математической модели взаимосвязи между значениями латентных переменных θ , β и наблюдаемыми результатами выполнения теста в IRT выбрана условная вероятность правильного выполнения обучаемыми заданиями теста. В частности, можно рассматривать условную вероятность P_j правильного выполнения m испытуемым с уровнем подготовки θ_i , различных по трудности заданий теста, считая θ_i , параметром i -го ученика β - независимой переменной.

$$P_i\{x_{yl} = 1|\theta\} = f(\theta_i - \beta), l = 1, 2, \dots, N. \quad (2.1)$$

Аналогично вводится P_j для обозначения вероятности правильного выполнения i -го задания трудностью β_j ; различными испытуемыми группы. Здесь θ - независимая переменная, а β_j ; - параметр, определяющий трудность j -го задания теста:

$$P_j\{x_{ij} = 1|\beta_j\} = \varphi(\theta_i - \beta_j), l = 1, 2, \dots, N, \quad (2.2)$$

где $X_{ij} = 1$, если ответ i -го испытуемого на j -е задание верный, или $X_{ij} = 0$, если ответ i -го испытуемого на j -е задание не верный. F и φ - символы

функциональной зависимости; N - число испытуемых; n - количество заданий в тесте.

В основе реализации аппарата IRT всегда будет использована одна из следующих моделей: однопараметрическая модель Раша, двухпараметрическая модель Бирнбаума, трёхпараметрическая модель или модель РСМ.

Чем больше информации мы получаем из ответов тестируемого, тем точнее наши сведения о уровне его знаний, то есть меньше ошибка. Из этого можно сделать вывод, что увеличение количества информации означает повышение эффективности тестирования. В IRT количеством информации называют величину, обратную дисперсии ошибок, а информационной функцией – соответствующую аналитическую зависимость.

$$I = \frac{1}{D} = \frac{1}{\sigma^2}. \quad (2.3)$$

Для трёх базовых моделей IRT количество информации рассчитывается по формуле:

$$I = \frac{1}{\sigma_{\theta_i}^2} = \sum_{j=1}^m \left[\left(\frac{1-P_{ij}}{P_{ij}} \right) \left(\frac{P_{ij}-c_j}{1-c_j} \right)^2 \right], \quad (2.4)$$

где σ_{θ_i} - стандартная ошибка уровня подготовленности i -го испытуемого; m - количество тестовых заданий; P_{ij} - вероятность правильного ответа i -го тестируемого на j -е задание; a_j и c_j - дифференцирующая способность и параметр коррекции на угадывание правильного ответа j -го задания.

Для одного задания формула 2.4 примет вид:

$$I = a^2 \left[\left(\frac{1-P}{P} \right) \left(\frac{P-c}{1-c} \right)^2 \right]. \quad (2.5)$$

В модели Раша $a_j = 1$, а $c_j = 0$, что приводит к следующей аналитической зависимости:

$$I = 1^2 \left[\left(\frac{1-P}{P} \right) \left(\frac{P-0}{1-0} \right)^2 \right] = (1-P)P = \left(1 - \frac{e^{\theta-\beta}}{1+e^{\theta-\beta}} \right) \frac{e^{\theta-\beta}}{1+e^{\theta-\beta}}. \quad (2.6)$$

Из формулы 2.5 следует, что количество информации максимально при вероятности правильного ответа $P = 0,5$:

$$I'(P) = ((1-P)P)' = 1 - 2P = 0, P = 0,5. \quad (2.7)$$

Информативность заданий, существенно отличающихся от уровня подготовленности испытуемого, почти нулевая. Действительно, способность испытуемого решать очень простые задания (θ много больше β) или неудачи при решении заданий повышенной сложности (θ много меньше β) мало информативны, так как не дают возможности уточнить уровень подготовленности тестируемого. Данная зависимость представлена на рисунке 3.

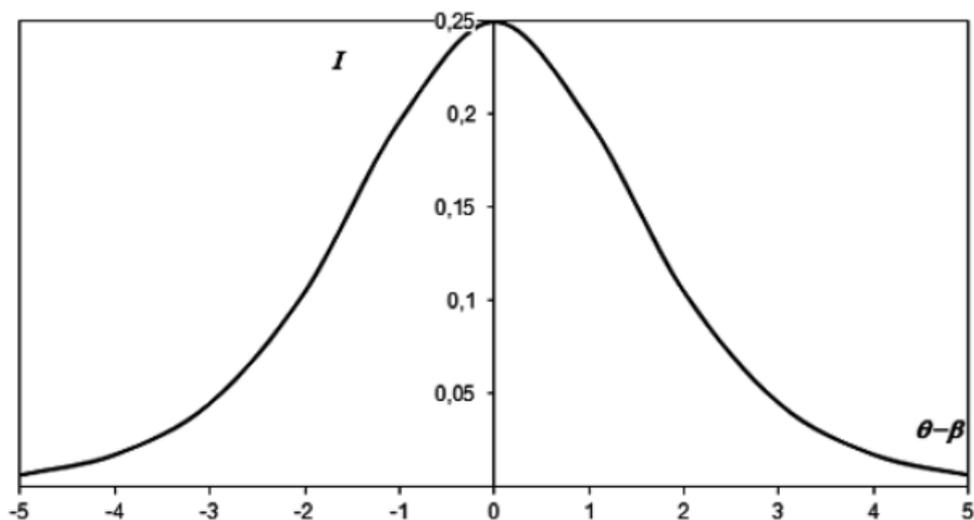


Рисунок 3 – Зависимость количества информации от разности уровня подготовленности испытуемого и уровня трудности задания

Для двухпараметрической модели ($c_j = 0$) формула 2.5 примет вид:

$$I = a^2(1 - P)P = a^2 \left(1 - \frac{e^{a(\theta - \beta)}}{1 + e^{a(\theta - \beta)}}\right) \frac{e^{a(\theta - \beta)}}{1 + e^{a(\theta - \beta)}}. \quad (2.8)$$

Количество информации для одного тестового задания в Partial Credit Model:

$$I = \sum_{l=1}^{x_{\max j}} l^2 * \pi_{ijl} - \left(\sum_{l=1}^{x_{\max j}} l * \pi_{ijl}\right)^2. \quad (2.9)$$

Сравнительный анализ моделей по количеству получаемой информации за ответ приведен в таблице 1.

Таблица 1 – Сравнение информационной ценности ответов

	Количество информации				
	Модель Раша	Двухпараметрическая модель ($a = 0,3$)	Partial Credit Model		
			двухбалльное задание	трёхбалльное задание	четырёхбалльное задание
Максимальное значение	0,25	0,0225	0,67	1,25	2
Незначительное расхождение уровня подготовленности испытуемого и уровней трудности заданий (0,1 лолита)	0,249	0,0225	0,644	1,2	1,901

Продолжение таблицы 1

	Количество информации				
	Модель Раша	Двухпара- метричес- кая модель ($a = 0,3$)	Partial Credit Model		
			двух- балльное задание	трёх- балльно е задание	четырёх- балльное задание
Значительное расхождение уровня подготовленности испытуемого и уровней трудности заданий (1 лолит)	0,197	0,0220	0,424	0,788	0,837

После анализа моделей адаптивного тестирования были сделаны следующие выводы:

1 тип математической модели оказывает существенное влияние на эффективность адаптивного тестирования;

2 наиболее эффективной в условиях адаптивного тестирования является двухпараметрическая модель, при условии достаточного количества заданий с высокой дифференцирующей способностью, равномерно распределённых по всему диапазону измерения уровня подготовленности испытуемых;

3 модель РСМ и модель Раша являются предпочтительными, если указанное выше условие не выполняется;

Для реализации системы адаптированного тестирования будет использована модель Раша из-за своей эффективности. Так же основным преимуществом модели Раша является низкая зависимость от дифференцирующую способности заданий, поэтому в условиях низкой наполненности тестовой базы данная модель будет показывать максимальную эффективность.

Рассмотрим теперь алгоритм тестирования (рисунок 4), который предлагается испытуемым. Введем обозначения: X – количество выданных заданий; Y – количество правильных ответов, данных испытуемым; Θ – текущий уровень подготовки испытуемого; i – номер текущего шага; R – оценка испытуемого (неудовлетворительно, удовлетворительно, хорошо или отлично); G – группа сложности, g – вопрос теста. Также для описания алгоритма необходимо ввести понятие критического количества вопросов $X_{\text{крит}}$ – количество вопросов для группы сложности, на которое испытуемый должен ответить (без изменения группы сложности), чтобы выставить ему оценку R .

Значение $X_{\text{крит}}$ зависит от общего числа вопросов в базе и в каждой группе сложности и от времени тестирования; на начальном этапе определяется преподавателем. Во избежание постоянных смен уровней сложности при чередовании верных и ошибочных ответах тестируемого введём переменную η – коэффициент верных ответов для сохранения уровня сложности (изменяется от 0 до 1), также определяется преподавателем. Рекомендуется на начальном этапе использовать значение $\eta = 1/3$.

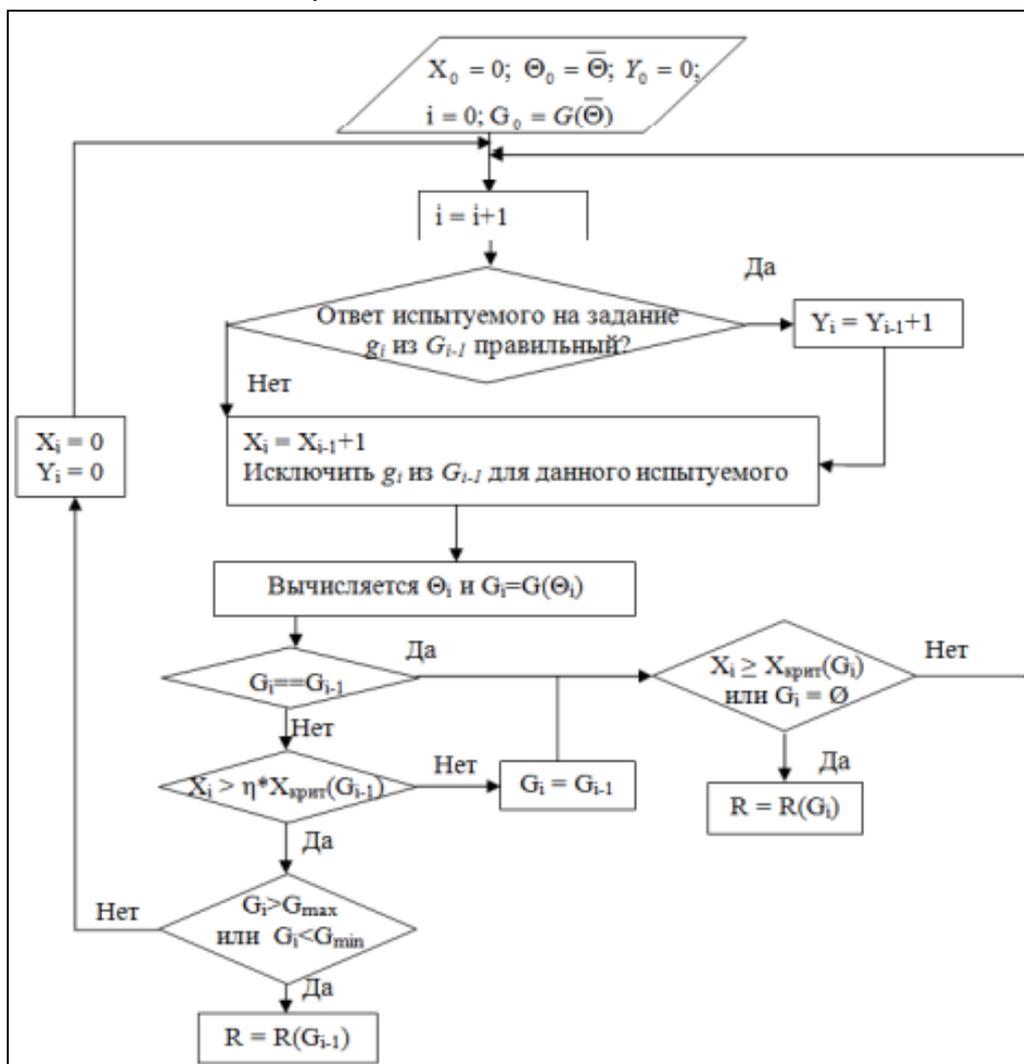


Рисунок 4 – Алгоритм адаптивного тестирования

Процедура оценки уровня подготовки испытуемых (рисунок 4):

1 выбирается задание из текущей группы сложности вопросов $g_i \in G_i$ предлагается для ответа испытуемому;

2 при успешном ответе испытуемого, увеличивается количество правильных ответов;

3 увеличивается количество выданных заданий, выданное задание исключается из набора заданий для данного испытуемого;

4 вычисляется новый уровень подготовки испытуемого Θ_i по модели Раша ранее описанным способом;

5 вычисляется группа сложности в зависимости от значения Θ_i . Принадлежность Θ_i к группе сложности определяется путем сравнения значения Θ_i с максимальным и минимальным значениями трудности заданий для каждой группы сложности. Если $\Theta_i > \beta_{\min}(G_k)$ и $\Theta_i < \beta_{\max}(G_k)$, то $G_i = G_k$ (k – индекс перебора по всем группам сложности, k изменяется от 1 до 4);

6 если группа сложности изменилась, количество выданных заданий превышает η -ую часть от $X_{\text{крит}}$ и новая группа сложности больше максимальной (оценка 5) или меньше минимальной (оценка 2), то выставляется оценка в соответствии с группой сложности на предыдущем шаге (5 или 2);

7 если группа сложности изменилась, количество выданных заданий превышает η -ую часть от $X_{\text{крит}}$ и новая группа сложности не больше максимальной (оценка 5) и не меньше минимальной (оценка 2), то количество выданных вопросов и правильных ответов обнуляется и происходит возврат к шагу 1;

8 если группа сложности изменилась, а количество выданных заданий не превышает η -ую часть от $X_{\text{крит}}$, то группа сложности возвращается к предыдущему значению (изменения не принимаются);

9 если группа сложности не изменилась, не превышено количество критических вопросов и текущая группа сложности содержит еще хотя бы 1 вопрос, то происходит возврат к шагу 1;

10 если группа сложности не изменилась, но при этом превышено количество критических вопросов или все вопросы текущей группы сложности уже были выданы испытуемому ранее, то выставляется оценка в соответствии с текущей группой сложности.

В третьей главе описывается процесс проектирования и разработки программной компоненты на основе наиболее эффективной математической модели. При проектировании приводятся процессная, даталогическая, инфологические модели приложения. Описывается концептуальная модель и архитектура программной компоненты. Демонстрируются результаты работы программы.

Целевые автоматизируемые процессы ИС представлены в процессной модели на рисунке 5.

Основными функциональными подсистемами в ИС адаптивного тестирования являются: авторизация; регистрация; база тестов; база заданий; база учебных групп; проигрыватель теста.

Для разработки пользовательской части приложения использовался javascript фреймворк Vue.js. Для разработки серверной части приложения использовался PHP фреймворк Laravel. Для хранения данных использовалась реляционная база данных PostgreSQL. Общая архитектура приложения представлена на рисунке 7.

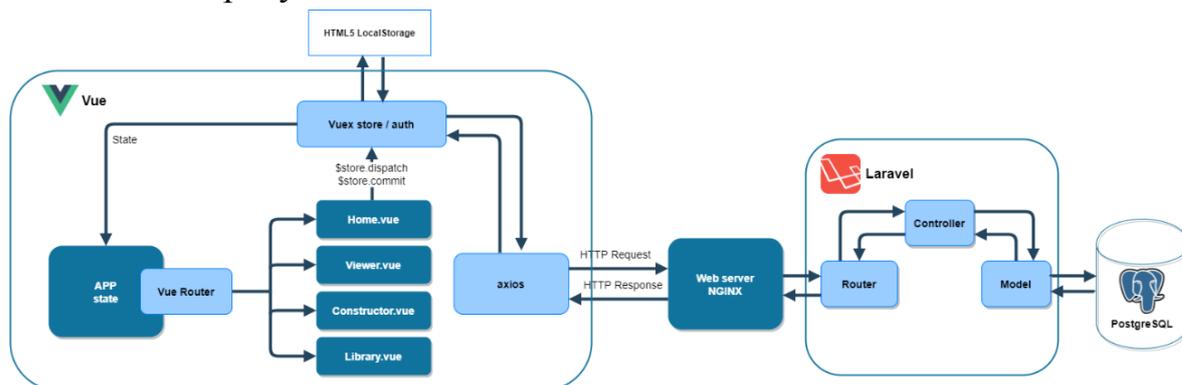


Рисунок 7 – Архитектура ИС адаптивного тестирования

На основе спроектированных моделей была разработана программная компонента адаптивного тестирования учащихся. Результат представлен на рисунке 8.

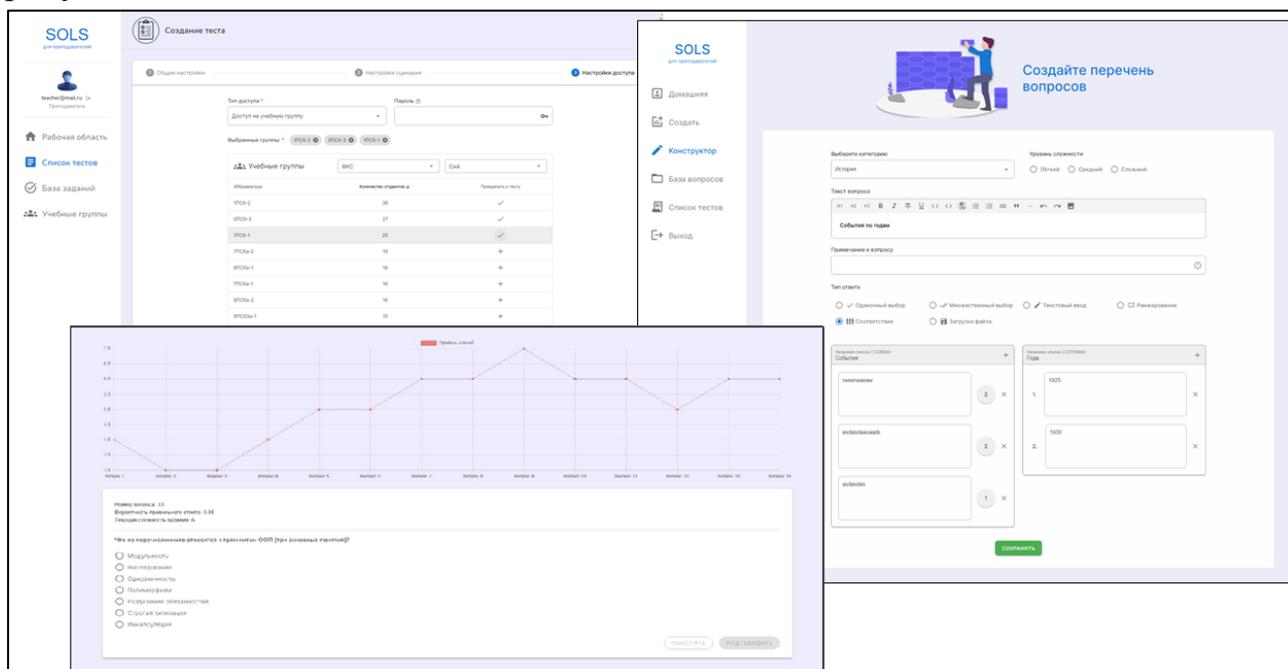


Рисунок 8 – Программная компонента

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В результате работы проведен анализ уже существующих моделей адаптивного тестирования, на основе которых выявлены перспективные направления в области их развития. Концептуально обоснован выбор объекта и предмета исследования. После проведения исследования эффективности работы существующих моделей адаптивного тестирования было установлено, что существенным недостатком большинства моделей является большая зависимость количества информации от разности уровня подготовленности испытуемого и уровня трудности задания. При этом мы пришли к выводу, что наиболее устойчивой к изменению этого показателя является модель Раша.

В области комплексов программ на основе авторского алгоритма и выбранной модели (Раша) был разработан программный комплекс для адаптивного тестирования с индивидуальной траекторией прохождения для каждого испытуемого.

Практическая значимость и ценность полученных результатов исследования заключается в том, что разработанное научно-методическое сопровождение и программно-инструментальная среда по реализации адаптивного тестирования могут быть применены в качестве системы контроля знаний в различных учебных заведениях, использоваться в системе повышения квалификации профессорско-преподавательского состава.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ИССЛЕДОВАНИЯ

Закусило А. М. Система адаптивного тестирования учащихся / А. Н. Намоконов, Е.Б. Абарникова // молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований материалы V Всероссийской. нац. науч. конф. студентов, аспирантов и молодых ученых, Комсомольск-на-Амуре, (Комсомольск-на-Амуре, апрель 2022 г.).

Закусило А. М. Разработка интеллектуальной системы анализа и формирования информационных кадров / А. М. Закусило, Е.Б. Абарникова / молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований материалы IV Всероссийской. нац. науч. конф. студентов, аспирантов и молодых ученых, Комсомольск-на-Амуре, (Комсомольск-на-Амуре, апрель 2021 г.).