

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

**«Сибирский государственный аэрокосмический университет
имени академика М. Ф. Решетнева»
(СибГАУ)**

ИНСТИТУТ	Информатики и телекоммуникаций
НАПРАВЛЕНИЕ	230100
	Информатика и вычислительная техника
МАГИСТЕРСКАЯ ПРОГРАММА	230100.68-01
	Теоретическая информатика
КАФЕДРА	Прикладной математики

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Комплексное использование ИКТ и

методов математического моделирования

Магистрант	А. В. Саяпин
	ФИО
Научный руководитель	К. В. Сафонов
	ФИО
Рецензент	
	ФИО
Ответственный за нормоконтроль	
	ФИО
Допускается к защите	
Зав. кафедрой	К. В. Сафонов
	ФИО
«___» _____ 2012 г.	

Красноярск 2012 г.

АННОТАЦИЯ

Саяпин Александр ВладимировичКомплексное использование ИКТ и методов математического
моделированияНаучный руководитель д-р физ-мат- наук, доцент Сафонов Константин
Владимирович

Направление подготовки: 230100 – Информатика и вычислительная техника

Магистерская программа: 230100.68-01 Теоретическая информатика

Институт информатики и телекоммуникаций

Кафедра прикладной математики

Год защиты: 2012

В магистерской диссертации рассматриваются вопросы проведения тестирования, оценки результатов выполнения респондентом контрольно-измерительного материала, оценки объективной сложности и дифференцирующей способности теста. Разработана объектная библиотека, предназначенная для разработки, оценивания и проведения компьютерного тестирования. На основе библиотеки построены приложения, предназначенные для компьютерной разработки, хранения, оценки и проведения теста.

Ключевые слова: тест, контрольно-измерительный материал, мера, характеристика, статистика, имитационное моделирование.

ABSTRACT

Sayapin Alexander V.The using of informational technologies and the mathematical simulating
Scientific adviser – Doctor of physics and mathematics, associate prof. Safonov
Konstantin V.

Course of qualification: 230100 – Computer science and computer techniques

Master's Program: 230100.68-01 – Theoretical Computer Science

Institution of Computer Science and telecommunications

Applied Mathematics Chair

Year of defense: 2012

In this master's work the computer-based methods of knowledge estimation and evaluation is described. The main part of the work is devoted to the method of the test's complexity and differentiation ability evaluation. The method is based on the approaches of mathematical statistics and simulation. On the basis of the described method the set of applications is developed. This software allows to develop, to estimate characteristics and to perform test both in printed and on-line form.

Key words: knowledge estimation, mathematical statistics, simulation

Содержание

ВВЕДЕНИЕ.....	4
1. Анализ задачи и способов ее решения.....	7
1.1. Инженерная педагогика.....	7
1.2. Контроль результатов обучения и их измерение.....	9
1.2. Числовые меры теста	17
1.3. Существующие средства автоматизированного тестирования.....	19
1.4. Обзор методов решения.....	21
1.4.1. Требования к числовым характеристикам теста и оценки	21
1.4.1. Вид распределения ответов студентов	22
1.5. Выводы по главе.....	27
2. МЕТОДИКА ОЦЕНКИ УРОВНЯ ЗНАНИЙ.....	28
2.1. Определение оценки за тест.....	28
2.2. Оценка объективной меры сложности задания	31
2.3. Оценивание дифференцирующей способности задания.....	33
2.4. Выводы по главе.....	34
3. РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННОЙ МЕТОДИКИ В ВИДЕ НАБОРА ПРИЛОЖЕНИЙ ДЛЯ РАЗРАБОТКИ, ОЦЕНКИ ХАРАКТЕРИСТИК И ПРОВЕДЕНИЯ ТЕСТА	35
3.1. Основные принципы построения системы тестирования.....	35
3.2. Основные технические решения	37
3.3. Анализ предметной области и выбор технических решений	40
3.5. Реализация системы компьютерного тестирования	43
3.6. Выводы по главе.....	53
ЗАКЛЮЧЕНИЕ	55
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	56

ВВЕДЕНИЕ

Вопросы объективной оценки параметров, которые зависят от многих факторов, является достаточно сложной задачей, особенно в случае, если параметры являются трудно формализуемыми.

Особенно это справедливо для такого сложного параметра, как уровень знаний студента. В частности, это обусловлено как тем, что сам параметр является трудно формализуемым, так и тем, что он зависит от большого количества факторов, часть из которых, в свою очередь, является трудно формализуемыми и зачастую субъективными.

Изменения, происходящие в системе высшего образования в течение последнего десятилетия, в том числе переход к двухуровневой системе образования, присоединение России к Болонским соглашениям, переход к кредитно-модульной системе все более остро ставят вопрос об оценке качества образования, как в целом, так и для отдельного индивида.

В идеале такая оценка должна отражать уровень компетенций, знаний, умений и навыков индивида в каждый конкретный момент времени.

В большинстве случаев используемые в настоящее время оценки компетенций и знаний студентов являются относительными, а не абсолютными, то есть знания и компетенции отдельного индивида выражаются не абсолютным значением, а в сравнении с неким заранее заданным уровнем. Чаще всего в качестве такого уровня выступает средний уровень знаний, показанный некой контрольной группой, уровень знаний которой, в свою очередь, был оценен ранее другими способами.

Очевидно, такая оценка не может быть названа объективной в полном смысле этого слова.

Целью данной работы является модернизация способов оценки знаний студентов, с целью обеспечения повышения качества текущего и рубежного оценивания знаний студентов с использованием современных информационно-компьютерных технологий и методов математического моделирования.

Поставленная цель определила следующие основные этапы ее достижения:

- анализ опубликованных работ по данной теме;
- анализ методов оценки знаний студентов;
- анализ и разработка различных методов оценки как знаний студента, так и численных характеристик самого теста;
- выработка рекомендаций по объективной оценке параметров теста и знаний студента;
- разработка программного обеспечения, которое можно использовать для оценки знаний студентов.

В основу данной работы положены концепции имитационного моделирования, математической статистики, проверки статистических гипотез.

Фундаментальные работы в области педагогики были опубликованы Коменским Я. А., Дьюи Д., Бим-бадом Б. Э., Краевским В. В., Хуторским А.В. Кумариным В. В..

Задача оценки уровня знаний индивида решалась в работах Жукова Д.О., Ефремовой Н.Ф., Поддубной Л.М., Аванесова В. М. и других. Данными авторами рассматривались следующие вопросы:

- построение систем обучения и диагностики знаний;
- конструирование и параметризация тестовых контрольно-измерительных материалов;
- построение математических моделей процесса тестирования;
- выработка рекомендаций по методикам тестирования.

В настоящей работе подробно рассматривается проблема оценивания уровня знаний индивида на основе тестирования, а так же определения качества и объективных характеристик теста и контрольно-измерительного материала.

Научная новизна работы состоит в следующем:

1. Сформулирована проблема объективной оценки уровня знаний индивида с использованием тестирования в терминах математической статистики;
2. С использованием методов математической статистики и имитационного моделирования показана возможность объективной оценки таких

параметров теста (контрольно-измерительного материала), как мера трудности и дифференцирующая способность.

3. Предложен новый алгоритм решения задачи оценки уровня знаний индивида с использованием методов имитационного моделирования.

Новизна полученных результатов подтверждается публикациями автора в изданиях по математическому моделированию, информационным технологиям, заключениями специалистов по профилю работы, выступлениями на семинарах и конференциях.

Практическая ценность работы состоит в следующем:

— На базе предложенных математических моделей и алгоритмов разработана методика оценки уровня знаний индивида;

— Разработанное программное обеспечение оценки уровня знаний индивида позволили реализовать автоматизированную систему оценивания уровня знаний студентов.

Разработанная автором методика и система оценки уровня знаний студентов в настоящее время используется ФГБОУ ВПО Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева.

Результаты работы прошли всестороннюю апробацию в докладах и обсуждениях на конференциях и семинарах по системе высшего образования и информационным технологиям [7, 8, 9].

На защиту выносятся:

1. Новый метод объективной оценки меры трудности теста (контрольно-измерительного материала) и дифференцирующей способности теста (контрольно-измерительного материала).

2. Новая методика объективной оценки уровня знаний индивида при помощи теста (контрольно-измерительного материала) с применением методов математической статистики и имитационного моделирования.

3. Программное обеспечение для объективной оценки уровня знаний индивида.

1. Анализ задачи и способов ее решения

1.1. Инженерная педагогика

В процессе обучения обычно различают следующие категории целей, которых должны достичь обучаемые [5]:

- когнитивная область охватывает все те учебные цели, которые содержат интеллектуальные результаты (преподавание, проектирование приборов и машин и т.д.). Здесь идет речь об усвоенных знаниях и приобретенных умениях;
- к психомоторной области относятся цели, которые касаются развития моторных способностей (машинопись, обслуживание приборов и машин и т.д.). Здесь речь идет преимущественно о навыках, регулирующих движение тела;
- аффективная область содержит цели, которые проявляют себя в чувствах и эмоциях. Здесь речь идет о чувственных ценностях (радость, взволнованность и т.д.).

Обычно, говоря о высшем профессиональном образовании, рассматриваются когнитивная и психомоторная области, однако с введением государственных образовательных стандартов высшего профессионального образования, включающих общекультурные компетенции, аффективная область так же включается в учебный процесс.

Модель выпускника, имеющая в качестве основного образовательного конструкта компетентность будущего специалиста, определяет в качестве стратегической задачи вуза проблему формирования и развития компетенций – готовности и стремления к деятельности. Тактической задачей педагогической системы вуза в таком аспекте становится проблема развития познавательной активности студента в каждом из представленных блоков модели компетентности выпускника. Такая модель служит основой для разработки и реализации модели подготовки специалиста [3]. Однако, основной целью подготовки компетентного выпускника по-прежнему является получение специалиста, компетентного

прежде всего в профессиональном плане, что обуславливает высокую важность когнитивной и психомоторной областей целей обучения.

Каждая из этих областей включает в себя иерархию частных целей. Так, когнитивная область включает следующие уровни или ступени:

Ступень 1. Категория «знать» определяется как способность помнить выученный ранее материал. Речь идет о самом низком уровне этой таксономии, практически о воспроизведении учебного материала: фактов, феноменов, законов и пр.

Некоторые типичные глаголы: знает, определяет, описывает, обозначает, называет, изображает схемой и др.

Ступень 2. Категория «понимать смысл» определяется как способность усвоить суть учебного материала.

Некоторые типичные глаголы: различает, сравнивает, идентифицирует, выбирает, доказывает и др.

Ступень 3. Категория «сознательно использовать» понимается как способность применять выученный материал в новых ситуациях на уровне «рутинных заданий».

Некоторые типичные глаголы: изменяет, предсказывает, решает, употребляет, оценивает, находит, объясняет, рассчитывает и т.д.

Ступень 4. Категория «анализировать» определяет способность осуществлять расчленение материала на компоненты для лучшего его уяснения и уточнения его структуры. Здесь начинается уровень решения «нерутинных задач».

Некоторые типичные глаголы: анализирует, разлагает, дифференцирует, охватывает, отделяет, противопоставляет и пр.

Ступень 5. Категория «синтезировать» означает способность соединять отдельные элементы в единое целое (систему). Здесь речь идет о способности формировать новые структуры для решения «нерутинных задач».

Некоторые типичные глаголы: составляет, разрабатывает, развивает, по-новому формулирует, планирует и пр.

Ступень 6. Категория «оценивать» понимается как способность определять значимость материала с точки зрения известной цели. Цели на этой ступени соответствуют самому высокому иерархическому уровню в когнитивной области. Некоторые типичные глаголы: определяет, интерпретирует, критикует и т.д. [5].

1.2. Контроль результатов обучения и их измерение

В настоящее время различают следующие функции контроля над результатом и измерения результата обучения: диагностическую и прогностическую.

Целью контроля над результатом в диагностическом аспекте является обеспечение основ для самоконтроля учащихся и для стимулирования их учебы. Для этого необходимо проводить соответствующее измерение результата, являющееся составным элементом контроля. Измерения должны время от времени пронизывать весь процесс обучения и научения.

Целью контроля над результатом в прогностическом аспекте является обеспечение основ для ранжирования учащихся по уровням знаний, а также для присуждения наград [5].

Оценивание деятельности учащихся – важный стимул обучения и воспитания, которое выполняет следующие функции [4]:

- констатация — фиксируется фактический уровень достижений;
- уведомление — информация о результатах сообщается заинтересованным сторонам;
- контроль — позволяет определить направления и объёмы дальнейшей работы;
- прямое воздействие — непосредственно регулирует учебную деятельность учащихся.

Практически можно различать три масштаба измерения результата: субъективный, относительный и абсолютный.

Субъективное измерение результата

«Субъективным» мы будем называть такое измерение результата, при котором преподаватель одни и те же результаты оценивает различными мерами. Такая оценка опирается обычно на факторы, которые находятся вне доказуемых результатов обучающегося. Так, субъективные оценки модифицируются в зависимости от старательности, способности студента, его характера и т.д. Примером такой оценки может являться то, что преподаватель ставит более высокую оценку студенту с большим прилежанием, но меньшими способностями, чем более способному, но менее прилежному студенту.

Относительное измерение результата

«Относительной» мы будем называть такую оценку результата, когда она используется в качестве масштаба среднего результата группы.

Этот вид оценки результата основан на представлении о «естественном» распределении результата, при котором только небольшая часть студентов может добиться очень хороших результатов, большая часть может достичь средних результатов и небольшая часть студентов закончит обучение с плохими результатами.

Эта ориентация на «нормальное распределение» имеет ряд недостатков, в частности:

- успех группы четко задает уровень результата, невозможно сравнивать оценки разных групп;
- мотивация результата менее одаренных студентов снижается; если все члены группы одинаково прилежны, то менее одаренные студенты все время находятся в конце реестра;
- предположение о нормальном виде распределения не является строго обоснованным. В действительности, не во всех случаях такое предположение справедливо, как будет показано далее.

Измерение результатов, ориентированное на цель.

«Абсолютное», или ориентированное на цель, измерение результата базируется на заранее заданной учебной цели. При этом соответствие измеренных значений нормальному распределению не является существенным. Существенно

лишь то, насколько точно контрольные задания идентифицируют достижение учебной цели [5].

Нормативно-ориентированный (norm-referenced) подход позволяет сравнивать учебные достижения отдельных учащихся друг с другом. Математические методы, применяемые при конструировании и оценке качества нормативно-ориентированных тестов, рассчитаны на модель нормального распределения и основываются в основном на корреляционном анализе [11].

Критериально-ориентированный (criterion-referenced) подход позволяет оценивать, в какой степени каждый учащийся овладел необходимым учебным материалом. Критериально-ориентированные тесты представляют собой систему заданий, позволяющую измерить уровень учебных достижений относительно полного объема знаний, умений и навыков, которые должны быть усвоены учащимися [11].

Подход к интерпретации тестового балла является основным критерием для разделения тестов на нормативно- и критериально-ориентированные. Вместе с тем эти два типа тестов различаются и по методам конструирования, и по особенностям применения.

В рамках нормативно-ориентированного подхода результат каждого испытуемого трактуется по отношению к некоторым нормам выполнения теста, которые устанавливаются эмпирически на выборке испытуемых [11].

В критериально-ориентированном подходе результаты интерпретируются по отношению к некоторой содержательной области и описываются в терминах усвоенных знаний, умений и навыков. В этих тестах закладывается то, что можно принять за 100%. Одним из важнейших принципов оценивания является объективность. Она выражается в том, что:

- одинаковые оценки характеризуют одинаковую успеваемость учащихся, близкую к целям учебных программ.
- разные учащиеся получают при одинаковой успеваемости одинаковые оценки.

- в основе оценки лежит фактическая успеваемость, а не субъективные мнения, возможности, условия учебы.

Выставление оценок носит конкретный характер:

Решения об оценках всегда ориентируются на совершенно конкретную ситуацию. [6].

Таким образом, для объективности результатов, и возможности сравнения результатов различных групп между собой необходимо использовать объективные оценки в критериально-ориентированном подходе.

В настоящее время в системе образования выделяют следующие виды контроля знаний [13]:

итоговый,
рубежный,
текущий.

В настоящее время в системе высшего профессионального образования используются три основных метода оценки знаний:

экзамен
опрос
собеседование
тест.

Каждый из этих методов решает свои собственные задачи и обычно имеет одну область применения.

Мы будем рассматривать тестирование, как обладающее следующими достоинствами: однозначность, документальность. Рассмотрим способы проведения тестирования в группе.

В настоящее время в различных источниках наблюдается большое количество названий форм тестовых заданий. Выделим четыре основные формы тестовых заданий [11].

Задания на множественный выбор (задания с закрытой формой). Преподаватель формулирует основную часть задания и ответы к заданию. Среди ответов, как правило, один правильный. Для контроля такие задания очень

удобны, т.к. можно охватить большой объем материала и затраты на проверку результатов невысоки. Задания закрытой формы наиболее технологичны.

Открытые задания. В открытых заданиях учащийся сам формулирует ответ в виде числа, формулы, слова и т. д. Эти задания очень нетехнологичны, однако делать их достаточно легко. И они часто используются педагогами в процессе текущего контроля.

Задания на соответствие. Обычно инструкция к таким заданиям выглядит так: «дополните», «установите соответствие».

Задания на установление правильной последовательности действий. Типовая инструкция представляется так: «установите правильную последовательность», «выстройте в правильной последовательности». Данный вид заданий удобно использовать при проверке знаний, например, по истории [11].

По способу предъявления различают тесты:

- бланковые (машиночитаемые бланки, тестовые тетради);
- предметные (манипуляция материальными объектами рассчитана на быстроту реакции и четкость действий);
- аппаратурные (с использованием специальной аппаратуры — датчиков для фиксации сигналов);
- практические (аналоги лабораторных работ, но с тестовыми условиями заданий);
- компьютерные (как частный случай — адаптивные).

По характеру действий тесты делят на:

- вербальные;
- невербальные.

По ведущей ориентации выделяют:

- тесты на скорость;
- тесты на результативность (мощность);
- смешанные тесты.

По степени однородности задач тесты делят на:

гомогенные, позволяющие измерить одно качество (уровень подготовки по одному предмету) однородными по составу заданиями;

гетерогенные (многомерные) — для измерения уровня подготовленности учащихся по нескольким учебным предметам и (или) свойствам личности.

Таким образом, по способу проведения непосредственно теста можно выделить два основных способа:

компьютерные;

бланковые.

Компьютерное тестирование так же может быть разделено на локальное (когда тестирующая программа установлена на одном или нескольких компьютерах, не связанных между собой) и сетевое (когда тестирующая программа использует интернет/интранет сетевые технологии).

Выбор бланкового или компьютерного вариантов тестирования определяется требованиями к документальности теста и наличием необходимых компьютерных ресурсов.

Согласно общепринятым представлениям, все тестовые задания могут быть отнесены к одному из следующих видов [4, 6]:

закрытое задание с выбором одного ответа из предложенных;

закрытое задание с выбором нескольких ответов из предложенных;

закрытое задание в котором необходимо установить порядок элементов;

закрытое задание с необходимостью установления взаимосвязей между элементами;

открытое задание с кратким ответом (в котором необходимо записать ответ словом, словосочетанием или числом);

открытое задание с развернутым ответом (в котором необходимо записать развернутый ответ в виде текста и/или последовательности формул).

В большинстве случаев, без ущерба для качества самого теста, все типы закрытых заданий могут быть заменены закрытым заданием с выбором нескольких ответов из предложенных. Кроме того, следует учитывать, что

автоматизированной проверке могут быть подвергнуты все типы тестовых заданий, за исключением открытого задания с развернутым ответом [8].

Отдельно следует отметить, что ценность обоих типов открытых заданий может быть поставлена под сомнение, поскольку снижение уровня владения русским языком и культурой письменной речи могут приводить к неправильной оценке компетенций, знаний, умений и навыков респондента. Это вызвано тем, что ошибка в написании ответа для открытого теста будет воспринята как ошибочный ответ.

Таким образом, использование контрольно-измерительных материалов, содержащих только закрытые задания, является не только допустимым, но в большинстве случаев и предпочтительным. Как будет показано далее, тесты, содержащих такие типы заданий, могут быть оценены объективно.

В общем случае процесс тестирования представляет собой оценку количества правильно и неправильно данных студентом ответов. Существует несколько подходов, используемых для такой оценки.

1. Оценка в виде "сдал/не сдал".
2. Оценка пятибалльная.
3. Оценка в процентах правильных ответов.
4. Оценка в условных единицах.

Рассмотрим подробнее каждый из подходов.

Оценка в виде "сдал/не сдал" сама по себе более подходит для зачета, чем для рубежной проверки знаний. Она лишена каких-либо градаций, и это серьезно ограничивает область ее применения.

Оценка пятибалльная в целом является общепринятой, и до сих пор распространена шире других. Она наиболее привычна как студентам, так и преподавателям. Тем не менее, существует проблема для данной оценки, и заключается она в сложности ее получения. Данная оценка в терминах вопросов и ответов теста является искусственной, она не выводится непосредственно из правильных или неправильных ответов студента. Например, если студент дал 40%

правильных ответов, какой оценке по пятибалльной шкале это должно соответствовать?

Оценка в процентах правильных ответов является наиболее понятной и с точки зрения "здорового смысла" наиболее "объективной", то есть, вызывает наибольшее доверие. В случае такой оценки требуется как минимум установление проходного процента, границы, отделяющей тех, кто сдал тест, от тех, кто тест не сдал. Установка такой границы субъективна и может вызывать споры.

Оценка в условных единицах – наиболее сложная оценка. Это связано с необходимостью разработки алгоритма, вычисляющего условную оценку по набору ответов студента. С одной стороны, это позволяет сделать такую оценку наиболее объективной, с другой стороны, студент, получающий такую оценку, часто с ней не согласен, каков бы ни был алгоритм ее вычисления.

1.2. Числовые меры теста

В настоящее время профессиональными тестологами выделяются следующие критерии, характеризующие как тест, так и КИМ [13]:

валидность;
мера трудности задания;
дифференцирующая способность;
вариация и дисперсия;
надежность теста.

Следует отметить, что часть данных критериев характеризует тест, а часть – прохождение теста (КИМ) конкретной группой лиц.

Под валидностью принято понимать соответствие тестового задания оцениваемым способностям. Фактически, валидность показывает, проверяет ли тест те знания, умения и навыки, которые он призван проверять. Валидность теста определяется по окончании этапа создания пула тестовых заданий, и не может измениться в момент формирования теста или контрольно-измерительного материала из пула заданий. Валидность теста может быть оценена только при условии глубокого владения материалом. Единственным способом оценки валидности теста является экспертная оценка. Критерий валидности теста является качественным, и для конкретного теста может принимать лишь значения «валиден/не валиден». Объективных способов оценки валидности теста не существует, поскольку в оценке валидности всегда принимает участие эксперт.

Мера трудности задания фактически определяет возможность прохождения теста всеми респондентами. Принято считать, что при слишком малой трудности задания его проходит большинство (либо все) респонденты, при слишком высокой сложности – проходит меньшинство респондентов (либо вообще никто). Такое определение показывает относительную природу этого показателя. Мера трудности не является характеристикой теста, а характеризует комбинацию теста и группы тестируемых лиц.

На практике меру трудности задания чаще всего определяют умозрительно (экспертная трудность), либо экспериментально, путем апробации тестового задания.

Следует отметить, что государственные стандарты высшего профессионального образования подразумевают совершенно определенный минимальный уровень компетенций, знаний, умений и навыков выпускника. Таким образом, при использовании тестирования для оценки знаний студентов возникает проблема оценки не относительного, а абсолютного уровня знаний.

Кроме того, при внедрении новых стандартов может возникнуть ситуация, когда тест будет использован для проверки знаний и умений студентов без возможности его предварительной апробации. Авторы отмечают, что в таких условиях особую важность приобретают возможности оценки характеристик теста без его апробации на реальных респондентах. Вопрос объективной оценки меры трудности тестового задания рассмотрен более подробно в [9].

1.3. Существующие средства автоматизированного тестирования

Рассмотрим существующие системы компьютерного тестирования.

На данный момент наиболее популярными системами компьютерного тестирования являются MyTest 3.0.4, OpenTest, тестовая подсистема Moodle, TestBuilder, Adaptive Computer Testing System, x-TLS, RATOS, Компьютерная Система Тестирования Знаний, АСТ-Тест Plus, МастерТест, УСАТИК.

В большинстве случаев, возможности всех этих систем достаточно близки, все эти системы обеспечивают базовую функциональность, включая создание и проведение тестов. Однако имеются и существенные различия.

Одним из базовых различий является ориентация системы на проведение локального или сетевого тестирования (часть систем способна работать и в том, и в другом режиме).

Различия систем определяются в том числе и дополнительными возможностями. К таковым можно, например, отнести возможность статистического анализа ответов тестируемых, возможность использовать в тексте вопросов и вариантов ответов мультимедийную составляющую, а именно, графические изображения, анимацию, видео, звук, и.т.д..

Некоторые из систем имеют ограничения на количество тем/вопросов/вариантов ответов (Adaptive Computer Testing System).

Из рассмотренных систем только Adaptive Computer Testing System обладает возможностью формировать бумажный вариант тестового задания.

Многие из рассмотренных систем представляют собой не среды разработки и проведения тестов, но и среды разработки обучающих систем, как, например, x-TLS, RATOS, Moodle.

Некоторые системы способны работать только в специально подготовленной среде. В первую очередь это относится к системам, использующим технологию клиент-сервер. Примерами таких систем являются x-TLS, RATOS, УСАТИК и АСТ-Тест Plus.

Системы, ориентированные на web-технологии (интернет/интранет тестирование) способны функционировать как в среде Linux, так и в среде

Microsoft Windows (OpenTest, x-TLS), причем для корректной работы могут потребоваться некоторые доработки.

Использование некоторых систем (OpenTest) требует участие коллектива специалистов.

Таким образом, ни одна из рассмотренных систем не обладает всем комплексом необходимых характеристик, что и обусловило необходимость разработки собственной системы тестирования.

1.4. Обзор методов решения

1.4.1. Требования к числовым характеристикам теста и оценки

Говоря о педагогических измерениях, нельзя не коснуться шкал измерения педагогических величин.

Принято различать следующие шкалы измерений [12].

Шкала наименований характеризуется только отношением эквивалентности какого-либо качественного проявления свойства. Пример такой шкалы измерений – классификация (оценка) цвета объекта по наименованиям (красный, белый, сине-зелёный и т. д.), опирающаяся на стандартные атласы цветов.

Шкала порядка описывает свойства, для которых имеют смысл не только отношение эквивалентности, но и отношение порядка по возрастанию или убыванию количественного проявления свойства. Характерный пример шкал порядка – шкалы чисел твёрдости тел, шкалы баллов землетрясений, шкалы баллов ветра и т. д. В такого рода шкалах в принципе нет возможности введения единицы измерений, также не имеют смысла суждения, во сколько раз больше или меньше проявления конкретных свойств. В шкалах порядка может быть (иметь смысл) нуль или его может не быть. Так, шкалы твёрдости начинаются с некоторого ненулевого значения, сейсмическая шкала начинается с одного балла, а шкала Бофорта для силы ветра - с нулевого значения.

Шкала разностей (интервалов) отличается от шкалы порядка тем, что для описываемого ею свойства имеют смысл не только отношения эквивалентности и порядка, но и пропорциональности или суммирования интервалов (разностей) между различными количественными проявлениями свойства. Характерный пример – шкалы времени; интервалы времени можно суммировать или вычитать, складывать же даты каких-либо событий бессмысленно. Шкалы разностей имеют условный нуль, опирающийся на какой-либо репер (например, шкала Цельсия).

Шкала отношений описывает свойства, ко множеству количественных проявлений которых применимы отношения эквивалентности, порядка, пропорциональности или суммирования (а следовательно, и вычитания, и

умножения). В шкале отношений существует естественный критерий нулевого количественного проявления свойства, т. е. нуль имеет не условное значение, а вполне определенный физический смысл. Примеры шкал отношений - шкала массы, термодинамическая температурная шкала.

Абсолютные шкалы обладают всеми признаками шкал отношений, но дополнительно в них существует естественное однозначное определение единицы измерения. Такие шкалы измерений соответствуют относительным величинам – отношениям одноимённых физических величин, описываемых шкалами отношений. Среди абсолютных шкал выделяются ограниченные по диапазону шкалы, значения которых находятся в пределах от 0 до 1. Они характерны для КПД, амплитудной модуляции и т. п. величин.

Очевидно, если мы говорим об объективной силовой оценке знаний, особенно о критериально-ориентированной оценке, мы должны использовать абсолютную шкалу ограниченного диапазона. Действительно, к уровню знаний применимы отношения эквивалентности, порядка и пропорциональности. В качестве естественного нуля выступает уровень знаний, статистически неотличимый от уровня знаний, демонстрируемого человеком, далеким от данной области знаний. Кроме того, для данной шкалы существует ограничение сверху. Это ограничение – наивысший уровень знаний, проверяемый тестом, в практическом выражении – это уровень знаний респондента, стабильно дающего правильные ответы на все вопросы контрольно-измерительного материала.

1.4.1. Вид распределения ответов студентов

Будем рассматривать процент правильных ответов, данных студентом, как случайную величину A , в силу многообразия влияющих на эту величину факторов.

В этом случае первая задача сводится к определению различий между двумя выборками – баллами, полученными студентом в ходе прохождения теста, и баллами, полученными при случайных ответах на вопросы теста. Вторая задача сводится к проверке гипотезы об отнесении студента к одной из нескольких

групп по полученным за тестирование баллам. Третья задача может рассматриваться с использованием различных критериев.

Рассмотрим подробно задачу определения факта прохождения студентом теста.

Задачу можно сформулировать как статистическую проблему выявления различий между двумя выборками – выборкой результатов, полученных случайным образом, и выборкой результатов конкретного студента.

В зависимости от вида распределения случайной величины для решения вышеуказанной задачи могут быть использованы различные виды статистических критериев. Особое значение приобретает небольшой объем выборки, поскольку именно три попытки являются разумным компромиссом между увеличением времени тестирования и повышением точности.

Таким образом, первоначальной задачей является определение нормальности распределения случайной величины A .

Следует рассматривать распределение процента правильных ответов как в выборке, включающей все три попытки, так и выборки, содержащие отдельно результаты первой, второй и третьей попыток.

Распределение процентов правильных ответов студента приведено на рис. 2.1. Данные получены по результатам тестирования студентов 1-3 курса факультетов АКФ и ФИСУ Сибирского государственного аэрокосмического университета. Общее количество тестов составляет 2498. Вычисленные оценки среднего и дисперсии составляют соответственно 65,31% правильных ответов и 18,70.

Количество интервалов определено в соответствии с формулой Стерджеса [14].

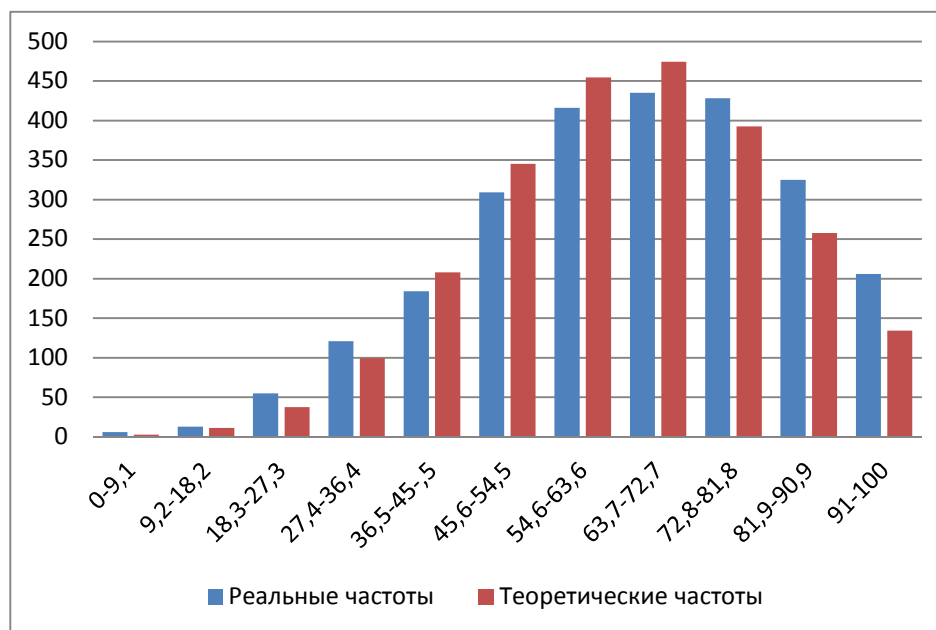


Рис. 2.1. Распределение частоты ответов по процентному количеству правильных ответов 1-3 попыток.

Визуальное сравнение не дает весомых оснований для отклонения гипотезы о нормальном виде распределения.

Для проверки статистической гипотезы H_0 о нормальности данного распределения против гипотезы H_1 о распределении величины, отличном от нормального, был использован критерий χ^2 Пирсона, как удовлетворяющий всем требованиям [14]. Число степеней свободы принято равным 8, поскольку среднее и дисперсия теоретического распределения оценивались по предложенной выборке. В результате анализа получено значение критерия Пирсона $\chi^2_{\text{эмп}}=73,72$, при теоретическом значении $\chi^2_{\text{теор}}=23,21$ при уровне доверительной вероятности $\alpha=0,01$, что позволяет с уверенностью отвергнуть гипотезу H_0 . Таким образом, распределение величины A не является нормальным при рассмотрении всей совокупности попыток сдачи теста. Независимое подтверждение данного вывода было получено при разбиении диапазона процентов правильных ответов на 39 интервалов. Поскольку использование критерия Пирсона в данном случае недопустимо [14], был использован λ -критерий Колмогорова-Смирнова. В результате анализа получены значения

$\lambda_{\text{эмп}}=0,036$ с уровнем доверительной вероятности $\alpha=0,01$ при теоретическом значении $\lambda_{\text{теор}}=0,027$.

Таким образом, различия следует признать достоверными.

При рассмотрении отдельно первой, второй и третьей попыток картина меняется.

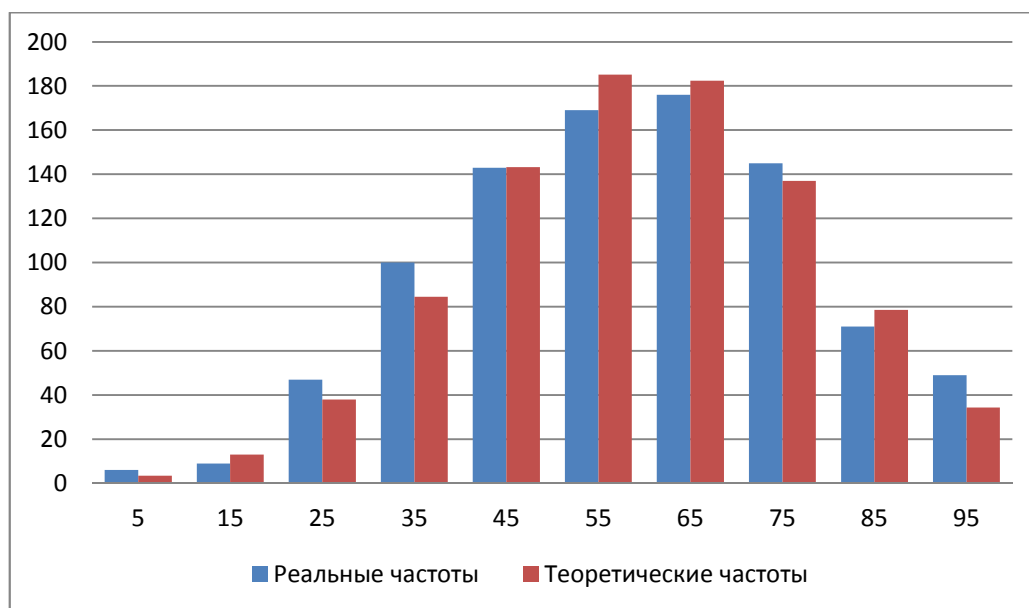


Рис. 2.2. Распределение частоты ответов по процентному количеству правильных ответов 1 попытки.

Распределение в первой попытке (915 тестов из 2498, среднее 59,46%, дисперсия 18,98) (рис. 2.2) не отличается от нормального (критерий χ^2) при уровне доверительной вероятности $\alpha=0,01$ ($\chi^2_{\text{эмп}}=17,38$, $\chi^2_{\text{теор}}=21,67$), но при $\alpha=0,05$ ($\chi^2_{\text{теор}}=16,92$) различия могут быть признаны достоверными. Критерий Колмогорова-Смирнова различий не выявляет при уровнях доверительной вероятности $\alpha=0,01$ и $\alpha=0,05$. Таким образом, для первой попытки однозначных выводов сделать нельзя.

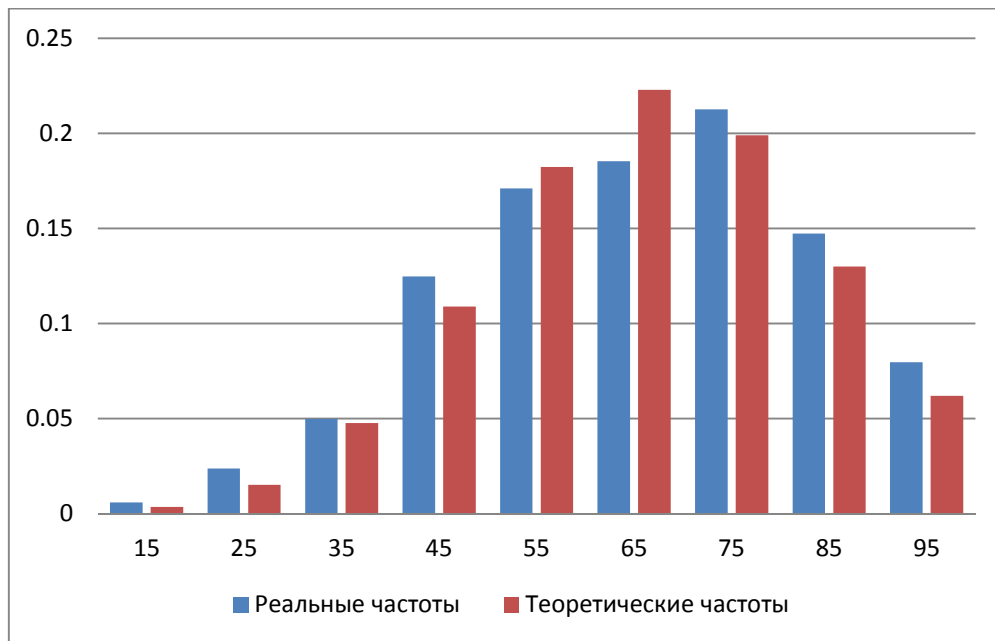


Рис. 2.3. Распределение частоты ответов по процентному количеству правильных ответов 2 попытки.

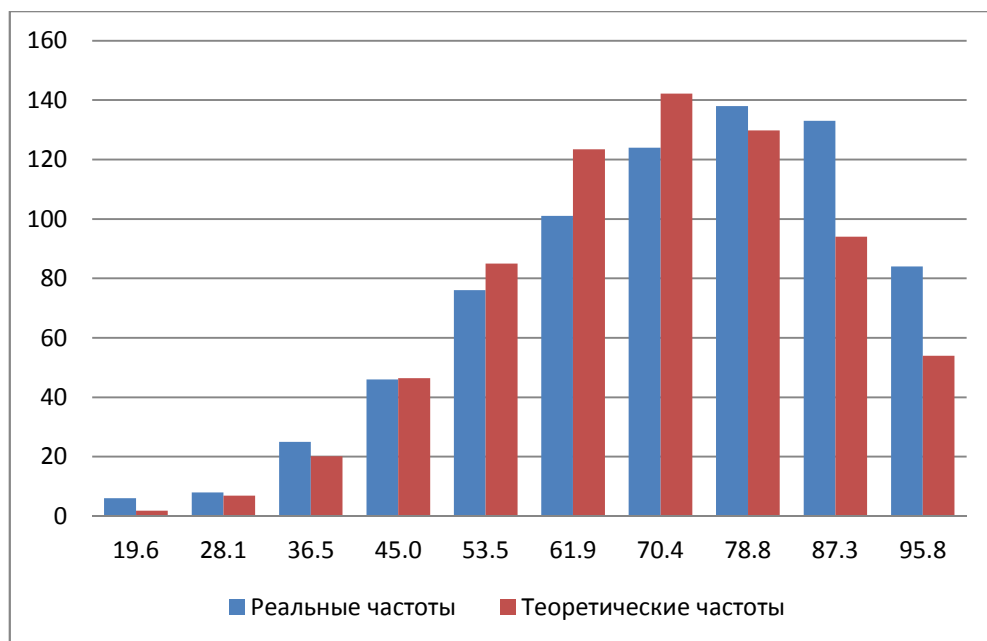


Рис. 2.4. Распределение частоты ответов по процентному количеству правильных ответов 3 попытки.

Распределение во второй попытке (рис. 2.3.) было проанализировано с использованием критерия Колмогорова-Смирнова (842 тестов из 2498, среднее 66,41%, дисперсия 17,62, , уровень доверительной вероятности $\alpha=0,05$). Отличий от нормального распределения не выявлено ($\lambda_{\text{эмп}}=0,029$, $\lambda_{\text{теор}}=0,047$).

Распределение в третьей попытке (рис. 2.4) (741 тест из 2498, среднее 71,30%, дисперсия 17,40, критерий χ^2 , уровень доверительной вероятности $\alpha=0,01$) не соответствует нормальному ($\chi^2_{\text{эмп}}=51,15$, $\chi^2_{\text{теор}}=21,67$).

1.5. Выводы по главе

В первой главе работы был проведен анализ основных методов и подходов инженерной педагогики, на основе данного анализа рассмотрены методы контроля результатов обучения и измерения этих результатов.

Рассмотрены числовые характеристики (меры), описывающие как тест, так и контрольно-измерительный материал.

Рассмотрены существующие системы компьютерного тестирования, в результате анализа установлено, что существующие системы не обеспечивают в полной мере объективной оценки знаний респондентов.

На основе требований, предъявляемых к методам оценки знаний вообще и к тестам и контрольно-измерительным материалам в частности, описаны требования к числовым характеристикам теста, контрольно-измерительного материала и результатам теста.

2. МЕТОДИКА ОЦЕНКИ УРОВНЯ ЗНАНИЙ

2.1. Определение оценки за тест

Прежде всего, определим, как следует учитывать отдельные пункты ответа. В различных системах учет ответов построен по-разному. В некоторых системах, если в одном из пунктов вопроса студент допустил ошибку, считается, что он ответил на вопрос неправильно. В других системах каждый пункт ответа учитывается индивидуально. Например, если в вопросе 4 пункта ответа, и студент допустил ошибку в одном из них (например, не пометил правильный пункт), считается, что ответ на вопрос правилен на 75%.

С точки зрения объективности оценки более предпочтительным следует считать второй вариант.

Оценка за тест носит относительный, а не абсолютный характер, поскольку нас интересует не само количество вопросов, на которые студент ответил правильно, а их доля в общем количестве вопросов.

Для нахождения относительной оценки будем использовать 2 характеристики. Первая – минимум из двух значений: отношения правильно помеченных студентом пунктов ответов к общему их количеству и отношения не помеченных студентом пунктов ответов к общему их количеству. Обозначим эту величину R (2.1).

$$R = \min\left(\frac{M_s}{M}, \frac{U_s}{U}\right), \quad (2.1)$$

где M_s – количество пунктов ответов, которые студент пометил как правильные;

M – количество пунктов ответов, которые студент должен был пометить;

U_s – количество пунктов ответов, которые студент не пометил;

U – количество пунктов ответов, которые студент не должен был пометить.

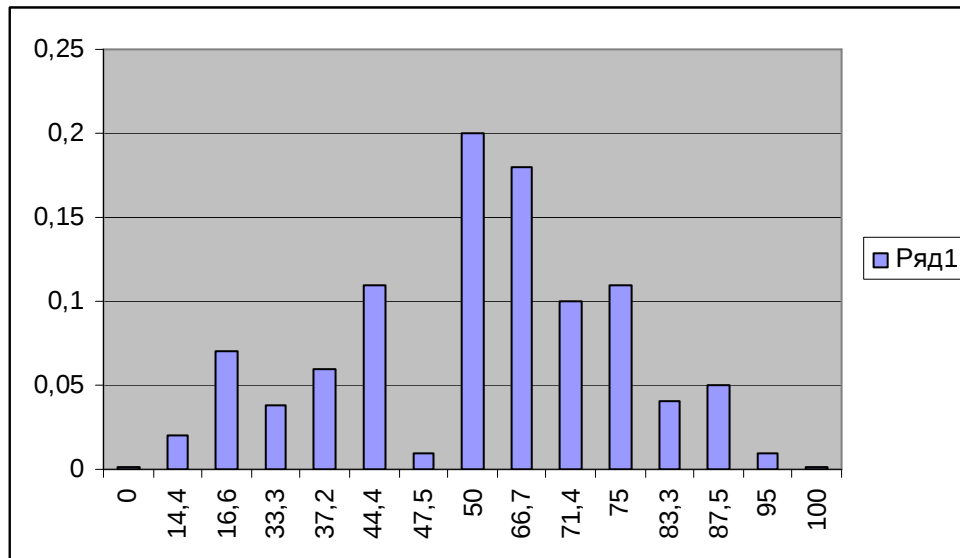


Рис. 2.5. Распределение вероятностей получения того или иного количества процентов правильных ответов.

Следует отметить, что величина R не является непрерывной, а представляет собой конечное множество фиксированных значений. Причем для заданного набора вопросов, предъявленных студенту в процессе тестирования, вероятность получения каждого конкретного значения величины R разная (рис. 2.5).

На рис. 2.5. приведен пример распределения вероятностей набрать то или иное количество баллов в процессе тестирования. Обратите внимание на то, что распределение не совпадает с классическим распределением Гаусса. Это вызвано, в частности, тем, что на отдельные вопросы студент должен выбирать отдельные варианты ответов, а в других вопросах – выбирать несколько правильных вариантов из предложенных.

Вторая характеристика – это значение величины R , которое может быть получено с заданной вероятностью α . Обозначим эту величину R_α . Эта величина – граничная, она отделяет тех, кто сдал тест, от тех, кто не сдал. Данная величина представляет собой значение величины R , соответствующее определенному, выбранному преподавателем значению α .

Фактически α – это вероятность того, что студент наберет заданное количество баллов. Зная R , мы можем определить соответствующее ему значение

α , и наоборот, задав значение α , можем определить, какая величина R может быть получена с такой вероятностью.

В математической статистике и теории вероятностей эта величина обычно называется уровнем значимости, и ее типичные значения – 0.05 или 0.01. Значение 0.05 соответствует одному шансу из двадцати набрать заданное количество баллов, 0.01 – одному шансу из ста.

Для практических целей при проведении тестирования достаточно ограничиться значением 0.05, поскольку при уменьшении значения до 0.01 диапазон значений баллов, которые может получить студент за тест, резко сужается, что приводит к уменьшению числа градаций и в итоге к снижению точности оценки, получаемой студентом.

Каким образом можно определить величину R_α при заданном значении α ? Теоретически, для этого достаточно определить все возможные варианты ответов студента на конкретный тест, определить, какой балл соответствует каждому из вариантов ответа, и определить вероятность каждого из них.

Однако расчеты показывают, что даже при 10 вопросах в тесте, и 3 вариантах ответа на каждый вопрос общее число комбинаций достигает 8^{10} , или 1073741824. Разумеется, все эти варианты сводятся к конечному (и довольно небольшому) числу баллов за тест, но чтобы определить вероятность выпадения определенного числа баллов, необходимо учесть все варианты. Учитывая, что для каждого теста граница вычисляется индивидуально в зависимости от заданных студенту вопросов, решить такую задачу в обозримый срок не представляется возможным.

Для разрешения этого противоречия предлагается использовать методы имитационного моделирования. В частности, имитируются ответы студента на те вопросы, которые реально были заданы в ходе теста.

Использование имитационного моделирования позволяет получать более точные оценки за счет определенных особенностей. Например, часть вопросов предполагает выбор одного из вариантов ответов, а часть – выбор нескольких

вариантов. Моделируя такое поведение студента, можно получить более точные оценки R_α .

Таким образом, после каждого прохождения студентом теста вычисляется индивидуальное для заданного студенту набора вопросов значение R_α , на основе которого и вычисляется балл студента.

Таким образом, задача определения того, пройден ли тест респондентом, сводится к проверке статистической гипотезы, которая может быть сформулирована следующим образом: процент правильных ответов, данных респондентом, превышает процент правильных ответов, который может быть получен случайно, с заданной доверительной вероятностью α .

Сформулируем соответствующие статистические гипотезы:

H_0 – средние выборок C и C_c совпадают и,

H_1 – средние выборок C и C_c различаются, где

C – выборка из генеральной совокупности всех возможных вариантов ответов на тест (контрольно-измерительный материал),

C_c – выборка из генеральной совокупности всех возможных ответов данного индивида.

В некоторых случаях возникает задача не только определения, обладает ли индивид каким-либо набором знаний в данной области, но и каков уровень этих знаний. Решение этой задачи удобнее представить в какой-либо числовой шкале, имеющей характер не ниже порядковой.

Как перевести полученные характеристики R и R_α в шкалу 0-10?

Определим разность $R_s = R - R_\alpha$. При величине $R_s \leq 0$ тест не пройден, оценка «неудовлетворительно». В противном случае масштабируем величину R_s таким образом, чтобы она принимала значения из интервала $(0;10]$ [9]. В результате получаем искомую оценку.

2.2. Оценка объективной меры сложности задания

Мера сложности теста, в общем случае, определяет, какое количество респондентов способны пройти данное тестовое задание. Это означает, что мера

сложности теста зависит от респондентов, проходящих тест. Таким образом, одно и то же тестовое задание для разной аудитории будет иметь разную сложность.

Такая оценка сложности теста является неудовлетворительной, в случае, если мы говорим о тесте, как средстве проверки соответствия респондента некоторому заранее заданному уровню знаний, умений и навыков.

Следовательно, необходим объективный критерий, позволяющий оценить сложность теста вне зависимости от респондентов, а так же отделить респондентов, обладающих необходимыми знаниями, навыками и умениями от тех респондентов, кто не обладает ими в полном объеме.

В качестве такого показателя может быть использован процент правильных ответов респондента, который может быть получен случайно с заданной вероятностью. Поскольку такая оценка не может быть получена явно, для определения ее значения предлагается использовать методы имитационного моделирования.

Для этого предлагается использовать следующую процедуру: случайным образом ответим на вопросы теста. Для учета того факта, что в части вопросов предусмотрен выбор одного варианта из нескольких, а в другой части – выбор нескольких вариантов ответа из нескольких, будем пометать как правильные несколько вариантов ответа в тех вопросах, где разработчиком теста предусмотрены несколько правильных ответов, и один ответ – в остальных. Для выполненного таким образом задания рассчитаем процент правильных ответов. Повторив процедуру несколько раз, для каждого из возможных значений процента правильных ответов его частоту встречаемости. Выбрав наибольшее значение процента правильных ответов с минимально устраивающей нас частотой встречаемости, мы получим искомое значение его объективной меры трудности. Чем выше полученное значение, тем большее количество правильных ответов может быть дано случайно, и тем более простым является тест. Отдельно следует отметить, что данное значение не показывает, насколько сложно будет респонденту ответить на вопросы теста, оно лишь показывает, насколько большой процент правильных ответов может быть получен случайно.

2.3. Оценивание дифференцирующей способности задания

Дифференцирующая способность теста понимается как способность тестового задания дифференцировать сильных (способных) респондентов от слабых [3, 5, 11]. В частности, наиболее удобным способом дифференциации респондентов является использование числовых оценок в определенной шкале. Автором предлагается использование для сравнительной оценки респондентов (именно это и предполагает дифференцирующая способность) как минимум интервальной шкалы оценок. Наиболее естественным выглядит использование числовой шкалы в заданном диапазоне, например, шкалы целых или вещественных чисел в пределах от 0 до 100. В настоящее время дифференцирующую способность теста чаще всего [11, 13] оценивают в той или иной мере в зависимости от дисперсионных показателей результатов теста для большого числа респондентов. Однако, такая оценка дифференцирующей способности обладает двумя недостатками: необходимость апробации теста на возможно большем числе респондентов перед началом его реального использования и относительный, а не абсолютный характер этой оценки.

Собственно дифференцирующая способность теста может быть определена, как количество уровней знаний, потенциально выявляемых тестом. Принимая во внимание, что величина R может принимать только строго определенные значения, их количество и собственно величины определяют количество выявляемых уровней знаний. Например, рассматривая рис. 2.5 и принимая во внимание, что $R_{\alpha}=75$, количество выявляемых данным тестом уровней знаний составляет 5 шт.: 75 и менее (что соответствует незначимым отличиям между уровнем знаний студента и его отсутствием), 83.3, 87.5, 90 и 100. Необходимо учитывать, что для каждого контрольно-измерительного материала значение уровня R_{α} и набор возможных уровней, превышающих это значение, является индивидуальным.

Таким образом, числовым показателем дифференцирующей способности контрольно-измерительного материала является количество возможных значений

величины R , превышающих по своей величине $R\alpha$. Это количество соответствует количеству выявляемым контрольно-измерительным материалом уровней знаний.

2.4. Выводы по главе

Проведен анализ вида распределения оценок студентов в ходе тестирования, в результате анализа установлено, что данное распределение не подчиняется нормальному закону.

Предложен метод определения оценки знаний, отличающийся объективностью. Основными чертами данного метода являются статистическое оценивание отличия знаний, продемонстрированных респондентом в ходе теста от случайных знаний по теме теста, а так же комплексность, выражающаяся в учете частичных знаний по теме отдельного вопроса (в том случае, если на вопрос дан частично правильный ответ).

Так же предложен метод объективной оценки меры сложности контрольно-измерительного задания на основе методов математической статистики и имитационного моделирования.

Предложен метод оценки дифференцирующей способности задания на основе методов математической статистики и имитационного моделирования.

3. РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННОЙ МЕТОДИКИ В ВИДЕ НАБОРА ПРИЛОЖЕНИЙ ДЛЯ РАЗРАБОТКИ, ОЦЕНКИ ХАРАКТЕРИСТИК И ПРОВЕДЕНИЯ ТЕСТА

Внедрение системы тестирования студентов подразумевает комплексный подход к процессу проверки знаний. Фактически, это означает внедрение информационной системы, задачей которой будет как проверка знаний студентов, так и накопление и непрерывное использование тестовых заданий по различным дисциплинам.

3.1. Основные принципы построения системы тестирования

Такая информационная система должна быть выполнена в соответствии с основными принципами построения и функционирования информационных систем, среди которых важнейшими в контексте данной задачи являются принцип соответствия и интегральности.

Рассмотрим эти принципы более подробно.

Принцип соответствия в контексте данной задачи означает, что разработанная система должна предоставлять следующие возможности:

- хранение вопросов теста;
- хранение вариантов ответов на вопросы с указанием их правильности;
- хранение сформированных индивидуальных тестовых заданий;
- хранение информации о студентах;
- хранение ответов студентов на вопросы тестовых заданий;
- определение процента правильных ответов студентов на вопросы тестовых заданий;
- определение границы процентов правильных ответов, которые можно получить случайным образом;
- определение и нормирование превышения студентом границы случайных ответов;
- определение количества ошибок, допущенных студентом при прохождении теста по темам тестового задания в процентном отношении;

сохранение результатов прохождения студентом тестового задания;

возможность проведения тестирования в произвольном компьютерном классе с минимальной предварительной подготовкой и минимально возможными конфликтами с уже установленным на компьютере программным обеспечением.

Принцип интегральности в общем случае требует многократного использования однажды введенной в систему информации. Для нашей задачи этот принцип будет означать многократное использование вопросов теста в различных тестовых заданиях, в том числе и в тестовых заданиях по различным дисциплинам, а также хранение ответов студентов на вопросы тестовых заданий. Помимо этого, возникает вопрос использования вопросов и тестовых заданий в различных вариантах теста, а именно, в бумажном и безбумажном вариантах. Следует так же учесть, что тестирование в безбумажном варианте может предполагать как локальное, так и удаленное (интранет- или интернет-тестирование). Учитывая широкое распространение операционных систем семейства Linux наряду с семейством Microsoft Windows, оправданной представляется разработка программного модуля, допускающего использование во всех перечисленных вариантах. Важной также представляется возможность анализа ответов студентов по различным статистическим критериям.

Кроме того, принцип интегральности в данном контексте подразумевает использование такого способа хранения информации в системе, который мог бы обеспечить импорт и экспорт информации в систему в каком-либо универсальном формате.

Вновь разрабатываемая система должна соответствовать принципам соответствия и интегральности, в том виде, как они были рассмотрены ранее.

Рассмотрим способы хранения информации в системе.

Принцип интегральности подразумевает высокие требования к формату хранения данных в системе. Выбранный формат должен обеспечивать возможность извлечения информации и помещения информации в систему различными средствами и системами, относительно высокое быстродействие,

универсальность и открытость формата, а так же возможность его преобразования.

3.2. Основные технические решения

В качестве способов хранения информации рассмотрим базы данных, а так же различные форматы файлов, в том числе XML, YAML, JSON, CSV.

Реляционные СУБД были исключены из рассмотрения по причине закрытого формата данных и использования некоторого сервера БД (требующего установки).

Соответственно, наиболее предпочтительным является хранение базы тестовых заданий и ответов студентов в одном из человеко-читаемых форматов.

Из таких форматов наиболее распространенными являются CSV, XML, JSON и YAML.

Рассматривать данные форматы мы будем с точки зрения следующих характеристик: поддерживаемые возможности, простота чтения и редактирования формата человеком, наличия стандартных библиотек по работе с данным форматом в распространенных языках и средах программирования, возможность конвертирования в другие форматы.

Формат CSV является одним из старейших форматов хранения информации в текстовом файле. Данный формат не является самодокументируемым в отличие от XML, JSON и YAML. Однако большинство существующих языков и сред программирования поддерживают данный формат. Практически отсутствуют стандартные средства конвертирования в другие форматы представления информации, однако при необходимости такое конвертирование легко может быть выполнено программным путем.

Файл формата XML образует древовидную структуру, что позволяет легко хранить в этом формате такие структуры данных, как записи, списки и деревья. XML является платформо-независимым языком. При необходимости формат допускает хранение бинарных данных.

Формат JSON может реализовать такие структуры данных, как запись, словарь, и ассоциативный массив (а так же некоторые производные от них). По простоте создания и редактирования человеком данный формат превосходит такие форматы, как XML и тем более CSV. Однако по распространенности библиотек для работы с данным форматом он существенно уступает XML и CSV.

Как и форматы XML и JSON, YAML представляет собой человекочитаемый язык представления данных. В данном формате можно сохранять такие структуры данных, как запись и массив (в том числе, массив записей, а так же производные от них). Часть из имеющихся библиотек являются платформозависимыми, что сужает диапазон его применения. Конвертация данного формата в другие осуществляется в основном программным путем.

На основе анализа вышеприведенных характеристик каждого из языков, было принято решение использовать в проекте язык XML.

Рассмотрим способы реализации системы.

К системе в целом предъявляются следующие требования: работа в среде Microsoft Windows/Linux, по возможности, без перекомпиляции; возможность использовать как в локальном, так и в сетевом варианте, по возможности, без дополнительного ПО или с минимальным набором. С целью соблюдения принципа интегральности было принято решение вести разработку системы в виде библиотеки классов. Такое решение позволяет в дальнейшем использовать данную библиотеку как в локальном варианте программы тестирования, так и в сетевом. Кроме того, правильное построение набора классов позволяет также легко реализовать как безбумажный, так и бумажный вариант тестирования.

Выбранный способ реализации системы и особенности поставленной задачи предъявляют определенные требования к языку и среде разработки. В частности, необходимыми особенностями являются объектная ориентированность языка, поддержка языка XML стандартными средствами или посредством сторонних библиотек, возможность создания как локальных, так и сетевых приложений, возможность исполнения программы как в среде Microsoft Windows, так и Linux. Крайне желательным является наличие в языке строгой типизации для

минимизации ошибок времени исполнения. С целью повышения безопасности системы предпочтительным является автоматическое управление памятью. Крайне важным требованием так же является обработка исключительных ситуаций средствами языка. Желательной также является возможность компиляции созданной библиотеки.

К современным кроссплатформенным языкам относятся языки C/C++, C#, Java, Python, PHP. С точки зрения универсальности проекта предпочтительнее использовать один из этих языков. Рассмотрим их с точки зрения следующих вышеприведенных критериев.

Язык C++ – компилируемый язык, что подразумевает необходимость перекомпиляции библиотеки под целевую платформу. В языке поддерживается строгая типизация, однако отсутствует автоматическое управление памятью. Поддерживается обработка исключений.

Программа на языке C# компилируется в специальный байт-код (аналогично Java), исполняемый в специальной среде .Net Framework CLR. На данный момент времени среды выполнения существуют в Microsoft Windows и в Linux. Язык полностью поддерживает строгую типизацию и использование разнообразных коллекций на уровне стандартных, встроенных в среду библиотек. Управление памятью полностью автоматическое. Имеется поддержка исключительных ситуаций.

Язык Java аналогично .Net Framework использует специальную среду выполнения, называемую JVM (Java Virtual Machine). Имеются реализации JVM для Microsoft Windows и Linux. Язык использует строгую типизацию и автоматическое управление памятью. Имеется обработка исключительных ситуаций средствами языка. Различные коллекции реализованы как набор стандартных классов. Для работы с XML имеются разнообразные наборы классов, как от разработчика языка, компании Sun Microsystems, так и от независимых разработчиков. В отличие от C#, изначально язык Java такой поддержкой не обладал.

Язык Python является объектно-ориентированным, хотя поддерживает и другие парадигмы разработки. Используется динамическая типизация. Изначально язык является интерпретируемым, однако существует возможность компиляции. Основными коллекциями в Python являются списки и словари. Поддержка XML реализована через стандартную библиотеку или сторонние библиотеки. Управление памятью – автоматическое.

Интерпретируемый язык PHP первоначально был разработан для создания динамических веб-страниц. Язык изначально не поддерживает строгую типизацию и обработку исключений, и даже в версиях 5.x.x существует возможность разработки, не используя данные принципы. Управление памятью – автоматическое, начиная с версии 5.x.x, но оно не является обязательным. Встроенные коллекции представлены исключительно словарями и массивами. Обработка исключений так же присутствует, начиная с версии 5.x.x. языка.

Исходя из вышеприведенных особенностей языков разработки, основными претендентами были выбраны Java и C#. Учитывая, что начиная с Microsoft Windows Vista среда .Net Framework является стандартным компонентом операционных систем семейства Microsoft Windows, а так же большое количество встроенных типов коллекций и возможности по работе с XML, языком разработки был выбран C#.

3.3. Анализ предметной области и выбор технических решений

Структура набора классов определяется спецификой задачи, поэтому перед реализацией системы необходим анализ предметной области с целью выделения сущностей предметной области и связей между ними.

В ходе анализа были выделены следующие сущности:

студент;

группа;

набор вопросов теста;

тестовое задание;

вопрос теста;

вопрос тестового задания;

вариант ответа на вопрос теста;

вариант ответа на вопрос тестового задания;

иллюстрация;

тестовое задание с ответами студента на вопросы.

На основании выделенных сущностей и их связей была построена объектная модель предметной области на языке C#.

Объектная модель включает в себя следующие классы и интерфейсы:

TestImage, TimesUpException, ITestObject, BaseObject, Student, Group, CommonTest, CommonAnswerVariant, WholeTest, WholeTestQuestion, WholeTestQuestionVariant, Test, TestQuestion, TestQuestionAnswerVariant, StudentTestAnswer, TestQuestionAnswer.

Рассмотрим их более подробно.

Абстрактный класс BaseObject является базовым для всех объектов библиотеки. В нем описаны члены, представленные во всех объектах предметной области, а так же он реализует виртуальные методы, описанные в интерфейсе ITestObject, и не специфичные для отдельных объектов, а именно SaveToXmlStream, LoadFromXmlStream, SaveToXmlFile, LoadFromXmlFile.

Класс Student описывает отдельного студента. Он содержит данные студента (Ф.И.О., номер студенческого билета), а так же ссылку на группу, в которой обучается студент.

Класс Group описывает группу. Группа характеризуется названием, специальностью, годом начала обучения, а так же списком студентов. В классе реализован итератор, позволяющий получить студента по его порядковому номеру в группе.

Класс WholeTest описывает полный набор вопросов, которые могут содержаться в тестовом задании студента. Он является потомком класса CommonTest. Он содержит список вопросов, а так же несколько конструкторов, позволяющих загрузить набор вопросов из файла, потока ввода-вывода, а так же

создать новый набор вопросов с заданным названием, специальностью, для заданной дисциплины и с заданной продолжительностью в секундах.

Кроме того, данный класс содержит метод `GenerateTest`, создающий новое тестовое задание для студента с заданным числом вопросов в нем. Вопросы тестового задания выбираются случайным образом из всех вопросов теста.

Класс `Test` описывает тестовое задание студента и наследует от класса `CommonTest`. Дополнительно данный класс содержит верхнюю границу оценки количества процентов правильных ответов, которую можно набрать случайным образом с уровнем достоверности 0,95. Кроме того, класс содержит метод, позволяющий получить случайный ответ на вопросы теста. Так же класс содержит функции, определяющую процент правильных ответов студента на вопросы теста (по заданным ответам студента, см. объект `StudentTestAnswer`) и его балл. Более подробно вопрос оценки ответов студента будет освещен далее.

Класс `TestQuestion` наследует от `CommonQuestion` и описывает отдельный вопрос тестового задания, выдаваемого студенту. Класс содержит порядковый номер данного вопроса в тестовом задании, а так же номер данного вопроса в полном наборе вопросов. Так же класс содержит варианты ответов на вопрос.

Класс `TestQuestionAnswerVariant` наследует от `CommonAnswerVariant`, и описывает вариант ответа на вопрос отдельного тестового задания. Он содержит порядковый номер варианта в вопросе и букву, описывающую вариант ответа.

Класс `StudentTestAnswer` наследует от `BaseObject` и описывает ответ тестируемого на вопросы тестового задания. Содержит ссылку на объекты `Student`, `Test`, содержит дату и время выполнения студентом теста, номер текущего вопроса, а так же количество секунд, прошедшее с начала тестирования (используется при компьютерном тестировании). Кроме того, данный объект содержит варианты ответов на вопросы тестового задания, которые студент отметил как правильные (`TestQuestionAnswer`).

Класс `TestQuestionAnswer` наследует от `BaseObject`, и содержит информацию о студенте, давшем данный ответ, тестовом задании, к которому

относится вопрос, на который ответил студент, а так же списка букв вариантов ответов, которые студент пометил как правильные.

3.5. Реализация системы компьютерного тестирования

С использованием разработанной библиотеки была построена комплексная система, включающая в себя:

модуль разработки теста;

генерации на основе теста набора контрольно-измерительных материалов в виде rtf или pdf документов;

проведения тестирования в среде локальной сети или глобальной сети интернет.

Интерфейс программы представлен рис. 3.1.

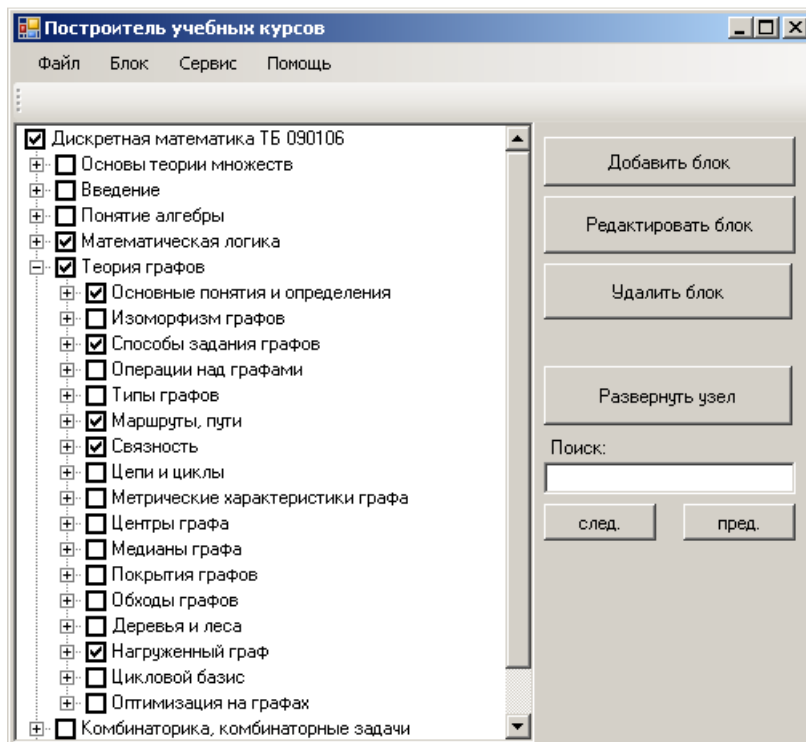


Рис. 3.1. Модуль создания тестов.

Структура учебного курса представляется в иерархическом виде. Имеется возможность добавлять, удалять, а так же редактировать выбранные блоки учебного курса.

Так же для выбранного блока имеется возможность добавлять, удалять и изменять как вопросы, так и варианты ответов к ним (рис. 3.2). Для каждого из выбранных вариантов ответа можно указать его правильность или неправильность. Как к вопросу в целом, так и к вариантам ответов могут быть добавлены графические иллюстрации. Добавление возможно как из файла (поддерживаются форматы JPEG и PNG), так и из буфера обмена.

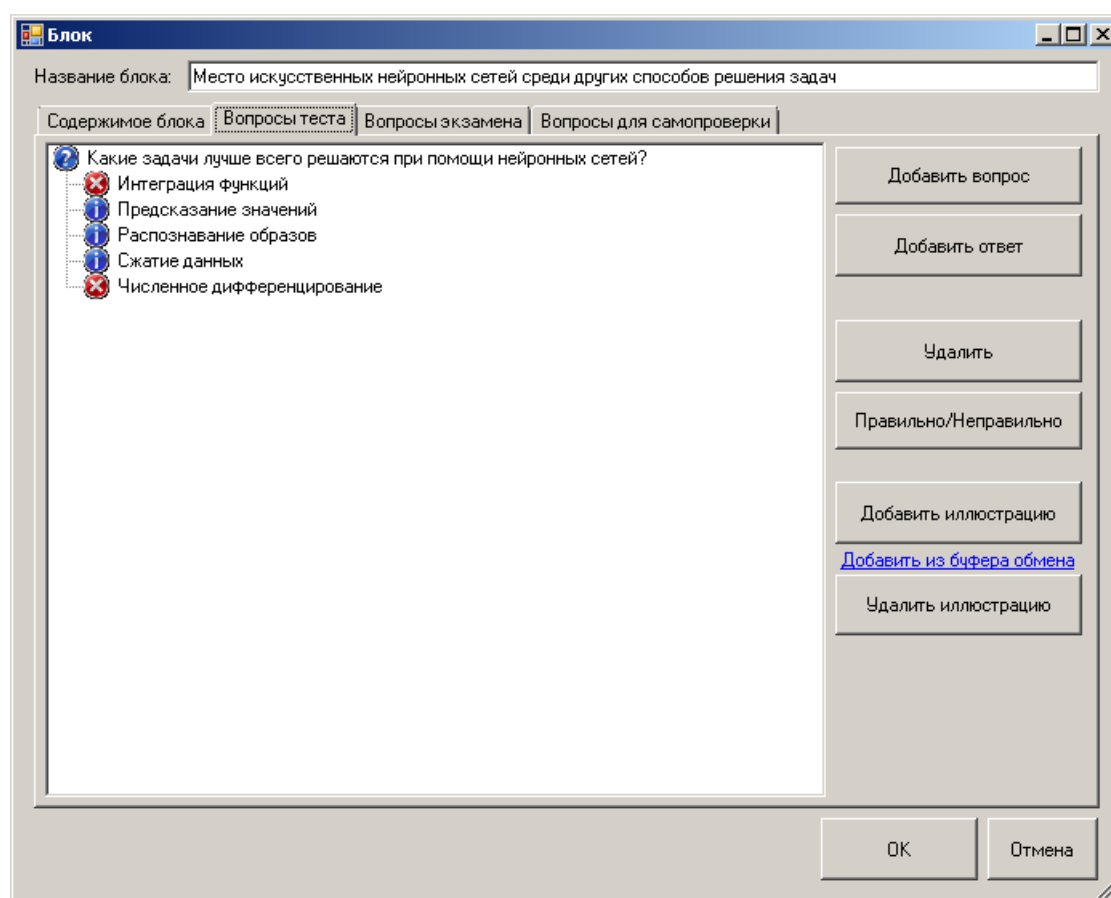


Рис. 3.2. Вопросы теста в выбранном блоке.

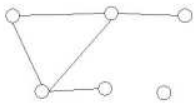
После того, как выбран материал теста, необходимо сгенерировать набор контрольно-измерительных материалов, а затем и проверить его. Для этой цели служит отдельный программный модуль, который может быть использован независимо (рис. 3.3).

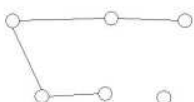
Рис. 3.3. Общий вид интерфейса модуля генерации и проверки КИМ.

Данное программное средство позволяет на основе материала теста построить набор контрольно-измерительных материалов. При этом можно выбрать как количество контрольно-измерительных материалов, которое будет сгенерировано, так и количество заданий в каждом контрольно-измерительном материале. В качестве выходного формата файла можно выбрать либо PDF, либо RTF-формат. При этом необходимо учитывать, что формат RTF в большей степени ориентирован на дальнейшее редактирование и печать бумажного варианта теста. В то же время, формат PDF пригоден как для печати, так и для интерактивного заполнения контрольно-измерительного материала. Пример PDF-файла приведен на рис. 3.4.

Тест N1

1. Выберите остовное дерево графа



☐ а) 

☐ б) 

☐ в) 

Рис. 3.4. Пример PDF-файла с тестом.

Как можно заметить, разработанные программные средства позволяют использовать графические изображения как в тесте, так и в контрольно-измерительном материале. Прямоугольник в верхней части страницы представляет собой интерактивное поле, предназначенное для указания личных данных респондента (фамилии, имени, отчества и группы).

Прямоугольники рядом с вариантами ответов так же представляют собой интерактивные поля типа `CheckBox`, то есть, респондент может пометить правильные с его точки зрения варианты ответов в файле PDF, а затем сохранить его для дальнейшей проверки программным обеспечением комплекса.

Кроме того, PDF-файл так же может быть распечатан и заполнен респондентов вручную.

Помимо генерации контрольно-измерительных материалов на основе теста, данное программное средство так же может быть использовано для проверки ответов респондента. Для этого служит раздел «Оценка результатов теста». При использовании этой функции, необходимо загрузить xml-файл с описанием набора контрольно-измерительных материалов («Файл набора тестов»), и выбрать

номер проверяемого контрольно-измерительного материала из выпадающего списка справа (рис. 3.5).

Рис. 3.5. Проверка результатов прохождения теста респондентом.

Далее, в зависимости от того, в каком виде получены ответы от респондента, проверка ответов может быть выполнена автоматически и вручную.

При автоматической проверке необходимо использовать экранную кнопку «Анализировать ответы PDF», и выбрать сохраненный респондентом файл формата PDF с ответами на вопросы.

При ручной проверке необходимо использовать нижнюю часть окна программы.

Экранная кнопка «Начать» используется для начала процедуры проверки ответов респондента, при этом перемещение между различными вопросами осуществляется с помощью экранных кнопок «Назад» и «Далее», текст текущего вопроса показывается в разделе «Текст вопроса», а варианты ответов – в разделе «Варианты ответов». Для определения результатов прохождения теста респондентом используется экранная кнопка «Закончить» (рис. 3.6).

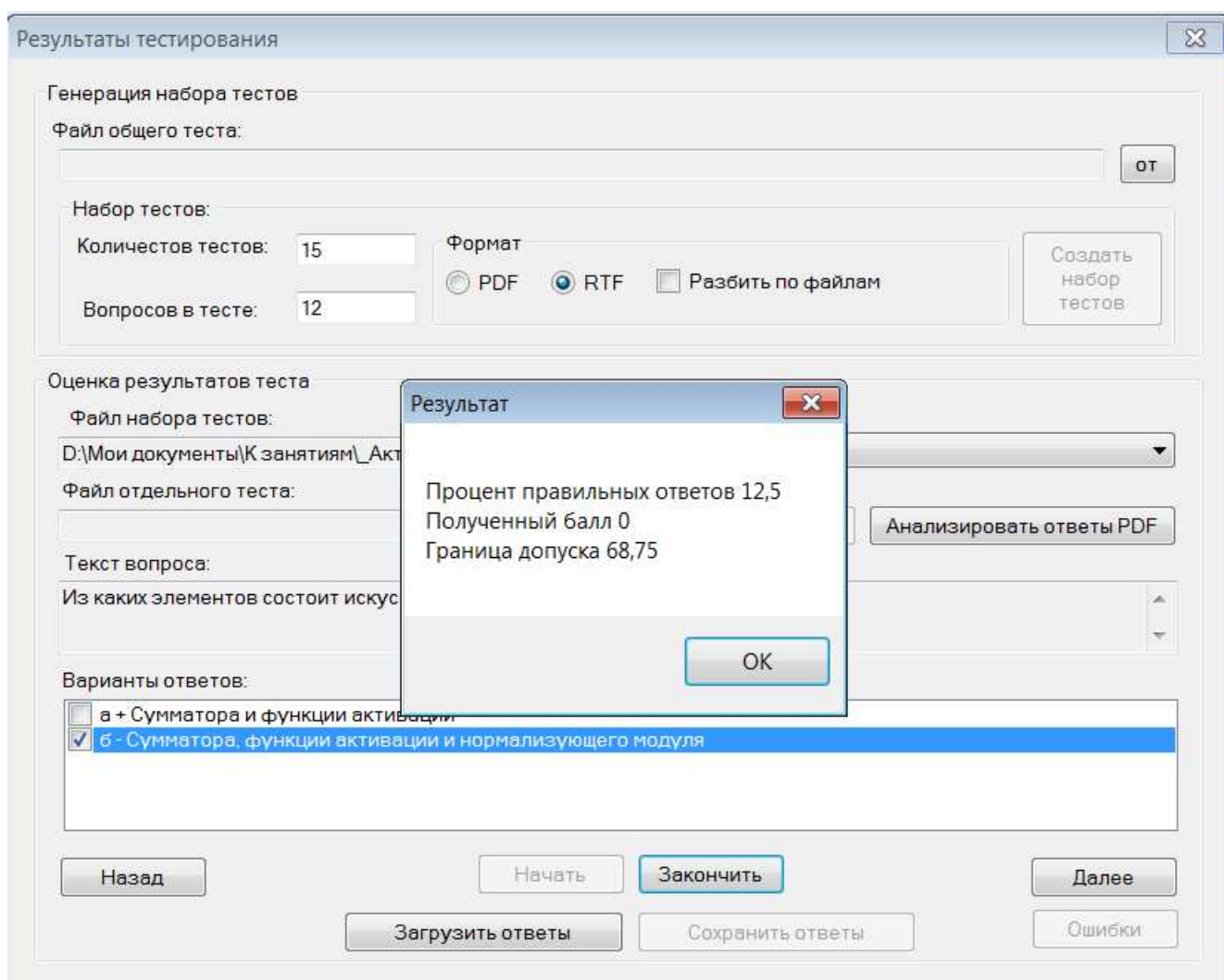


Рис. 3.6. Результаты прохождения теста респондентом.

Оператору выдается как величина набранных респондентом в ходе тестирования баллов, так и процент правильных ответов, а так же граничное значение, которое может быть получено путем выбора случайных вариантов ответов на вопросы в ходе тестирования с вероятностью 0.05.

Существует так же возможность просмотра тем, в которых респондент допустил ошибки. Для этой цели служит экранная кнопка «Ошибки» (рис. 3.7).

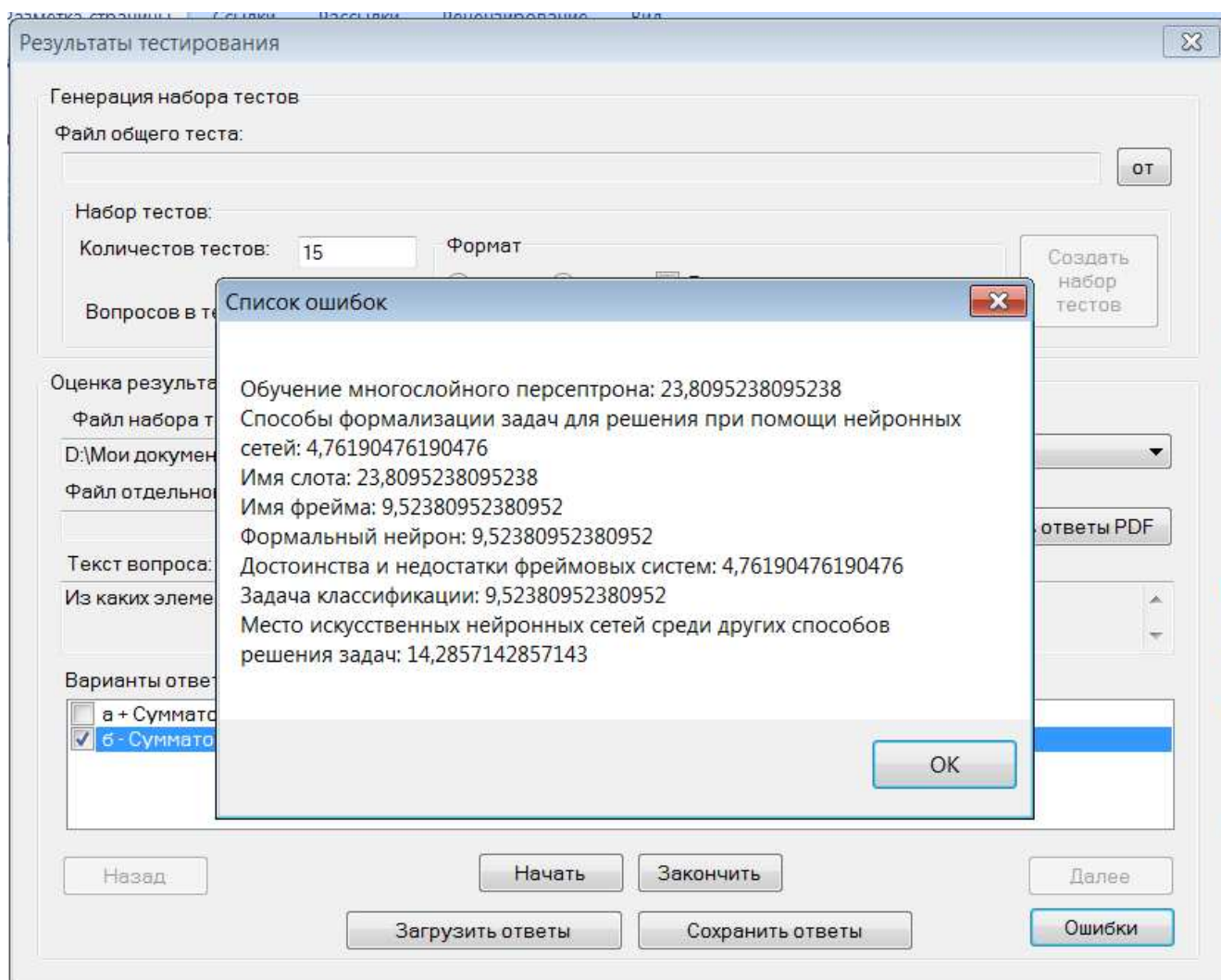


Рис. 3.7. Список ошибок, допущенных респондентом в ходе тестирования.

Для каждой темы указывается доля ошибочно отвеченных вопросов, относящихся к этой теме.

Кроме того, имеется возможность сохранить ответы респондента в файл формата XML (а так же загрузить их в дальнейшем, например, для уточнения). Для этой цели служат экранные кнопки «Сохранить ответы» и «Загрузить ответы».

Во многих случаях тестирование респондентов удобнее проводить в аудитории, оснащенной компьютерами. В этом случае можно либо использовать

сгенерированные контрольно-измерительные материалы в виде PDF-файлов, либо использовать специально разработанную систему.

Для проведения on-line тестирования был разработан сайт.

Веб-сайт олимпиады мог быть реализован различными способами. В качестве основы для разработки были рассмотрены PHP, Zend Framework, CMS Joomla, Microsoft ASP.NET, Microsoft ASP.NET MVC, Java Servlets.

Основными требованиями при выборе платформы были простота сопровождения системы, наличие удобных сред разработки и отладки приложений, модульность получаемого в результате разработки решения, скорость разработки. Требования простоты сопровождения и скорости разработки однозначно указывали на использование объектно-ориентированных языков программирования со строгой типизацией. По этой причине технологии, основанные на языке PHP, были исключены из рассмотрения, хотя Zend Framework и обладает почти всеми перечисленными свойствами. По этой же причине была исключена CMS Joomla. Кроме того, администрирование Joomla требует определенных навыков – настройка системы может потребовать существенного времени.

Таким образом, реальный выбор стоял между технологиями ASP.NET, ASP.NET MVC и Java Servlets. Для упрощения разработки, последующего сопровождения и облегчения доступа пользователей к системе было принято решение использовать шаблон разработки MVC (Model-View-Controller, Модель-Вид-Контроллер). При выборе между ASP.NET MVC [10] и Java Servlets [1] был сделан выбор в пользу первой. Одной из причин такого решения была возможность использования LINQ в среде ASP.NET MVC, а так же поддержка этой технологии некоторыми драйверами MongoDB. Использование языка LINQ существенно облегчает извлечение данных из NoSQL баз данных, а так же получение информации о статистике соревнований.

Использование среды ASP.NET MVC практически однозначно определяет операционную систему и веб-сервер, используемые системой. Однако использование серверной ОС и Internet Information Server противоречит

требованию простоты развертывания системы. Для устранения данного противоречия был использован веб-сервер, основанный на продукте Cassini. Это решение так же использует среду Microsoft .Net, однако не требует установки, и функционирует в любой ОС семейства Microsoft Windows, начиная с Microsoft Windows XP.

Кроме того, среда Microsoft .Net предоставляет широкие возможности по управлению выполнением сторонних программ. Microsoft .NET Framework 2.0 содержит специализированное пространство имен System.Diagnostics.Process, предоставляющее широкие возможности по созданию, запуску, остановке и уничтожению процессов, а так же управлению стандартным вводом-выводом процессов. Эти возможности позволяют гибко управлять тестированием программных решений студентов (рис. 3.8).

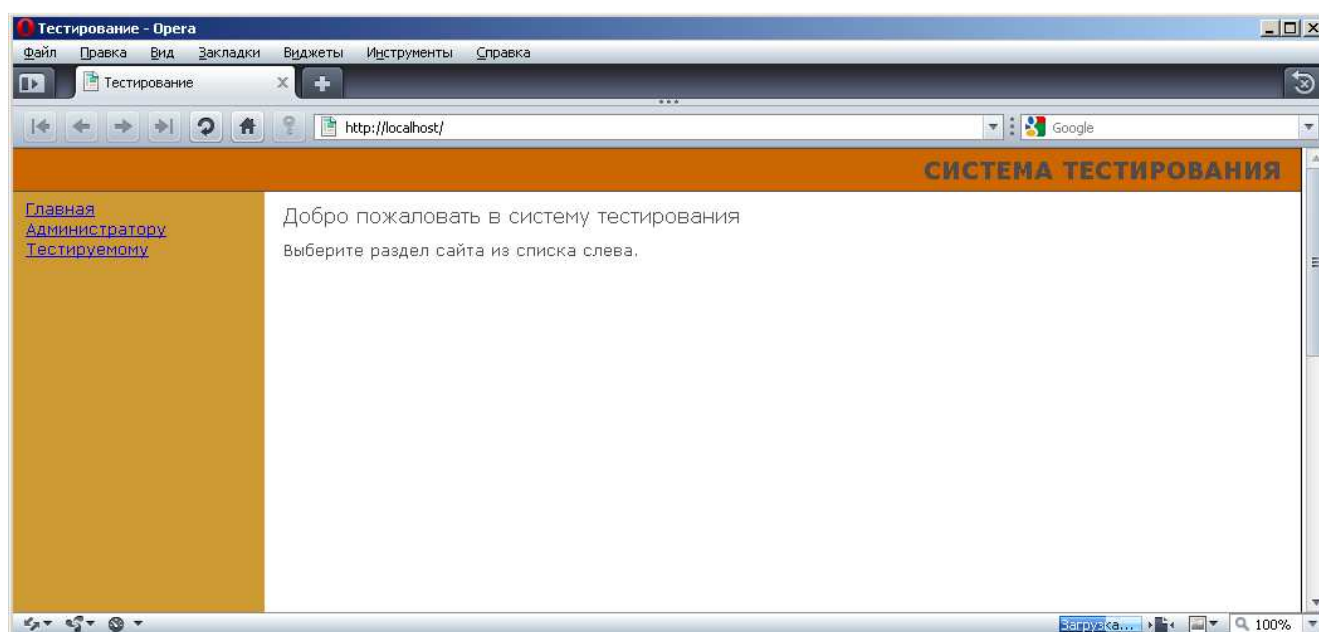


Рис. 3.8. Система on-line тестирования.

Интерфейс системы разделен на две части – интерфейс администратора и интерфейс респондента. Интерфейс администратора доступен только при условии ввода правильного имени пользователя и пароля.

Интерфейс администратора (рис. 3.9) предоставляет следующие возможности:

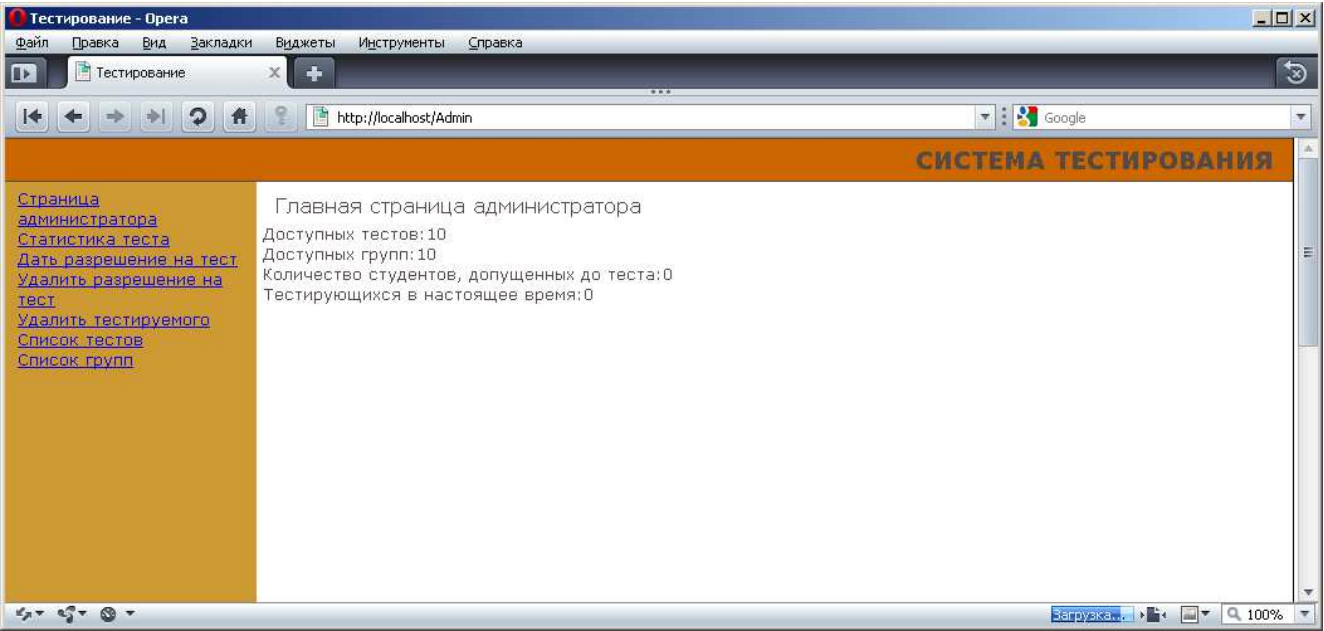


Рис. 3.9. Интерфейс администратора системы on-line тестирования.

просмотр статистики тестирования, включая имена и фамилии респондентов, как прошедших тестирование, так и тестирующихся в настоящий момент;

выдача разрешений на прохождение теста с указанием количества возможных попыток (контрольно-измерительный материал генерируется независимо для каждой из попыток) (рис. 3.10);

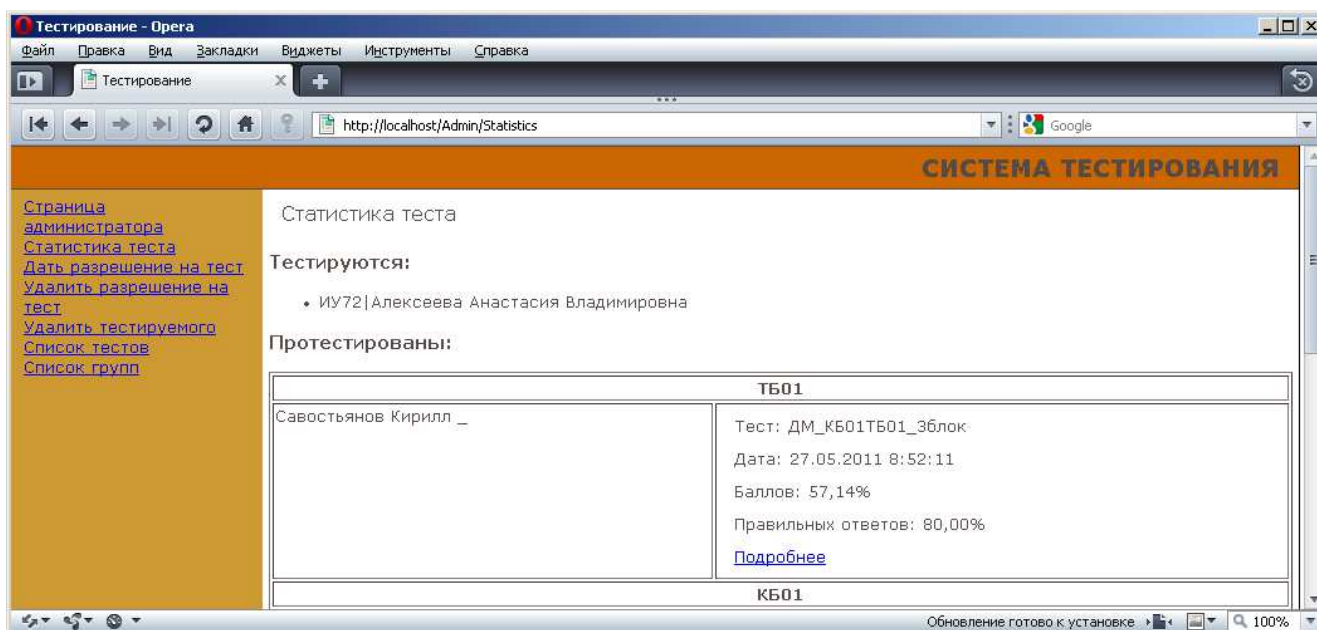


Рис. 3.10. Выдача разрешения на прохождение теста.

Студент, проходящий тест, взаимодействует с сайтом, выбирая правильные на его взгляд варианты ответов на предложенный вопрос.

При ответе студент имеет возможность перейти к произвольному вопросу контрольно-измерительного материала.

Кроме того, при проведении тестирования по предложенному варианту время, в течение которого принимаются ответы, ограничено. Оставшееся время демонстрируется в левой части окна.

По окончании теста выдается результат, респондент может оценить как полученные баллы, так и процент данных им правильных ответов, а так же нижнюю границу прохождения.

3.6. Выводы по главе

На основе методов, предложенных в главе 2 данной работы, разработана объектная библиотека, предназначенная для описания, составления, оценивания характеристик теста и контрольно-измерительного материала, сохранения результатов тестирования. Библиотека реализована на языке C# и предназначена для работы в среде Microsoft .Net Framework 2.0.

На основе предложенной библиотеки разработаны приложения для составления теста и генерации на основе теста контрольно-измерительного материала, проверки результатов тестирования как в бланковом, так и в электронном виде, для проведения on-line тестирования в среде локальной сети и сети интернет.

ЗАКЛЮЧЕНИЕ

В результате проведенной работы решена актуальная научно-техническая задача модернизации существующей методики оценки уровня знаний респондентов при помощи компьютерного тестирования.

В процессе решения задачи разработана инженерная методика автоматизированной процедуры оценки знаний респондентов, и таким образом поставленная цель достигнута.

В ходе исследований получены следующие основные научные и практические результаты:

1. Проведен анализ вида распределения процента правильных ответов респондентов, установлено, что такое распределение отличается от нормального.

2. Предложен метод оценки объективной сложности теста с использованием методов математической статистики и имитационного моделирования.

3. Предложен метод оценки объективной дифференцирующей способности теста с использованием методов математической статистики и имитационного моделирования.

4. Разработана библиотека, предназначенная для проведения и анализа тестов.

5. На основе предложенной библиотеки разработаны программные средства для создания, оценки и проведения тестов в виде компьютерного тестирования, а так же в бланковом варианте.

6. Результаты работы используются ФБГОУ ВПО Сибирский государственный аэрокосмический университет им. академика М. Ф. Решетнева.

Разработанная методика оценки теста и контрольно-измерительного материала может быть использован для широкого класса задач, в том числе возможно дальнейшее развитие методики и разработанного программного обеспечения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Java Servlet [Электронный ресурс] // Wikipedia, the free. – Электрон. дан. – Режим доступа: http://en.wikipedia.org/wiki/Java_Servlet (дата обращения: 15.05.2012). – Загл. с экрана.
2. Введение в инфраструктуру Microsoft .NET Remoting [Электронный ресурс] // GotDotNet.Ru: сообщество .NET. – Электрон. дан. – Режим доступа: <http://www.gotdotnet.ru/blogs/msdn/6362/> (дата обращения: 15.05.2012). – Загл. с экрана.
3. Ильязова М.Д. Модель выпускника вуза в рамках компетентностного подхода к целям и результатам ВПО [Текст] // Педагогические науки № 3 2006 г.
4. Мащенко Т. П. Рейтинговая система оценивания знаний учащихся в современном образовательном пространстве [Текст] // XII Всероссийский интернет-педсовет. – Электрон. дан. – Режим доступа: <http://12.pedsovet.org/> (дата обращения: 15.05.2012). – Загл. с экрана.
5. Мелецинек Адольф. Инженерная педагогика [Текст]. - М.: МАДИ(ТУ), 1998. - 185 с.
6. О классном журнале замолвите слово... [Электронный ресурс]. – Электрон. дан. – Режим доступа <http://upr.1september.ru/2005/09/4.htm> (дата обращения: 15.05.2012). – Загл. с экрана.
7. Саяпин А. В. О реализации объектной библиотеки для компьютерного тестирования как средства определения уровня знаний студента [Текст] / Международный научно-технический журнал «Информационные технологии моделирования и управления» (ISSN 1813-9744) № 2(61)2010.
8. Саяпин А.В. Использование сетевых технологий для проведения тестирования студентов [Текст] / Актуальные проблемы экономики, информатики и права: Сб. науч. статей X межвузовской научно-практической конференции «Актуальные проблемы экономики,

- информатики и права». – Красноярск: Красноярский филиал МЭСИ, 2010.
9. Саяпин А.В. Оценка объективной меры сложности компьютерного теста методами имитационного моделирования [Текст]. - Международная научно-практическая конференция «Инновационная интегрированная система профессионального образования. Проблемы и пути развития-2011» (статья) – Красноярск: СибГАУ, 2011.
10. Сравнение ASP.NET MVC [Электронный ресурс] . – Электрон. дан. – Режим доступа <http://www.aspnet.com.ua/Blog/sravnenie-Asp-Net-mvc.aspx/203> (дата обращения: 15.05.2012). – Загл. с экрана.
11. Тестовый контроль знаний учащихся [Электронный ресурс] . – Электрон. дан. – Режим доступа: http://www.interwrite.ru/teacher/lesson/knowledge_control/ (дата обращения: 15.05.2012). – Загл. с экрана.
12. Физическая энциклопедия. В 5-ти томах [Текст]. — М.: Советская энциклопедия. Главный редактор А. М. Прохоров. 1988.
13. Челышкова М.Б. Теория и практика конструирования педагогических тестов [Текст] /- М., 2001
14. Сидоренко Е.В. Методы математической обработки в психологии [Текст]. – СПб.: ООО «Речь», 2003.