

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН

ИННОВАЦИОННЫЙ ЕВРАЗИЙСКИЙ УНИВЕРСИТЕТ
МАГИСТРАТУРА

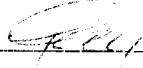
Кафедра «Автоматизированные системы обработки информации и
управления»

Магистерская диссертация

**РАЗРАБОТКА МЕТОДИКИ ТЕСТИРОВАНИЯ ЗНАНИЙ СТУДЕНТОВ
ИИЕУ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

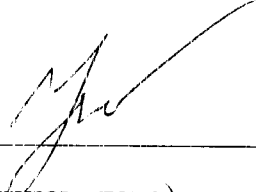
6N0607 «Вычислительная техника и программное обеспечение»

Исполнитель  25.06.18 Бирюкова А.В.
(подпись, дата)

Научный руководитель
к.т.н., профессор  Фандюшин В.И.
(подпись, дата)

Допущена к защите:

Зав. кафедрой «АСОИиУ»

к.т.н., профессор  Наумов В.В.
(подпись, дата)

Павлодар, 2008

0-4986/1

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН**
Учреждение «Инновационный Евразийский Университет»

Факультет И.В.В.
Специальность Инженерная физика и прикладная физика
Кафедра Инженерная физика и прикладная физика

**ЗАДАНИЕ
на выполнение магистерской диссертации**

Магистранту Бирюков Александр Александрович
(фамилия, имя, отчество)

Тема магистерской диссертации Исследование влияния температуры на свойства полимерных материалов

утверждена приказом по вузу № _____ от «___» _____ 20__ г.

Срок сдачи законченной диссертации «___» _____ 20__ г.

Исходные данные к диссертации _____

Перечень подлежащих разработке в магистерской диссертации вопросов или краткое содержание диссертационной работы:

- а) Исследование влияния температуры на свойства полимерных материалов
б) Исследование влияния температуры на свойства полимерных материалов
в) Исследование влияния температуры на свойства полимерных материалов

Перечень графического материала (с точным указанием обязательных чертежей)

Рекомендуемая основная литература 1. Бирюков А.А. Исследование влияния температуры на свойства полимерных материалов. Диссертация на соискание степени кандидата технических наук. М.: Издательство МГУ, 1998.
2. Бирюков А.А. Исследование влияния температуры на свойства полимерных материалов. Диссертация на соискание степени кандидата технических наук. М.: Издательство МГУ, 1998.

Аннотация

Диссертационная работа посвящена вопросам разработки новой методики тестирования знаний студентов ИнЕУ.

Данная работа состоит из 3 разделов: компьютерное тестирование как эффективная система контроля знаний, интеллектуальное тестирование и реализация системы тестирования уровня знаний студентов.

В первом разделе были рассмотрены основная цель и функции системы контроля знаний, определены важнейшие свойства и формы педагогического теста, перечислены основные достоинства и недостатки тестовых технологий, также были выделены преимущества компьютерного тестирования перед бланковыми технологиями и выработаны основные требования к компьютерному тестовому комплексу.

Раздел «Интеллектуальное тестирование» посвящен обзору методов и моделей интеллектуального тестирования. В данном разделе подробно описаны алгоритм адаптивного тестирования и методы обработки данных в тестологии.

Третий раздел содержит описание этапов реализации системы компьютерного тестирования, а именно описание структуры базы данных, выбор среды разработки, а также интерфейса приложений системы тестирования.

В заключении приведены выводы о целесообразности внедрения разработанной автоматизированной системы тестирования уровня знаний студентов ИнЕУ с применением методов искусственного интеллекта.

Түйіндеме

Диссертациялық жұмыс ИнЕУ студенттерінің білімін тестілеудің жаңа әдісін әзірлеу мәселелеріне арналған.

Бұл еңбек 3 тараудан тұрады: білімді бақылаудың тиімді жүйесі ретіндегі компьютерлік тестілеу, зияткерлік тестілеу және студенттердің білім деңгейін тестілеу жүйесін іске аыру.

Бірінші тарауда білімді бақылау жүйесінің негізгі мақсаты мен міндеттері қаралды, педагогикалық тестінің маңызды қасиеттері мен нысандары анықталды, тест технологияларының негізгі артықшылықтары мен кемшіліктері тізбектелді, бланкілік технологиялар алдындағы компьютерлік тестілеудің артықшылықтары ерекше аталып көрсетілді және компьютерлік тестлік кешенге қойылатын негізгі талаптар дайындалды. «Зияткерлік тестілеу» тарауы зияткерлік тестілеудің әдістері мен модельдеріне сипаттауға арналған.

Осы тарауда бейімдеп тестілеу алгоритмі мен тестологияда деректерді өңдеу әдісі егжей-тегжейлі сипатталды.

Үшінші тарауда компьютерлік тестілеу жүйесін іске асыру кезеңдері, атап айтқанда дерекқор құрылымының сипаттамасы, әзірлеу ортасын таңдау, тестілеу жүйесі қосымшаларының интерфейсі берілген.

Қорытындыда жасанды интеллект әдісін қолдану арқылы ИнЕУ студенттерінің білім деңгейін тестілеудің әзірленген автоматтандырылған жүйесін енгізудің орынды екендігі туралы тұжырым жасалған.

The summary

The master's thesis is devoted to the development of new methodology for testing of InEU student learning.

This work consists of 3 sections: computer testing as an effective monitoring system of knowledge, intellectual testing and realisation of system testing of student learning.

In the first section have been considered a main purpose and functions of the control system of knowledge, the major properties and forms of the pedagogical test are defined, listed the main advantages and disadvantages of testing technologies, advantages of computer testing before blank technologies also have been allocated and the basic requirements to a computer test complex are developed.

Section «Intellectual testing» is devoted to reviewing the methods and models of intellectual testing. In the given section the algorithm of adaptive testing and data processing methods in test's science are in detail described.

The third section describes the stages of development of testing system of knowledge, namely the description of the database structure, choice of development environment, as well as interface test system applications.

In conclusion outputs about expediency introduction of the developed automated testing system of knowledge's levels of InEU students with application of methods of an artificial intelligence are resulted.

РЕФЕРАТ

Диссертация состоит из введения, трех глав, заключения и списка использованных источников (20 наименований). Основное содержание диссертационной работы изложено на 87 страницах машинописного текста, иллюстрированного таблицами и рисунками.

В настоящее время происходит компьютеризация всех сторон функционирования человека и общества, в связи с этим необходимо переосмысление педагогических основ. Поэтому анализ вопросов использования современных информационных технологий для оценки знаний студентов ВУЗов является актуальным направлением педагогических исследований.

Объектом исследования является тестовые системы контроля знаний студентов ВУЗов.

Предметом исследования является компьютерное тестирование знаний с применением методов искусственного интеллекта, его преимущества и недостатки.

Целью диссертационной работы является исследование существующих методов контроля знаний учащихся и разработка методики тестирования знаний студентов ИнЕУ с применением методов искусственного интеллекта.

В связи с намеченной целью поставлены следующие основные задачи:

- анализ контроля знаний в системе обучения ВУЗа;
- выявление недостатков существующих методик контроля знаний студентов;
- выработка требований к специализированному программному обеспечению, ориентированному на выявление уровня знаний учащихся;
- изучение методов и моделей интеллектуального тестирования;
- исследование теорий обработки результатов тестирования;
- реализация программного обеспечения компонентов системы тестирования знаний студентов.

Методы исследования основаны на использовании положений теории оценивания латентных качеств личности и параметров тестовых заданий – IRT теории, теории вероятности и методов искусственного интеллекта. В разработке программного обеспечения использовалась технология объектно-ориентированного программирования.

В диссертационной работе предложен новый подход к организации контроля знаний студентов, учитывающий особенности оценивания знаний учащихся ИнЕУ.

Программное обеспечение, разработанное в процессе компьютерной реализации тестирования знаний с применением методов искусственного интеллекта могут быть использованы в качестве системы оценивания приобретенных знаний в процессе обучения студентов ИнЕу.

СОДЕРЖАНИЕ

Введение	7
1. Компьютерное тестирование как эффективная система контроля знаний	8
1.1 Контроль знаний в системе обучения	8
1.2 Педагогическое тестирование	11
1.2.1 Валидность и надежность – важнейшие свойства педагогического теста	15
1.2.2 Дополнительные свойства педагогического теста	19
1.2.3 Четыре формы тестовых заданий	25
1.2.4 Общие достоинства и недостатки тестовых технологий	30
1.3 Компьютерное тестирование	32
1.3.1 Преимущества компьютерного тестирования	33
1.3.2 Требования к компьютерному тестовому комплексу	36
2 Интеллектуальное тестирование	38
2.1 Понятие «Искусственный интеллект»	38
2.2 Методы и модели интеллектуального тестирования	38
2.2.1 Модели распознавания образа уровня знаний	39
2.2.2 Предметно - критериальная методика составления тестов	40
2.2.3 Метод определения количества образовательной информации	42
2.2.4 Информационно-генетические алгоритмы	44
2.2.5 Модель Раша	45
2.2.6 Абсолютная временная шкала измерения знаний	47
2.2.7 Методика статистического анализа качества обучения	48
2.2.8 Технология рейтинговых исследований качества образования с применением нейронных сетей	49
2.2.9 Модель адаптивного тестового контроля	51
2.3 Адаптивное тестирование	51
2.4 Обработка данных в тестологии	55
2.4.1 Классическая теория тестирования	55
2.4.2 IRT - теория тестирования	59
2.4.3 Сравнительный анализ IRT и классической теории тестирования	64
3 Реализация системы тестирования уровня знаний студентов	67
3.1 Модель системы тестирования уровня знаний студентов	67
3.2 Структура базы данных разрабатываемого тестового комплекса	68
3.3 Выбор среды разработки автоматизированной системы контроля знаний	69
3.4 Язык SQL как стандартный язык баз данных	70
3.5 Программная реализация системы тестирования знаний студентов	71
3.5.1 Приложение «Организация и обработка результатов тестирования»	72
3.5.2 Приложение «Базовое тестирование»	80
3.5.3 Приложение «Адаптивное тестирование»	82
Заключение	84
Список использованных источников	86

ВВЕДЕНИЕ

В настоящее время, с целью формирования более обширного объема знаний у студентов и устранения субъективизма в их оценке, широко применяются новые прогрессивные формы электронных методов обучения и контроля эффективности подготовки. Во многом это объясняется явной необходимостью и потребностью сегодняшнего дня ВУЗов, так как применяемый большинством кафедр традиционный устный или письменный контроль знаний студентов, не отвечает современным требованиям учебного процесса. Преподаватель при этом имеет весьма ограниченные возможности оперативно провести унифицированный текущий контроль уровня знаний с максимальным вовлечением в этот процесс всех студентов.

В работе студента с контролирующей компьютерной программой, помимо неоспоримых её достоинств, примешивается весомый фактор новизны, повышенного интереса и внимания. Использование контролирующей компьютерной программы позволяет осуществить стопроцентный опрос всех студентов в процессе вводного контроля знаний на практическом занятии, при отработке пропущенных занятий, приёме зачётов и экзаменов. Внедрение компьютерного контроля в учебный процесс дисциплинирует студентов, учит навыкам выделения главного и улучшает процесс усвоения учебного материала. Компьютерные программы тестирования знаний позволяют использовать их и для проверки знаний и для самоконтроля студентов при изучении учебных дисциплин.

1 КОМПЬЮТЕРНОЕ ТЕСТИРОВАНИЕ КАК ЭФФЕКТИВНАЯ СИСТЕМА КОНТРОЛЯ ЗНАНИЙ

Система контроля знаний является составной частью процесса обучения.

По определению контроль это соотношение достигнутых результатов с запланированными целями обучения. От его правильной организации во многом зависят эффективность управления учебно-воспитательным процессом и качество подготовки специалиста. Проверка знаний учащихся должна давать сведения о правильности или неправильности конечного результата выполненной деятельности.

Правильно поставленный контроль учебной деятельности учащихся позволяет преподавателю оценивать получаемые ими знания, умения и навыки, вовремя оказать необходимую помощь и добиться поставленных целей обучения. Все это в совокупности создает благоприятные условия для развития познавательных способностей учащихся и активизации их самостоятельной работы на занятиях.

Хорошо поставленный контроль позволяет преподавателю не только правильно оценить уровень усвоения учащимися изучаемого материала, но и увидеть свои собственные удаchi и премахи.

1.1 Контроль знаний в системе обучения

Основная цель контроля знаний и умений состоит в обнаружении достижений, успехов учащихся ВУЗов, в указании путей совершенствования, углубления знаний, умений, с тем, чтобы создавались условия для последующего включения студентов в активную творческую деятельность. Эта цель в первую очередь связана с определением качества усвоения учащимися учебного материала - уровня овладения знаниями, умениями и навыками предусмотренных программой [1].

Во - вторых, конкретизация основной цели контроля связана с обучением приемам взаимоконтроля и самоконтроля, формированием потребности в самоконтроле и взаимоконтроле.

В - третьих эта цель предполагает воспитание у учащихся таких качеств личности, как ответственность за выполненную работу, проявление инициативы.

Если перечисленные цели контроля знаний и умений учащихся ВУЗов реализовать, то можно говорить о том, что контроль выполняет следующие функции:

- контролирующую,
- обучающую (образовательную),
- диагностическую, прогностическую,
- развивающую,
- ориентирующую,

- воспитывающую

Рассмотрим эти функции более подробно.

Контролирующая функция состоит в выявлении состояния знаний и умений учащихся, уровня их умственного развития, в изучении степени усвоения приемов познавательной деятельности, навыков рационального учебного труда.

При помощи контроля определяется исходный уровень для дальнейшего овладения знаниями, умениями и навыками, изучается глубина и объем их усвоения. Сравнивается планируемое с действительными результатами, устанавливается эффективность используемых преподавателем методов, форм и средств обучения.

Обучающая функция контроля заключается в совершенствовании знаний и умений, их систематизации. В процессе проверки учащиеся повторяют и закрепляют изученный материал. Они не только воспроизводят ранее изученное, но и применяют знания и умения в новой ситуации. Проверка помогает выделить главное, основное в изучаемом материале, сделать проверяемые знания и умения более ясными и точными. Контроль способствует также обобщению и систематизации знаний.

Сущность диагностической функции контроля - в получении информации об ошибках, недочетах и пробелах в знаниях и умениях студентов и порождающих их причинах затруднений учащихся в овладении учебным материалом, о числе, характере ошибок. Результаты диагностических проверок помогают выбрать наиболее интенсивную методику обучения, а также уточнить направление дальнейшего совершенствования содержания методов и средств обучения.

Прогностическая функция проверки служит получению опережающей информации об учебно-воспитательном процессе. В результате проверки получают основания для прогноза о ходе определенного отрезка учебного процесса: достаточно ли сформированы конкретные знания, умения и навыки для усвоения последующей порции учебного материала (раздела, темы).

Результаты прогноза используют для создания модели дальнейшего поведения учащегося, допускающего сегодня ошибки данного типа или имеющего определенные пробелы в системе приемов познавательной деятельности.

Прогноз помогает получить верные выводы для дальнейшего планирования и осуществления учебного процесса [2].

Развивающая функция контроля состоит в стимулировании познавательной активности студентов, в развитии их творческих способностей. Контроль обладает исключительными возможностями в развитии учащихся. В процессе контроля развиваются речь, память, внимание, воображение, воля и мышление студентов. Контроль оказывает большое влияние на развитие и

проявление таких качеств личности, как способности, склонности, интересы, потребности.

Сущность ориентирующей функции контроля - в получении информации о степени достижения цели обучения отдельным студентом и группой в целом - насколько усвоен и как глубоко изучен учебный материал. Контроль ориентирует учащихся в их затруднениях и достижениях.

Вскрывая пробелы, ошибки и недочеты учащихся, он указывает им направления приложения сил по совершенствованию знаний и умений. Контроль помогает учащемуся лучше узнать самого себя, оценить свои знания и возможности.

Воспитывающая функция контроля состоит в воспитании у учащихся ответственного отношения к учению, дисциплины, аккуратности, честности. Проверка побуждает более серьезно и регулярно контролировать себя при выполнении заданий. Она является условием воспитания твердой воли, настойчивости, привычки к регулярному труду.

Интегративная функция контроля представляет собой способствование выработке представлений о науке как едином комплексе знаний и методов [2].

Сущность стимулирующей функции заключается в поощрении творческой деятельности, побуждении обучаемых и педагогов к самообразованию и поиску новых форм и путей обучения.

Выделение функции контроля подчеркивает его роль и значение в процессе обучения. В учебном процессе сами функции проявляются в разной степени и различных сочетаниях. Реализация выделенных функций на практике делает контроль более эффективным, а также эффективней становится и сам учебный процесс.

Выделим также основные принципы контроля знаний:

- целенаправленность;
- объективность;
- всесторонность;
- регулярность;
- индивидуальность.

Раскроем эти принципы контроля подробнее.

Целенаправленность предполагает четкое определение цели каждой проверки. Постановка цели определяет всю дальнейшую работу по обоснованию используемых форм, видов, методов и средств контроля. Цели контроля предполагают ответы на следующие вопросы: что должно проверяться, кто должен опрашиваться, какие выводы можно будет сделать на основе результатов проверки, какой ожидается эффект от проведения проверки. При конкретизации целей контроля исходят из целей воспитания, развития и обучения учащихся, которые реализуются на данном этапе обучения.

Объективность контроля предупреждает случаи субъективных и ошибочных суждений, которые искажают действительную успеваемость учащихся и снижают воспитательное значение контроля. Объективность контроля зависит от многих факторов. Среди них выделяют следующие: четкое выделение общих и конкретных целей обучения, обоснованность выделения и отбора и содержания контроля, обеспеченность методами обработки, анализа и оценивания результатов контроля, организованность проведения контроля. От решения этих вопросов во многом зависит объективность и качество контроля.

Под всесторонностью контроля понимают охват большого по содержанию проверяемого материала. Этот принцип включает в себя усвоение основных идей данного курса, и усвоение учебного материала по определенным содержательным, стержневым линиям курса, и знание учащимися отдельных и существенных, фактов, понятий, закономерностей, способов действий и способов деятельности. При таком обилии проверяемого материала усложняется методика составления заданий, т.е. предъявляются повышенные требования к методике выделения и сбора объектов проверки.

Под регулярностью подразумевается систематический контроль, который сочетается с самим учебным процессом.

Индивидуальность контроля требует оценки знаний, умений, навыков каждого учащегося.

Итак, контроль знаний является неотъемлемой частью учебного процесса.

1.2 Педагогическое тестирование

Одной из форм массового контроля знаний студентов, который осуществляет преподаватель после изучения ими всей программы учебной дисциплины, является тестирование. Экзамен в форме тестирования обладает целым рядом преимуществ перед традиционной формой экзамена-диалога «преподаватель-студент».

Перед преподавателем, составляющим тест, всегда стоит трудноразрешимая задача: создать минимальное количество тестовых заданий, позволяющих получить максимально достоверную и обширную информацию о знаниях учащегося. Составление тестовых заданий – это путь проб и ошибок. Каждое задание необходимо проверить на больших группах испытуемых и только потом делается заключение о включении или не включении его в тест. Поскольку возможности эксперимента всегда ограничены, нет гарантии, что задание одинаково надежно будет играть свою роль всегда.

Ранние попытки автоматизировать систему контроля знаний во многих странах в начале XX века привели к созданию бланковых технологий тестирования и «индустрии тестов», а затем – к появлению международно-признанных организаций, проводящих ежегодно десятки миллионов сеансов бланковых тестирований. [1]

Несовершенства бланковых технологий тестирования стали очевидны после появления достаточно качественных компьютерных дисплеев (мониторов) в 80-х годах прошлого века.

К техническим несовершенствам бланковых технологий можно отнести:

- длительность тиражирования;
- высокая вероятность подлога;
- сложность сохранения конфиденциальности;
- высокая вероятность угадывания верных вариантов путем простого логического сравнения;
- трудоемкость тиражирования, хранения и транспортировки бланков и соответствующих вопросников;
- сложность обеспечения должного уровня секретности экзаменационных материалов и относительная простота подлога;
- сложность обеспечения верного восприятия правил претендентами и их выполнения;
- трудоемкость извлечения и обработки результатов;
- дороговизна сеанса и др.

К методическим несовершенствам бланковых технологий мы здесь отнесем еще одно, на наш взгляд, самое существенное. Суть этого несовершенства легко понять, если внимательно сравнить две популярные телевизионные игры — «Кто хочет стать миллионером?» и «Своя игра». В первом случае один играющий указывает верный ответ на вопрос в предъявленном списке. Во втором — игроки соревнуются друг с другом в скорости ответа на вопрос, зная лишь его тему и используя лишь свою память [1].

С методической точки зрения, эти два сценария кардинально различны. В первом случае указать правильный ответ несравненно легче:

во-первых, видя четыре варианта ответа, среди которых всего один верный, можно осуществить выбор просто логическим путем, не имея специального знания;

во-вторых, даже если нет возможности вспомнить верный ответ или догадаться путем логического перебора вариантов, можно выбрать любой ответ и с вероятностью 25% добиться успеха.

Во втором же случае, т. е. в «Своей игре», ни логика, ни удача не помогут победить — помогут только память и знание.

Применяемые тестовые задания называют:

- в первом случае — заданиями закрытой формы;
- во втором случае — заданиями открытой формы.

Надо сказать, что, согласно математической теории вероятностей, приведенную выше величину — 25% — можно уменьшить, т. е. уменьшить вероятность простой удачи. Например, четырьмя способами:

- увеличением количества вариантов выполнения заданий;
- помещением среди демонстрируемых вариантов не одного, а нескольких верных ответов, которые требуется указать все;
- созданием дистракторов.

Однако, даже при применении этих способов, при тестировании:

- сохраняется, хотя и уменьшается, степень возможности достижения успеха с помощью простой логики, поскольку испытуемый видит одновременно все варианты и имеет возможность их сравнить;
- в связи с наличием в результатах тестирований определенного высокого процента успеха, списываемого «на удачу», много труда составителей тестовых заданий оказывается напрасным (иногда даже свыше трети);
- значительно «раздувается» объем бумажных бланков, и, соответственно, увеличивается вероятность сбоев при проверке;
- увеличивается непроизводительное время знакомства испытуемых с заданиями, соответственно должно быть уменьшено количество заданий в связи с объективно накапливающейся усталостью и ограниченным временем сеанса.

В технологиях автоматизированного бланкового тестирования обычно заложен именно вариант выбора из списка (задания закрытой формы) с его главным недостатком – возможностью увидеть одновременно все правильные и неправильные варианты ответа и воспользоваться этим, проявив не знания, а умение логически мыслить.

На взгляд организаторов таких тестирований, все отмеченные выше недостатки тестов с заданиями закрытой формы окупаются простотой обработки результатов.

Для «обмана» человеческой логики создатели бланковых тестовых материалов придумали множество приемов. Эти приемы позволяют конструировать варианты ответов на вопросы так, что неверный ответ кажется правдоподобным, а верный – неправильным. Совокупность этих приемов иногда называют теорией дистракторов.

Процедура массового бланкового тестирования слабо защищена от подлога. Одно дело, когда педагог проводит тестирование для облегчения собственного труда и сам наблюдает за выполнением процедуры, и совсем другое – когда тестирование ведется большим коллективом людей, которые не заинтересованы в качестве результатов. Или, еще хуже, которые заинтересованы лишь в успехе «выделенных» испытуемых.

Возможных путей подлога в бланковом тестировании довольно много:

- «выделенного» учащегося можно вместе с заданием поместить в отдельный кабинет, где он все правильно выполнит под присмотром, например, репетитора;

- задание может выполнить репетитор без участия самого испытуемого и потом подложить бланк в общую пачку;
- в процессе своего изготовления бланки проходят через сотни рук и глаз, поэтому тестовые ключи «уплывают» на рынок.

При желании и опыте можно придумать еще больше вариантов теневых путей к успеху в бланковом тестировании.

В педагогическом тестировании можно выделить классификацию оцениваемых параметров претендентов:

1. Кругозор:

- знание основных фактов;
- знание истории науки;
- методологическая подготовка (классификации, внутридисциплинарные и междисциплинарные связи);
- умение применить философские методы, понимание противоречий, различение формы и содержания;
- знакомство с основной литературой;
- словарный запас.

2. Знания основ дисциплины (темы):

- знание законов;
- знание определений и ключевых понятий, номенклатуры;
- знание правил;
- знание основных концепций.

3. Умение решать стандартные задачи в рамках дисциплины (темы):

- умение применять знания в решении простых задач;
- умение защищать свои взгляды;
- умение логичного изложения;
- умение приложить теорию к практике, описать и объяснить простое явление;
- наличие стандартных умений и навыков (построение таблиц и графиков, формул, проведение экспериментов, анализ текстов и изображений);
- умение перевести стандартный текст на иной язык.

4. Умение разрешать проблемы:

- понимание взаимосвязанности тем и разделов курса, умение применить взаимосвязи;
- знакомство с периодической и не основной литературой;
- наблюдательность, умение вычленить главное;
- умение распознать структуру текста (проблемы, изображения, формулы, явления);

- умение применить межпредметные знания, «сквозные» учебные умения и навыки.

1.2.1 Валидность и надежность – важнейшие свойства педагогического теста

Педагогический тест – система заданий такая, что результат их выполнения группой претендентов позволяет достаточно надежно ранжировать их (присвоить им порядковые номера) по качеству обученности, количеству имеющихся знаний. Педагогическим тестам, а также результатам их применения (результатам тестирования) приписывают характеристики и свойства, обуславливающие качество реализации основной цели таких тестов – оценивания уровня знаний студентов. Ниже перечислим основные свойства педагогического теста.

Валидность теста. Результаты тестирования группы претендентов должны соответствовать объективным характеристикам, данным претендентам их руководителями, коллегами, преподавателями. Это важнейший принцип, которому должны быть подчинены усилия составителя теста.

Валидность отражает пригодность теста для измерения того, что он по замыслу должен измерять. Измерение валидности теста может включать ряд процедур, главной из которых является валидация.

Валидация производится путем математического сравнения результатов тестирования с успешностью выполнения соответствующей практической деятельности испытуемыми. При этом учитывается шкалированное мнение руководителей группы прошедших тестирование испытуемых об их конкретных исполнительских качествах.

Валидность считается достаточно высокой, если коэффициент корреляции будет более 0,6. При значении коэффициента корреляции 0,45 – 0,65 валидность считается вполне удовлетворительной [3].

Считается, что стопроцентно валидных тестов нет. Различают следующие виды валидности:

- содержательную – степень соответствия теста программам обучения и образовательным стандартам;
- критериальную – степень соответствия результатов тестирования внешнему, не относящемуся к тесту критерию;
- прогностическую – степень полноты достижения цели тестирования и др.

Наибольший «вес» здесь приходится на содержательную валидность.

Надежность теста и технологии тестирования. Результаты тестирований подобных групп претендентов с помощью одного теста должны быть одинаковыми (подобными) и не зависеть от времени. Это второй по важности принцип, которому должны быть подчинены усилия составителя теста. Чаще всего, надежность можно охарактеризовать лишь умозрительно, поскольку:

- повторное тестирование одной группы претендентов по одному тесту приводит к росту результатов за счет эффектов памяти, самообучаемости и пр.;
- очень трудно найти подобные группы претендентов и доказать, что это так;
- на результаты оказывает влияние предыстория групп претендентов, технические сложности исключения предварительного общения между ними и пр.

Надежность – принцип, пришедший в педагогику из техники. Это принцип, предъявляемый к инструменту технического измерения. Но педагогика – иная область с иными закономерностями, и поэтому принцип надежности часто обоснованно подвергается критике.

Надежность теста определяется как устойчивость результатов при повторном тестировании на той же (такой же) выборке испытуемых. Считается, что при коэффициенте корреляции результатов повторного тестирования более 0,75 уровень надежности теста приемлем.

Надёжность теста проверяется относительно:

- временных изменений;
- выбора конкретных заданий;
- аспектов процедуры тестирования.

Вычисление ошибки измерений – вероятных пределов колебаний измеряемой величины – основано на понятии надёжности.

Ни один тест не является абсолютно надёжным. Дисперсия ошибки (мера отклонения результатов сеансов тестирования) отражает случайные колебания, вызываемые неконтролируемыми факторами:

- тренинг и дообучение претендентов в течение периода измерения надежности;
- случайными отвлекающими моментами;
- обучаемостью претендентов в течение сеанса и в перерывах между сеансами;
- мотивированностью претендентов и ее изменениями;
- адаптацией, привыканием претендентов к форме проведения тестирования;
- изменениями состояния претендентов.

Первый из этих факторов, наиболее влияющий на результаты и «путаящий все карты», – обычное явление, создаваемое родителями, руководством учебных заведений, добросовестными педагогами, не желающими, чтобы их воспитанники предстали перед экспериментаторами в худшем виде [1].

Стандартный набор данных о тесте, предназначенном для широкого употребления, обязательно должен включать сведения о мере его надёжности.

Надёжность обычно вычисляется с помощью коэффициента корреляции произведения моментов К. Пирсона (его можно найти в статистических справочниках).

При измерении надёжности следует обращать внимание на критерий целесообразности, помнить об отношении «затраты — польза». Следует контролировать, в каком интервале времени измерялась надёжность, произошли ли за этот период какие-либо события, способные повлиять на результаты.

Существуют различные типы надёжности и подходы к их вычислению. Педагогу, апробирующему тест на своих группах, важнее всего владеть техникой измерения:

- ретестовой надёжности;
- надёжности эквивалентных форм.

Надёжность не обязательно предполагает высокую валидность. Среди тестологов распространено поверье, что на практике всегда выполняется следующее выражение: валидность $\leq$ надёжность. Иными словами, значение валидности теста не может превышать значение его надёжности, какие бы процедуры их определения не использовались.

Надёжность и валидность тестов могут быть значительно повышены, если их качественные и содержательные характеристики будут связаны со статистическими данными, полученными при обработке больших массивов результатов тестирования испытуемых.

Разработано несколько способов определения и повышения надёжности теста.

Метод повторного тестирования (метод ретеста). Двукратное или многократное использование одного и того же теста в одной группе испытуемых. Достоинство метода заключается в простоте его использования, ясности основных посылок, простоте сравнений и расчетов. К недостаткам относят неопределенность выбора временного интервала между опросами. Этот интервал может колебаться от нескольких минут до нескольких дней, месяцев и даже лет. Естественно, что при этом по-разному проявляются факторы: запоминание или, наоборот, забывание, влияние опыта, полученного в первом опросе на второй, влияние общения испытуемых между собой после первого опроса.

Метод параллельного тестирования (метод эквивалентных форм). Одной и той же группе испытуемых дается вначале одна форма теста, и после перерыва — другая. Затем вычисляется величина коэффициента корреляции верности выполнения заданий, которая и принимается за значение коэффициента надёжности. Если между предъявлением обеих форм имеется

значительный временной интервал, то коэффициент надежности называют по-другому: коэффициентом эквивалентности или коэффициентом стабильности.

Метод раздельного коррелирования. В основе лежит допущение о параллельности не только отдельных форм, но и частей внутри одной формы теста. Для получения величины коэффициента надежности сравниваются результаты выполнения частей теста. В зависимости от способа деления теста могут меняться значения коэффициента. Часто применяемая процедура деления теста – это сведение в одну часть результатов респондентов в нечетных номерах высказываний и в другую – четных. Суммирование баллов в этих половинах теста дает два вектора, коэффициент корреляции между которыми и служит коэффициентом надежности теста. Его называют еще коэффициентом внутренней состоятельности теста. Этот метод имеет преимущество перед другими, поскольку позволяет оценить надежность при однократном тестировании.

Чаще всего рассматривают три типа надежности.

Ретестовая надежность. Измеряется при повторном проведении того же самого теста на том же контингенте и в тех же условиях. Ее аналогом является проверочная контрольная работа, проводимая, когда результаты первой контрольной кажутся сомнительными. Ретестовая надёжность обычно подсчитывается с использованием коэффициента корреляции моментов Пирсона. Чтобы повысить ретестовую надёжность, надо отбирать из первоначального, апробируемого варианта теста те задания, на которые испытуемые дают устойчивые ответы. Чем выше ретестовая надёжность, тем менее чувствительны результаты к влиянию неконтролируемых факторов.

Надежность эквивалентных форм. Измерение надёжности эквивалентных форм проще измерения ретестовой надёжности. Это вычисление корреляции результатов выполнения двух форм одного и того же теста, считающимися равноценными. Их называют также параллельными, взаимозаменяемыми, сопоставимыми, подобными. Здесь коэффициент надёжности одновременно отражает временную стабильность теста и согласованность результатов сеансов по двум формам. Только если варианты сеансов следуют один за другим, можно точно измерить надёжность эквивалентных форм теста. Для снижения влияния эффекта самообучения претендентов, меняют способ предъявления эквивалентных форм сходных заданий.

Надежность теста на скорость. Для проверки надёжности тестов на скорость считаются наиболее эффективными метод ретеста и метод эквивалентных форм. Есть приёмы разделения полного времени выполнения теста на четыре части с регистрацией результатов отдельно для каждой четверти.

1.2.2 Дополнительные свойства педагогического теста

Научная достоверность. В тест включаются только те элементы знания и связи между ними, которые являются объективно истинными. Соответственно, спорные точки зрения, вполне нормальные в науке, не включают в тест.

Обобщенность. Тема тестирования не должны быть излишне обобщенной.

Значимость. Значимое тестовое задание отражает структурный информационный элемент дисциплины либо связь между структурными элементами, без которых знания становятся неполными, с пробелами.

Соответствие современному знанию. В тестовое задание включаются только современные элементы знания. Исключением могут быть тестовые задания, составленные специально для ведения учебных занятий процесса в рамках концепции «диалога культур».

Соответствие источникам знания. Все тестовые задания должны включать адрес источника, который может быть включен и в демонстрируемую часть текста тестового задания в случаях, когда имеются разночтения в рекомендованной литературе.

Репрезентативность. Репрезентативный тест содержит совокупность тестовых заданий, отражающую все структурные элементы содержания дисциплины и их связи. Репрезентативная (представительная) база тестовых заданий содержит совокупность тестовых заданий, достаточную для ведения вариативного контроля по любой из имеющихся в структуре дисциплины учебных тем.

Вариативность. Содержание теста может и должно варьироваться по мере:

- изменения содержания дисциплины;
- изменения и появления новых научных знаний;
- изменения цели тестирования;
- изменения направленности на особенные выборки испытуемых.

Для достижения достаточной вариативности теста в рамках учебной дисциплины необходимо, чтобы база тестовых заданий содержала тестовые задания соответствующие разным учебным пособиям и разным способам изложения учебного материала.

Сложность. Для каждого тестового задания в процессе контроля определяют усредненную, выраженную числом, сложность. Сложность задания № j определяют двумя способами. Например, так:

$$q_i = 1 - p_i, \quad (1)$$

где p_i – относительное количество претендентов, выполнивших данное задание.

Либо, как «логит»:

$$\beta = \ln\left(\frac{q_i}{p_i}\right) \quad (2)$$

Претендент, правильно выполняющий сложные задания, с большой вероятностью правильно выполнит и легкие. Сложность теста соответствует содержанию обучения и поддерживает высокий уровень мотивации претендентов. Не следует пренебрегать заданиями, которые выполняются большим количеством претендентов (от 90 до 100%). Иногда может показаться, что задание лишнее, поскольку слишком лёгкое. Надо проконтролировать выборку – вошли ли в нее слабо подготовленные претенденты. Они могут сознательно не допускаться к тестированиям руководством учреждения под благовидным предлогом. Аналогичная, но обратная ситуация может случиться и с наиболее сложными заданиями. Прежде, чем отвергать задания, которые никто не выполнил, необходимо проверять, проходили ли тестирование наиболее подготовленные учащиеся. Они могли в это время отсутствовать.

Трудоемкость. Для каждого тестового задания в процессе контроля определяют усредненную, выраженную числом, трудоемкость. Она характеризует количество элементарных операций (и мыслительных, и физических, и расчетных), которые необходимо выполнить в процессе работы над заданием. Отношение трудоемкостей заданий можно определить через усредненное отношение времен их выполнения. Относительное количество труда, который тестируемые затратят (в среднем) на выполнение каждого тестового задания, должно быть пропорционально относительной значимости отраженного в задании элемента курса.

Взвешенность. Для придания результату тестирования большей объективности, все тестовые задания после сеанса «взвешивают». Затем общий результат претендента вычисляют как сумму весов выполненных заданий. Иногда за вес тестового задания принимают его относительную сложность, т. е. долю претендентов, не сумевших его выполнить. Иногда вес приписывают заданиям волонтаристски, опираясь на преподавательский опыт. Оба подхода к вычислению весов не лишены недостатков. Например, при первом подходе возможен вариант, когда наивысший вес приобретет задание, которое показалось всем претендентам слишком легким, недостойным сиюминутного внимания, которое они, в погоне за рейтинговыми очками, оставили «на потом» и не успели к нему вернуться.

Системность. Задания в базе тестовых заданий связываются между собой общей структурой знаний. Такая связь может быть определена умозрительно либо методами факторного анализа.

Комплексность и сбалансированность. Следует гармонично сочетать в тесте задания на проверку знаний теоретического материала (понятия, законы, закономерности, гипотезы, факты, структурные компоненты теории), методов научной и практической деятельности, умений решать типовые задания. Отношение количеств тестовых заданий перечисленных типов должно соответствовать отношениям значимостей и объемов рекомендованной учебной информации.

Соответствие содержания и формы. Педагогический тест характеризуют как результат единения содержания заданий с наиболее подходящей формой.

Разрешающая способность теста. Сложность тестовых заданий не может быть только легкой (все выполнили), средней (половина группы выполнила) или трудной (никто не выполнил):

- выраженные численно сложности совокупности тестовых заданий в тесте должны равномерно заполнять тот интервал, который соответствует обученности претендентов;
- трудоемкость тестовых заданий должна быть такой, чтобы во временные рамки теста могли уложиться все претенденты, способные их выполнить.

От сбалансированности тестовых заданий по сложности и трудоемкости зависит способность теста дифференцировать претендентов в соответствии с их свойствами. Эта способность теста, выраженная числом, называется разрешающей способностью.

Дидактическая направленность теста и технологии тестирования. Одних претендентов трудные тестовые задания могут подтолкнуть к учебе, других – оттолкнуть от нее. Скучные, однообразные, излишне формальные совокупности тестовых заданий вредоносны в учении. Тест, в своей совокупности, должен включать элементы:

- привлекательности (юмористическая окраска условий и пр.);
- сюрпризности (разные и неожиданные словарные обороты, редко встречаемые в учебнике, обращение в условии заданий непосредственно к исполнителю с использованием данных регистрации и пр.);
- практической направленности (примеры из жизни, возможность применить результаты выполненного задания непосредственно в деятельности и пр.).

Сюрпризные задания включаются в тест для противодействия возникновению в процессе работы претендентов эффекта «ровной дороги». Трудоемкость и сложность тестовых заданий, в совокупности, должна быть такой, что хотя бы пару-тройку заданий должны правильно выполнить абсолютно все претенденты. Это заставит слабых претендентов поверить в свои

силы, увлечет, инициирует соревнование, даст основания для похвалы. Тест (технология тестирования), в соответствии с принципами дидактики, должен (должна) не только дифференцировать и измерять свойства претендентов, но и обладать свойством инициирования их самообучения и проявлять их стремление к повышению качества знаний, умений, навыков.

Оптимальность теста. В тест, как правило, включаются только те тестовые задания, содержание которых не дублируется и равномерно покрывает, в своей совокупности, учебный материал. Часть тестовых заданий исключают из теста, когда известно, что:

- претенденты не знают соответствующей части материала или, наоборот, знают ее гарантированно хорошо;
- претенденты не имеют навыков работы и поэтому не справятся с трудоемкими заданиями.

Тест считают более оптимальным, если с его помощью можно измерить такие же характеристики и провести дифференцирование претендентов за меньшее время.

Темперированность сложности. Последовательность заданий теста такова, что каждое последующее тестовое задание, по мнению составителя, труднее предыдущего. Этот принцип, применяющийся в бланковом тестировании, нередко подвергается критике, поскольку действует лишь «в среднем». Если претенденты учились у разных преподавателей и по разным учебникам, предсказать ряд возрастающей трудности для большинства из них – большая проблема. А «усредненный» ряд не приносит желаемого эффекта повышения разрешающей способности. Поэтому в компьютерном тестировании чаще применяют иной принцип: претендентам позволяют самостоятельно выбирать порядок выполнения заданий [1].

Темперированность трудоемкости. Последовательность заданий теста такова, что каждое последующее тестовое задание, по мнению составителя, более трудоемко, чем предыдущее. Применяется, когда оптимизируют тесты для измерения работоспособности, внимательности и подобных характеристик претендентов.

Свобода выбора. Принцип составления теста, пришедший на смену принципу «возрастающей сложности». Принцип свободы выбора в технологиях компьютерного тестирования проявляется в том, что претендентов не ограничивают в выборе последовательности выполнения заданий. Практика показывает, что при этом претенденты сами ищут и интуитивно находят для себя индивидуальный ряд тестовых заданий возрастающей сложности и трудоемкости.

Вариативность теста. Каждое обращение к базе тестовых заданий приводит к генерации нового варианта теста в соответствии с алгоритмом, ключ которого может формироваться в соответствии с регистрационными

данными, датой и временем сеанса. Алгоритм, ориентирующийся при составлении варианта теста на личность претендента, может исключать возможность повторения тестовых заданий при повторном тестировании.

Адаптивность теста. Содержание теста адаптируют для придания тесту большей разрешающей способности в известной группе претендентов. Для адаптации теста к свойствам группы претендентов и каждого претендента в отдельности, необходимо наличие в базе тестовых заданий такого набора классифицированных заданий, что возможно построение вариантов теста нацеленных на измерение конкретных свойств:

- тест дробят на субтесты, а сеанс – на ступени;
- перед построением субтеста очередной ступени проводится анализ результатов выполнения предыдущих субтестов;
- при анализе конкретизируются результаты произведенных измерений свойств претендентов, затем, если объективная дифференциация группы претендентов еще не возможна, составляется очередной субтест;
- действия повторяются на очередной ступени сеанса до тех пор, пока не будет достигнута требуемая степень дифференциации.

Полнота. В банке тестовых заданиях должны быть отражены абсолютно все структурные элементы содержания дисциплины и связи между ними.

Целесообразность. Содержание теста зависит от цели тестирования. Например, возможны ситуации:

- если нужно отобрать самых подготовленных учащихся на олимпиаду, то задания должны быть сложные, ибо только с помощью таких заданий можно отобрать лучших;
- если претенденты учились по разным учебным пособиям, то формулировки отобранных для теста заданий должны соответствовать каждому.

Верифицируемость результата. Тест может содержать равномерно распределенные среди всех заданий такие ТЗ, которые характерны разной формой и равным содержанием. Установив корреляцию результатов их выполнения можно установить, как претендент выполнял задания – вдумчиво либо методом «случайного тыка».

Сбалансированность. Количество разделов базы тестовых заданий должны коррелировать с содержанием, относительными объемами информации разделов и формой дисциплины.

Стилистическая непротиворечивость. Тестовые задания формулируются в привычной для претендентов форме, с использованием словарных оборотов из рекомендованных им учебных пособий. Исключения – сюрпризные задания, включаемые в тест для противодействия возникновению в поведении претендентов эффекта «ровной дороги».

Логическая и семантическая непротиворечивость. Среди логических требований – определенность предмета измерения. Близость дисциплин затрудняет определение предметной принадлежности тестовых заданий. Чем меньше пересечений дисциплин и их разделов, тем чище, определеннее выражается в тесте их содержание. Логически тестовые задания можно ассоциировать с утверждением. В случае правильной подстановки утверждение превращается в истинное, в случае неправильной – в ложное. Привычные для педагога вопросы не являются утверждениями, а потому содержание вопроса не может определяться как истинное или ложное. Вопросы применяются в тестовых заданиях закрытой формы. Следует избегать несоответствий рода, числа, склонения в формулировках заданий и вариантов их выполнения. В ТЗ нет места двусмысленностям, например:

- на вопрос «Где находится Москва?» можно ответить «на берегу Москвы-реки», «в Европе» и пр.;
- на вопрос «Кто первым полетел в космос?» можно ответить «советский человек», «Юрий Гагарин», «Лайка» и пр.;
- на вопрос «Когда началась Великая отечественная война?» можно ответить; «в июне», «в XX веке» и пр.;
- в заданиях открытой формы следует указывать род, падеж и пр. ожидаемых слов. Например, на вопрос «Как называется судно, колющее лед?» можно ответить и «ледокол», и «ледоколом»;
- среди заданий на установление соответствия или на установление последовательности встречаются варианты, трактуемые в разных источниках по-разному (причины и последовательности исторических событий) [4].

Дистрактивность. Варианты выполнения заданий формулируются так, чтобы правильные варианты были похожи на неправильные, и наоборот. Формулировки правильных вариантов должны точно соответствовать материалу учебного курса. Следует избегать характерной ошибки начинающих авторов, когда формулировки верных вариантов, в большинстве, длиннее формулировок неверных вариантов. Следование принципу дистрактивности значительно повышает валидность теста.

Избыточность. Часто случается так, что в группу претендентов попадают учащиеся, учившиеся по разным программам и учебникам. Тест составляется один для всех и поэтому, чтобы «сравнять шансы», предпринимаются следующие шаги:

- в тест включаются задания, составленные в формулировках разных учебных пособий, примерно в равных пропорциях;
- эквивалентные по информации задания группируются, и участники информируются, что необходимо и достаточно правильно выполнить только одно задание группы;

- количество заданий в тесте превышает норму, которую мог бы выполнить усредненный участник сеанса.

Иногда завышают количество заданий теста выше нормы и из других соображений:

- чтобы увеличить шанс слабо успевающим учащимся хоть в чем-то проявить себя;
- чтобы исключить при проведении образовательной олимпиады накладки связанной с тем, что несколько участников выполняют правильно все предложенные тестовые задания.

Деловая игра. Технологии тестирования конструируются исходя из рекомендаций дидактики, утверждающей, что наилучших успехов можно достичь при игровом обучении. Игровая технология позволяет сделать процесс тестирования привлекательным и достичь высоких успехов за счет:

- естественной индивидуализации;
- возникновения соревнования;
- коллективности;
- публичности.

1.2.3 Четыре формы тестовых заданий

В тестологической практике принято считать, что существуют всего четыре формы тестовых заданий. Все остальное разнообразие можно классифицировать как их вариации.

Вместе с тем, существует мнение, что четырьмя типами все описать нельзя. Тестология развивается. Например, не так давно появилась методика дифференцированной оценки выполнения заданий. Она известна под названием «grading» (градуирование, распределение). В этой методике предполагается, что оценка выполнения заданий может быть не только двузначной «правильно - неправильно», но и принимать целый спектр значений. Считается, что более всего эта методика применима в математике.

Вот названия (характеристики) всех четырех традиционных форм тестового задания:

- задания с выбором ответов (испытуемый выбирает правильные ответы из числа готовых, предлагаемых в задании теста) - задание в закрытой форме;
- задания на дополнение (испытуемый сам дает краткий или развернутый ответ) – задание в открытой форме;
- задания на установление соответствия между элементами двух множеств;
- задания на установление правильной последовательности в ряду элементов.

Выбор формы зависит от:

- цели тестирования;
- содержания теста;
- технических возможностей;
- уровня подготовленности преподавателей и персонала.

Приведем (с некоторыми сокращениями) объемную цитату, проясняющую взаимоотношение формы и содержания тестовых заданий:

«Овладение формой является необходимым, но недостаточным условием создания полноценных тестов. Форма придает заданиям лишь структурную целостность и определенность, внешнюю организованность. Задания в тестовой форме только внешне похожи на тестовые задания, но это недостаточно для их включения в тест; нужна еще проверка свойств, позволяющих включить их в тест. Вместе с тем, меняющееся от дисциплины к дисциплине содержание позволяет абстрагировать форму в качестве самостоятельного предмета и рассматривать ее как подобие инварианта, независимого, в значительной степени, от содержания конкретной учебной дисциплины. Тем самым здесь подтверждается общий философский тезис о сравнительном постоянстве формы, остающейся устойчивой при изменении содержания...

Начиная с Аристотеля, форма понимается как идеальный принцип расположения элементов. Другое истолкование выдвинул И. Кант, согласно которому форма есть принцип упорядочения. Эти две идеи – организации и упорядочения, плюс идея существования содержания в каких-либо формах легли в основу нашего определения формы тестовых заданий: это способ организации, упорядочения и существования содержания теста.

Исследованием научных основ разработки тестов занимается педагогическая тестология. В западных странах получили развитие эконометрика, биометрика, социометрия и другие науки, имеющие своим предметом изучение количественных свойств и отношений, разработку объективных показателей состояния интересующих явлений...

Общие принципы разработки формы тестовых заданий рассматриваются в тестологии как важный предмет исследования, в то время как содержание, будучи не менее важным компонентом хорошего теста, относится все же не столько к тестологии, сколько к той науке, знание которой проверяется.» (В. С. Аванесов)

Первая форма – закрытая форма задания. Основой этой формы является закон исключенного третьего, сформулированный Аристотелем. Выбор правильного варианта порождает истину, выбор неправильного – ложь. Третьего не дано. Из этого закона следует: каждое задание должно иметь один верный вариант выполнения.

Чаще всего в литературе встречаются задания в закрытой форме, имеющие вопрос и четыре (или пять) варианта ответа, из которых один – верный. Это – уже классика.

Для выбора правильного варианта испытуемый должен совершить одно из действий:

- ввести код (цифры или буквы) требуемого ответа в специальном окне;
- щелкнуть мышью по «радиокнопке» с номером варианта;
- щелкнуть мышью по тексту варианта.

Главными недостатками классической закрытой формы являются:

- высокая вероятность угадывания верного варианта;
- непроизводительные затраты времени на прочтение всех вариантов;
- возможное непроизвольное подсознательное запоминание испытуемым неправильных вариантов как правильных.

Форма налагает логический запрет на применение таких вариантов, как: «правильного ответа нет», «все ответы правильные», «все ответы неправильные».

Он часто нарушается как в зарубежных, так и в российских тестах. Нарушения эти свидетельствуют о невнимательности либо об отсутствии культуры логического мышления среди составителей тестовых заданий.

В публикуемых в печати заданиях закрытой формы часто встречаются логические неувязки, несоответствия, многозначности. Это происходит в том случае, когда:

- вопрос содержит скрытый вопрос, который некоторые, особо искусственные в тестированиях, испытуемые могут принять за основной;
- вопрос содержит дополнительную контекстную информацию, могущую показаться неверной части испытуемых (особенно это касается истории, политологии и т. п.);
- вопрос многословен, содержит придаточные предложения;
- в задании отсутствует задание или вопрос, т. е. есть только текст в области задания и четыре варианта текста в области ответов;
- задание содержит утверждение с многоточием и варианты замещения многоточия.

Чаще всего, логические неувязки встречаются в четвертом из этих случаев.

Указанные и другие недостатки есть во всех тестах. Есть три пути борьбы с ними:

- повторное авторское редактирование;
- самотестирование автора;
- проведение тестирования в референтных группах, состоящих не только из учащихся, но и из преподавателей.

Наиболее быстрым, экономичным и эффективным следует признать третий путь.

Есть два достаточно известных пути развития закрытой формы, восполняющие указанный недостаток:

- задание с оцениванием верности, т. е. выбором одного, наиболее верного варианта из нескольких верных в разной степени;
- задание с выбором нескольких верных вариантов из нескольких предложенных.

Оба они так же считаются заданиями в закрытой форме.

Есть третий и четвертый пути развития задания в закрытой форме, еще более снижающие вероятность угадывания, описанию которых в данном пособии уделено особое внимание:

- задание с оцениванием степени верности каждого из вариантов;
- задание с оцениванием степени верности одного из вариантов, где испытуемый не может видеть все варианты одновременно.

В последнем случае испытуемому предоставляется возможность просматривать варианты, выпадающие в случайном порядке.

Третий и четвертый пути, кроме наличия очевидных преимуществ в оценивании степени владения знаниями, весьма перспективны для развития эвристичности мышления, а также для обучения студентов педагогических специальностей.

Вторая форма – открытая форма задания. Задание в открытой форме является обычно утверждением или большим текстом. В тексте есть специальные места, куда следует вносить информацию, оговоренную инструкцией. Обычно это «клетки», в каждую из которых вписываются буквы, цифры или пробелы. После заполнения клеток, текст превращается в истинное или ложное высказывание.

Обычно используется три вида тестовых заданий открытой формы:

- задания дополнения, где испытуемый должен сформулировать дополнения к предъявленным текстам;
- задания – кроссворды, ориентированные на выяснение знания претендентами определений, могущие содержать подсказки на пересечениях слов и игровой элемент;
- задания свободного изложения, где требуется изложить полное решение задачи, сочинение или перевод текста, где почти никакие ограничения на выполнение задания не накладываются.

Правильно сконструированные задания открытой формы полностью исключают догадку - главный недостаток заданий закрытой формы. Это - главное достоинство таких заданий.

Задания открытой формы часто вызывают недопонимания среди педагогов в связи с частотой одинаковых случаев: претенденты, выполняющие задания открытой формы по географии (истории, литературе, математике...) делают орфографические ошибки. Считать ли варианты с орфографическими

ошибками ошибочными? Ставить ли учащемуся, превосходно знающему теоремы и умеющему решать задачи, двойку по математике за то, что он неправильно пишет слова «прямая» и «плоскость»? [5]

Однако задания свободного изложения имеют существенный недостаток - трудно предсказуемая многовариантность. Результаты их выполнения невозможно полноценно проверить автоматически. По крайней мере, на сегодняшнем этапе развития компьютерных интеллектуальных систем. Можно с этим бороться:

- уделять больше внимания формулировке заданий, чтобы существенно сократить количество вариантов;
- указывать количество букв в допустимом варианте выполнения;
- предусмотреть все возможные верные варианты.

Все эти попытки приводят к резкому ограничению длины возможных текстов, вводимых претендентами в качестве вариантов ответа.

Учитывая, что количество цифр и букв ограничено, можно считать задание в открытой форме эквивалентным заданию в закрытой форме, только с очень большим количеством (миллионы, миллиарды) неправильных вариантов выполнения, скрытых от претендента.

Третья форма – задания на установление соответствия. Это задания, где элементам одного множества нужно верно сопоставить элементы другого. Вариантом выполнения задания является перечень связей (соответствий) элементов. Для обозначения этих связей поступают так. Элементы одного множества обозначаются цифрами, другого - буквами. Ответ конструируется как последовательность связанных цифр и букв. Например, для двух пятиэлементных множеств, нужно ввести в качестве результата выполнения задания кодовое слово: 1P2M3B4E5H6J.

Результаты выполнения задания на установлении соответствия:

- или пишутся в отведенной для этого строке по оговоренному инструкцией правилу;
- или указываются мышью графические изображения связей (соответствий).

Задания на установление соответствия рекомендуются для проверки ассоциативных и алгоритмических навыков пользования знаниями. Поиск аналогий на основе учебного материала, позволяет делать выводы не только о владении обучаемым теми или иными понятиями, но и о других характеристиках структуры знаний.

Задания на установление соответствия позволяют вести проверку усвоения сразу нескольких дисциплин. Поэтому они могут обладать большой проверочной емкостью.

Выполнение таких заданий легко проверять путем сравнения с единственным верным вариантом. Это жесткий подход. Но ведь претендент

может сделать одну ошибку – неужели из-за этого все задание он выполнил неправильно? Проверить выполнение таких заданий «мягко», дифференцированно - вручную нельзя, а с компьютером - можно.

Часто встречаемый недостаток заданий третьей формы – «особо начитанные» претенденты усматривают в таких заданиях не только изоморфные (однозначные) связи, но и гомеоморфные (многозначные). Это приводит к необоснованным снижениям оценок и справедливым апелляциям [1].

Поскольку результат выполнения задания третьей формы - кодовое слово с четко оговоренной длиной, в этом смысле эта форма совпадает со второй (открытой).

Четвертая форма – задания на установление последовательности. В этих заданиях требуется установить правильную последовательность вычислений, действий, шагов, операций, терминов.

Установив правильную, со своей точки зрения, последовательность, испытуемый вводит свой вариант в специально отведенном для этого месте путем щелканья мышью по последовательности mnemonicически понятных элементов (цифр и пр.) или ввода последовательности цифр с клавиатуры.

Эта форма заданий вызывает часто нарекания не из-за себя, а из-за недостатков интерфейса и инструкции, приводящим к разночтениям алгоритма ввода варианта.

Эту форму применяют для проверки верности понимания испытуемыми хода решения.

Четвертая форма заданий становится эквивалентной третьей, если представить одно из элементных множеств задания третьей формы простой последовательностью цифр.

1.2.4 Общие достоинства и недостатки тестовых технологий

С внедрением тестовых технологий в педагогике появилась возможность использования точных статистических методов анализа качества, позволяющих повысить объективность суждений о том, в какой степени усилия преподавателей и учащихся достигают цели.

Внедрение тестового контроля:

- ведет к совершенствованию приемов обучения, позволяя оперативно оценивать их результативность на основе объективных критериев;
- превращает педагогику в (допускающую измерение) точную науку;
- влечет появление новых воспитательных и обучающих методик.

Тестирование знаний нередко подвергаются критике со стороны педагогов. Изготовленные в России тестирующие программы мало отличаются

от тех, что предлагают зарубежные производители. К основным недостаткам тестирования часто относят:

- двузначность (не многозначность) логики вариантов ответа в наиболее популярных тестовых заданиях закрытого типа;
- невозможность контроля навыков устной речи;
- жесткость контроля, ведущая к психическому напряжению и сверхнормативной усталости претендентов;
- невозможность диагностики навыков общения (коммуникации);
- невозможность диагностики изобретательских качеств;
- невозможность диагностики оригинальности мышления в решении учебных проблем и задач;
- невозможность учета всех случайностей и факторов, которые всегда учит преподаватель-профессионал на экзамене – настроение, усталость, темперамент, возраст, пол, национальность экзаменуемого;
- невозможность диагностики ассоциативного и образного мышления, способности к обучению, желания обучаться.

Необходимо более подробно рассмотреть проблему двузначности логики тестеров.

В реальном учебном процессе, как правило, даже односложный ответ ученика грамотный учитель способен ранжировать по пятибалльной шкале, опираясь на свой опыт общения с ним и в зависимости от контекста урока.

Не бывает однозначной научной истины. Не бывает абсолютно точных ответов на все случаи жизни. Так устроена жизнь. Грамотный специалист, изобретатель, ученый может появиться лишь в такой учебной среде, где логика общения многозначна (с ответами, например: «да», «нет», «может быть», «не совсем точно», «не вполне»), где есть противоречия в знаниях, где есть место для фантазии [6].

Создавать тест, содержащий задания с многозначительными (многозначными) вопросами и ответами, оказывается, очень сложно!

Поэтому совместно с использованием тестовой формы контроля необходимо практиковать и традиционные формы – семинары, конференции, диспуты, обсуждения, деловые игры. Общепринято, что оптимального сочетания форм контроля можно достичь следующим образом.

С помощью компьютерных тестов проводится предварительный контроль (зачет) знаний определений и сведений, навыков решения задач, навыков восприятия речи на слух, грамотного написания слов и формул и т. п.;

После зачетных испытаний учащимся предлагается встреча с преподавателем, где они продемонстрируют уже не формальные знания, а навыки коммуникации и устной речи, остроту и оригинальность мышления, способность к учению.

1.3 Компьютерное тестирование

Среди продукции системы образования – знания, умения и навыки, переданные учащимся и воспринятые ими. Качество этой образовательной продукции играет в развитии общества первостепенную роль.

Без измерения и сравнения качества продукции невозможно развитие производства. В образовании для измерения качества применяют поурочный, рубежный, итоговый и прочие виды контроля. Каждый вид контроля характерен своей технологией.

Наиболее объективная технология массового контроля, по мнению многих, – автоматическая, однообразная, не зависящая от каких-либо субъективных влияний. Хотя, у этой точки зрения есть и противники, считающие, что контроль должен проводиться «глаза в глаза» [7].

Слабые места первой точки зрения: невозможно автоматическими средствами проверить и гарантированно оценить умения:

- излагать мысли;
- находить нестандартные решения;
- учить и учиться.

Вторая точка зрения также имеет слабые места:

- высокая себестоимость;
- отсутствие гарантии исключения субъективизма;
- невозможность гарантированно беспристрастно и единообразно оценить свойства большого количества претендентов.

В споре между точками зрения следует занимать «философскую» позицию, находить золотую середину. Например, такую:

- контролировать автоматическими средствами знания рутинных сведений, умений производить стандартные действия - т. е. то, что в вузе относят к «зачету»;
- контролировать силами специалистов умение мыслить и излагать мысли, а также другие свойства, не поддающиеся контролю со стороны автомата.

В мире разработано и внедрено множество подходов к оцениванию знаний, умений и навыков, основанных на применении компьютеров. Многие из них признаны достаточно объективными.

Однако среди известных компьютерных обучающих и тестирующих комплексов трудно найти абсолютно выдерживающий критику с точки зрения:

- объективности;
- удобства;
- привлекательности;
- полноты набора необходимых функций контроля и представления результата;
- объемности содержания.

Счастливым исключением - профессиональные военные тренажеры [1].

Перечислим недостатки традиционных систем оценки и аттестации ЗУН:

- субъективизм;
- нерегулярность;
- несогласованность требований и режимов контроля;
- сложность ведения статистической обработки (мониторинга);
- сложность обеспечения режима секретности хранения экзаменационных материалов и защиты от подлога;
- наличие случайных, несистематических ошибок измерений;
- отсутствие чётких математических критериев оценки.

Многое из перечисленного вполне может быть преодолено при внедрении компьютерных технологий. Однако, пока это - только пожелание, подтвержденное теоретически. Безупречной системы автоматического контроля знаний в настоящее время нет.

Связано это со многими причинами:

- разнородность, частая смена и относительная незрелость (молодость) как информационных технологий, так и соответствующего технического обеспечения;
- отсутствие должной государственной поддержки разработки и внедрения;
- отсутствие специалистов;
- отсутствие адекватной теоретической базы.

Путь преодоления этих проблем связан с новыми управленческими решениями и мероприятиями, в частности, такими:

- государственная поддержка развития тестологии как отрасли знаний и науки;
- создание индустрии систем тестового контроля [8].

В современном научном мире сложилась противоречивая ситуация с признанием тестологии как полноправной науки. С одной стороны, общее признание тестового контроля работоспособным направлением исследования широкого спектра педагогических и психологических проблем; с другой – затянувшаяся сдержанность, недостаток информации, непонимание сущности и возможностей.

Может быть, следует более конкретизировать понятие образования, более наглядно представить его в качестве товара? Там, где товар, там и сертификация, там и тестирование [1].

1.3.1 Преимущества компьютерного тестирования

Появление персонального компьютера, а также сопутствующих ему принтера и сканера, привело к значительной модернизации бланковой технологии и резкому подъему массовости бланковых тестирований.

Компьютер стал использоваться для подготовки (распечатки) бланков и для статистической обработки информации, которую испытуемые на бланки заносили в виде крестиков, галочек, и даже букв. По сравнению с ранними бланковыми технологиями, где результат обрабатывался вручную, снизились [10]:

- трудоемкость подготовки материалов и обработки результатов;
- вероятность внесения случайных ошибок в результат.

Вместе с тем повысились требования к аккуратности испытуемых, поскольку:

- используемое программное обеспечение не всегда верно позволяло идентифицировать нанесенные карандашом на бланк знаки;
- помятость бланка могла привести к его деформации в считывающих механизмах и появлению дополнительных погрешностей.

За последние два десятилетия в мире обозначилась тенденция к переходу от бланковых технологий к полностью электронным технологиям тестирования.

Компьютерные технологии тестирования позволили «забыть» о многих недостатках бланковых технологий. Рассмотрим преимущества использования компьютерного тестирования:

- Объективность. Исключается фактор субъективного подхода со стороны экзаменатора. Обработка результатов теста проводится с помощью компьютера.
- Возможность проверки различного объема знаний. Исключается фактор «лотереи» обычного экзамена, на котором может достаться «несчастливый билет» или задача. Большое количество заданий теста охватывает весь объем материала того или иного предмета, что позволяет тестируемому шире проявить свой кругозор и не «провалиться» из-за случайного пробела в знаниях. Генератор тестов позволяет по результатам тестовых испытаний посчитать процент усвоения каждого учебного элемента и всего курса (темы), как для отдельного студента так и средние характеристики для всей группы.
- Надежная обратная связь на всех этапах обучения. Тестирование позволяет выявлять лидеров по признаку успеваемости, обнаруживать массовые и индивидуальные пробелы в знаниях, своевременно корректировать содержание и форму учебного процесса и т.д.
- Простота. Тестовые вопросы конкретнее и лаконичнее обычных экзаменационных билетов и задач и не требует развернутого ответа или обоснования - достаточно выбрать правильный ответ и установить соответствие.
- Демократичность. Все тестируемые находятся в равных условиях.

- Массовость и кратковременность. Возможность за определенный установлений промежуток времени охватить итоговым контролем большое количество тестируемых.
- Побуждение интереса к изучаемой дисциплине. Содержащиеся в тестировании элементы игры - неопределенность, возможность случайности (повезет - не повезет) и др. Тестирование становится одним из рычагов приобщения студентов к компьютерной технике и технологиям;
- Быстрая обработка результатов. Проведение экзамена в форме компьютерного тестирования позволяет получить результаты тестирования сразу после его проведения.
- Внедрение компьютерного тестирования соответствует тенденции перехода к дистанционному образованию, представляя собой средство «удаленного» контроля знаний. Как известно, уже в настоящее время в Интернет функционирует развитая подсистема дистанционного образования – Internet Education, насчитывающая несколько миллионов компьютер-хостов [5].
- Быстрота сравнения результатов измерений для большого количества испытуемых на больших территориях.
- Внедрение и разработка новых технологий адаптивного тестирования.
- Возможность повышения дифференцирующей способности тестов.
- Оперативное использование методов статистической обработки для повышения уровня объективности результатов.
- Возможность повышения объективности результатов измерения путем применения новых способов демонстрации тестовых заданий, например, исключением возможности для испытуемого видеть все варианты ответов на вопрос задания закрытой формы и сравнивать их логическим перебором или заменой фиксированных вариантов наборов заданий случайными выборками.
- Оптимизация продолжительности тестовых измерений.
- Оперативное слежение за объемом и качеством знаний больших контингентов учащихся, сохраняя при этом все индивидуальные характеристики, выявленные для каждого.
- Увеличение скорости поиска и предоставления заданий в соответствии с алгоритмами адаптивного тестирования [9].
- Простота хранения, сортировки и формирования больших тестовых массивов в соответствии с алгоритмами.
- Реализация возможности автоматической массовой коррекции параметров заданий.
- Реализация возможности интерактивного адаптивного взаимодействия системы с претендентом.

- Упрощение ведения архива тестирований.

Таким образом, применение компьютерных технологий вполне может положительно повлиять на уровень качества образования.

1.3.2 Требования к компьютерному тестовому комплексу

Компьютерный тестовый комплекс – это совокупность компонент:

- база тестовых заданий, из которых по какому-либо алгоритму собирается тест;
- компьютерная программа, реализующая сценарий тестирования и обеспечивающая процесс тестирования испытуемых в сети или на одном персональном компьютере;
- пакет компьютерных программ обеспечивающих сбор, статистическую обработку и представление результатов тестирования;
- компьютерная база данных, аккумулирующая результаты тестирования.

К тестовым комплексам предъявляют, следующие требования:

1. Эффективность (оптимальность):

- точность измерения;
- скорость измерения.

2. Дифференцирующая способность:

- объективность;
- адаптируемость к условиям.

3. Мобильность:

- скорость доставки и развертывания до готовности;
- количество, доступность, стоимость, скорость предоставления необходимых для сеансов и их обработки материалов и услуг (доступ к сети, размножение инструкций и справочных материалов).

4. Дружественность интерфейсов:

- претендента;
- руководителя;
- редактора.

5. Качество БТЗ:

- наполненность (количество заданий);
- полнота (соответствие имеющихся ключевых заданий всем разделам дисциплины);
- вариативность, фасетность (количество возможных вариантов);
- сепарабельность (отделимость заданий в соответствии с темами и условиями).

6. Привлекательность:

процедуры измерения;
интерфейса претендента.

7. Скорость и простота

наполнения;

редактирования;

корректировки шкал и задания условий.

8. База результатов:

- представительность (наполненность, понятность) результатов;
- конвертируемость результатов в связи с изменениями системы оценки.

Таким образом, качество результата тестирования, т. е. степень соответствия его дидактическим целям и задачам, зависит от многих факторов.

2 ИНТЕЛЛЕКТУАЛЬНОЕ ТЕСТИРОВАНИЕ

2.1 Понятие «Искусственный интеллект»

В понятие «искусственный интеллект» вкладывается различный смысл - от признания интеллекта у ЭВМ, решающих логические или даже любые вычислительные задачи, до отнесения к интеллектуальным лишь тех систем, которые решают весь комплекс задач, осуществляемых человеком, или еще более широкую их совокупность.

Можно выделить две основные линии работ по искусственному интеллекту (ИИ). Первая связана с совершенствованием самих машин, с повышением «интеллектуальности» искусственных систем. Вторая связана с задачей оптимизации совместной работы «искусственного интеллекта» и собственно интеллектуальных возможностей человека.

Идея создания мыслящих машин «человеческого типа», которые, казалось бы, думают, двигаются, слышат, говорят, и вообще ведут себя как живые люди, уходит корнями в глубокое прошлое. Еще древние египтяне и римляне испытывали благоговейный ужас перед культовыми статуями, которые жестикулировали и изрекали пророчества (разумеется, не без помощи жрецов). В средние века и даже позднее ходили слухи о том, что у кого-то из мудрецов есть гомункулы (маленькие искусственные человечки) - настоящие живые, способные чувствовать существа. В настоящее время роботы, системы распознавания образов, экспертные системы и т.д. вызывают у непосвященного тот же трепет и восторг перед «думающей» машиной.

Но не зря в свое время были заморожены некоторые исследования в области искусственного интеллекта. Попытки создать машинный разум не удавались, и раз за разом энтузиазм ученых угасал, так как существующие на тот момент вычислительные средства не позволяли хотя бы приблизительно воссоздать взаимодействие нейронов головного мозга. Появление многопроцессорных систем и увеличение количества команд микропроцессоров и его тактовой частоты позволяет сейчас «построить» приближенное мышление человека с использованием параллельных процессов и нейронных сетей.

Обращаясь к проблеме роли ИИ в обучении и образовании, будем рассматривать этот процесс как одну из разновидностей взаимодействия человека с ЭВМ, и раскрывать среди перспективных возможностей те, которые направлены на создание так называемых адаптивных тестовых систем, имитирующих оперативный диалог учащегося и преподавателя-человека.

2.2 Методы и модели интеллектуального тестирования

Интеллектуальное тестирование предполагает наличие модели знаний, модели самого процесса тестирования и оценивания. Так можно

охарактеризовать в общем все разработки в этой области. Рассмотрим некоторые из них более подробно.

2.2.1 Модели распознавания образа уровня знаний

Под образом уровня знаний понимаются обучаемые, принадлежащие к множеству (группе), знания которых по «эталону уровня знаний» отнесены к лингвистическим оценкам неудовлетворительно (D), удовлетворительно (C), хорошо (B), отлично (A).

Под распознаванием образа уровня знаний понимается процедура принятия решения о принадлежности конкретного обучаемого к одному из указанных образов на основании сравнения его образовательных достижений при тестировании с характеристиками образа.

При тестировании по альтернативному признаку используется закрытая форма теста, характеристиками которой являются: функция плотности распределения неправильных ответов $f(d)$, приемлемый уровень неправильных ответов q_0 , неприемлемый уровень неправильных ответов q_1 , риск заниженной оценки знаний a , риск завышенной оценки знаний b , функция оценивания знаний $f(Q)$, объем образовательной информации N , объем выборки заданий теста n и критерий принятия решений в виде предельного числа неправильных ответов K .

Перечисленные характеристики являются взаимозависимыми, но не обладающими достаточным свойством четкости. В условиях их нечеткости для распознавания образа уровня знаний обучаемых вполне допустимо для нормально реализованной образовательной услуги принять модель распределения неправильных ответов по закону редких случайных событий Пуассона и функцию оценивания уровня знаний сформировать по этому же закону [11].

Поскольку образовательная информация в банке заданий теста N в их выборке n представляется как статистическая совокупность, а задания теста обучаемому в компьютерном варианте всегда для выполнения выдаются последовательно, то для распознавания образа уровня знаний возможно воспользоваться последовательным критерием Вальда. При этом примем дополнительное принципиальное условие, что задания теста однородны по количеству образовательной информации по конкретной учебной дисциплине, поскольку аналитических методов классификации заданий по мере их сложности или трудности пока не разработано.

Будем обозначать гипотезу о приемлемом уровне знаний H_0 , а гипотезу о неприемлемом уровне знаний H_1 . Пусть в результате последовательного поступления заданий теста в объеме n получены неправильные ответы $d_1, d_2, d_3 \dots d_n$. При известной функции оценивания знаний по закону Пуассона последовательный критерий Вальда позволяет по выборке объемом n

классифицировать обучаемых по уровню знаний на три подобраза по количеству областей принятия решений. Для того, чтобы иметь четыре образа необходимо произвести для каждой из трех областей повторное последовательное тестирование [11].

В предлагаемой процедуре рекомендуется использовать два способа распознавания образа уровня знаний: нормальный и усиленный. При этом задаются только четыре исходные характеристики теста q_1 , q_0 , a и b .

По нормальному способу по первой выборке заданий теста n_1 производится классификация обучаемых на три предварительные области (уровни): низкая, нормальная и высокая. По второй выборке заданий теста $n_2 = n_1$ или $n_2 < n_1$ для уровня низкий ужесточаются исходные характеристики q_0 и q_1 и обучаемые аттестуются по трем образам D, C и B. Для нормального уровня ужесточаются характеристики a и b и обучаемые аттестуются по трем образам C, B и A. Для высокого уровня тестирование осуществляется без изменения исходным q_0 , q_1 , a и b и обучаемые аттестуются на два образа B и A. К достоинствам нормального способа относится то, что обучаемые по второй выборке могут существенно улучшить свои образовательные достижения.

По усиленному способу по первой выборке обучаемые классифицируются только на два уровня: низкий и высокий. По второй выборке для низкого уровня ужесточаются характеристики q_0 и q_1 и обучаемые аттестуются только на два образа D и C. По второй выборке для высокого уровня ужесточаются характеристики a и b и обучаемые аттестуются только на два образа B и A. Достоинством усиленного способа является более уверенное распознавание образа уровня знаний и поэтому его рекомендуется использовать в тех случаях, когда снижено доверие к реализуемой образовательной услуге.

Существенным отличием предлагаемых методов распознавания образа уровня знаний при тестировании от известных является заранее заданная погрешность распознавания, заложенная в рисках принятия решений a и b , использование наиболее мощного критерия Вальда и достаточно простые и апробированные в статистическом приемочном контроле способы ужесточения планов тестирования, что позволяет использовать для решения практических задач распознавания образа уровня знаний международный стандарт ИСО 8423-91 «Статистические методы. Последовательные планы выборочного контроля по альтернативному признаку». Это способствует повышению достоверности компьютерного тестирования, что позволяет использовать предложенные теоретические разработки в качестве метода оценивания знаний [11].

2.2.2 Предметно - критериальная методика составления тестов

В каждом курсе есть ключевые моменты, особенно важные темы, без знания которых невозможно усвоение более сложного материала в процессе

учебы или которые будут необходимы в работе по специальности. На устном экзамене при личном контакте со студентом преподаватель обязательно оценивает понимание студентом этих тем. При автоматизированном тестировании можно учесть важность каких-либо разделов курса, увеличив долю вопросов по этим разделам в общем количестве вопросов. Но это не всегда удобно для составителя теста, потому что не всегда наиболее важные разделы содержат больше всего материала.

Предлагаемая методика предусматривает учет таких параметров, как степень важности и объем изучаемого материала в разделах курса.

При составлении теста преподаватель делит курс на темы $T_1, T_2, \dots, T_k$ и оценивает степень важности S_i и объем изучаемого материала V_i по каждой теме T_i . Количество вопросов n_i по каждой теме T_i должно соответствовать (быть пропорционально) объему изучаемого материала V_i .

Минимальное количество вопросов n_i по каждой теме T_i определяется в соответствии с методикой с учетом параметра V_i .

Знания по каждому разделу курса оцениваются по пятибалльной (а фактически по четырехбалльной) системе. Оценке «отлично» (5) соответствует вероятность правильного ответа от p_3 до 1; оценке «хорошо» (4) соответствует вероятность правильного ответа от p_2 до p_3 ; оценке «удовлетворительно» (3) соответствует вероятность правильного ответа от p_1 до p_2 ; оценке «неудовлетворительно» (2) соответствует вероятность правильного ответа менее p_1 . Следует отметить, что вероятности p_1, p_2 и p_3 ($0 < p_1 < p_2 < p_3 < 1$) задаются преподавателем с учетом структуры теста и могут быть изменены. Абсолютное количество (или доля) правильных ответов, достаточное для получения соответствующей оценки, определяется по специальной методике.

Итак, преподаватель:

- разбивает курс на темы (разделы) $T_1, T_2, \dots, T_k$;
- определяет их объемы $V_1, V_2, \dots, V_k$ и степень важности $S_1, S_2, \dots, S_k$;
- определяет структуру теста – количество m вариантов ответов на каждый вопрос;
- задает p_1, p_2, p_3 – уровни знаний студента (или вероятности выбора правильного ответа), соответствующие оценкам: «2» – $0 < p < p_1$, «3» – $p_1 < p < p_2$, «4» – $p_2 < p < p_3$, «5» – $p_3 < p < 1$.

p_1 должно быть заметно больше $1/m$ – вероятности выбора правильного ответа наугад.

После этого вычисляется минимальное количество вопросов n , необходимое для того, чтобы при заданных параметрах m, p_1, p_2, p_3 и заданном уровне значимости ϵ на основании испытания статистических гипотез можно было поставить оценку «5», «4», «3» или «2» за определенный раздел курса.

Минимальное количество вопросов n будет содержать тест по теме с минимальным значением $V_i = \min\{V_1, V_2, \dots, V_k\}$; $n_i = n$.

Минимальное количество вопросов по темам $T_1, T_2, \dots, T_k$ определяется пропорционально их объемам, $V_1, V_2, \dots, V_k$.

По ответам студента вычисляется оценка O_i по каждой теме T_i ($1 < i < k$) как результат испытания статистических гипотез

При вычислении итоговой отметки за тест (курс) O учитывается степень важности S_i каждого раздела T_i . Получившаяся итоговая оценка O округляется до целых.

Следует отметить, что описанная выше методика позволяет давать студентам тест поэтапно, по мере изучения и усвоения материала отдельных разделов курса, и выводить итоговую оценку с учетом результатов промежуточного тестирования [12].

2.2.3 Метод определения количества образовательной информации

Теория образовательного тестирования должна формироваться на частных законах и закономерностях таких научных направлений как информатиология, общая статистика, статистический приемочный контроль, квалиметрия, педагогика, психология, исследование операций, теория принятия решений и др. Прямое применение теоретических разработок из указанных научных направлений не дает заметных практических результатов по оцениванию знаний по причине нематериальности знаний, как объекта исследований. Задачу формирования теории образовательного тестирования можно сформулировать как задачу поиска оптимальной структуры специфических законов и закономерностей тестологии, позволяющую оценить знания с заданной погрешностью.

Для решения задач подобного класса наиболее успешно используются генетические методы, основанные на реализации генетических алгоритмов, позволяющих осуществить направленный перебор частных законов и закономерностей по наиболее приемлемым направлениям для формирования отечественной теории образовательного тестирования.

В отличие от традиционного случайного поиска приемлемых решений, алгоритмы генетического поиска используют аналоги или близость имеющихся решений во многих областях знаний к поиску оптимального набора специфических законов, обеспечивающих объективность, достоверность и точность оценивания уровня знаний, воспроизведенных обучаемыми в процедурах тестирования. Такой направленный перебор частных законов является эволюционным и имеет очень много сходств с операторами, применяемыми в генетических алгоритмах и процедурах, происходящих с живыми организмами в природе.

Рассмотрим применение генетических алгоритмов для формирования специфического закона о количестве образовательной информации. Исходные популяции: Государственный образовательный стандарт, учебная программа,

специфическая совокупность учебной информации, банк тестовых заданий. Репродукция: образовательная совокупность. Скрещивание: образовательная совокупность, статистическая совокупность. Мутация: образовательная совокупность информации.

Следующий генетический алгоритм направлен на поиск единицы образовательной информации. Исходные популяции: единица статистической совокупности, единица допуска, информации. Репродукция: единица образовательной совокупности. Скрещивание: единица образовательной совокупности, единица допуска, информации. Мутация: условная единица образовательной информации конкретной дисциплины.

Формирование специфического закона тестологии об образовательной информации: образовательная информация является первичной, поскольку независимо от образовательной услуги, формы теста, процедур тестирования и уровня подготовки обучаемых знания по конкретной дисциплине оцениваются только по их соответствию «образу знаний». Вся остальная информация является вторичной и третичной и не может претендовать на такую же роль как образовательная информация.

Любая информация, и в том числе образовательная, для ее последующего применения в заданиях теста должна быть представлена определенным количеством, рассчитанным с использованием условной единицы образовательной информации.

Следующий закон о сохранении образовательной информации определяет, что количество образовательной информации H_Q и количество ее энтропии I_Q величина всегда постоянная для всех процедур тестирования. Вычисление количества информации и количества энтропии производится по одной и той же формуле. При этом H_Q вычисляют только после создания тестов, а I_Q до их создания, что позволяет погрешности тестирования определить априорно. Такие вычисления невозможны без условной единицы образовательной информации, под которой понимается наиболее типичное и применяемое понятие в конкретной учебной дисциплине, поскольку обобщенного понятия пока получить не удастся. Например, в материаловедении это «свойство материала», в технологии машиностроения это «операция». В последующем количество образовательной информации пересчитывается по аналогии.

Реализация указанных законов об образовательной информации позволяет с достаточной для практической цели точностью определять количество информации в банке тестовых заданий, в одном задании теста, в выборке заданий теста и в выборке выполненных тестов и обеспечивать соблюдение минимально необходимого соотношения между объемом выборки и банком тестовых заданий, соответствующего выбранной погрешности оценивания уровня знаний по образовательным тестам [13].

2.2.4 Информационно-генетические алгоритмы

Основные свойства образовательных тестов предлагается формировать на популяциях частных законов таких научных отраслей как: информатология; психология, педагогика и психодиагностика; логика; теория вероятностей; теория поиска; теория нечетких множеств; теория игр; теория статистических решений; приемочный выборочный контроль.

Эти популяции позволяют реализовать информационно-генетический алгоритм и получить новое поколение специфических законов теории тестирования о (об): первичности образовательной информации; «образе знаний», воссозданном по первичной информации; «образе уровня воспроизведенных знаний»; количестве и энтропии образовательной информации; единстве количества образовательной информации; минимально допустимом соотношении между количеством образовательной информации в «образе знаний» и в выборке заданий теста; условной единице образовательной информации; энтропии нормальной образовательной услуги; не материальности знаний, как объекта исследований; соответствии формы теста и признака оценивания; формах существования функции оценивания знаний; характеристиках доверия к результатам тестирования; правах тестируемых на получение объективной оценки уровня знаний; защите прав тестируемых при воспроизведении ими знаний по образовательным тестам; переходе количественных результатов тестирования в качество «уровня знаний».

В качестве примера рассмотрим применение информационно-генетических алгоритмов на трансформацию международных и отечественных стандартов ГОСТ Р 50 779.71-99 и ГОСТ Р 50 779.72-99 на статистический приемочный контроль, применительно к задачам тестирования, позволяющих предложить способы выделения «образов уровня знаний» при тестировании по двум независимым выборкам заданий теста закрытой формы, и применению традиционных лингвистических оценок: отлично (I), хорошо (II), удовлетворительно (III) и неудовлетворительно (IV) [3].

Исходные данные для реализации «образа уровня знаний»: N – объем банка заданий (образ знаний); n – объем выборки заданий; a — риск занижения оценки; b — риск завышения оценки; $AQL(q_0)$ – приемлемый процент неправильных ответов; $RQL(q_1)$ – неприемлемый процент неправильных ответов; C_I – приемлемое число неправильных ответов; степень тестирования – абсолютный объем выборки; уровень тестирования – соотношение между объемом выборки n и числом C в зависимости от предшествующих результатов ответа на задания теста; QL – предельный процент неподготовленных обучаемых, которые могут получить завышенную оценку.

Для нормальной образовательной услуги характерен нормальный процесс восприятия и воспроизведения знаний обучаемыми, нормальный «белый шум».

В таких условиях неправильные ответы на задания теста предпочтительного соотношения 5-1 (пять ответов, из которых один правильный) вполне оправданно считать как редкие случайные события и функцию оценивания знаний сформировать по закону Пуассона.

Для конкретного примера задаем $N=250$; $AQL=10\%$; $RQL=20\%$; $QL=10\%$; $a < b$; степень II по ГОСТ Р 50 779.72-99. Объем первой выборки $n_1 = 20$ и критерии принятия решений $C_1 < 5$, $C_2 = 6$. Для второй выборки применяем усиленное тестирование $n_1 = 20$, $C_3 < 3$, $C_4 = 4$, а для нормального тестирования оставляем исходный план $n_2 = 20$, $C_1 < 5$, $C_2 = 6$.

По первой выборке уровень тестирования принят нормальный (классификация), а тестируемые разделяются на две группы: у — недостаточная подготовка и Е — достаточная подготовка. По второй выборке (аттестация) для группы у тестирование производится по усиленному уровню (ужесточенному) и деление производится на два образа IV и III.

Для группы Е уровень остается нормальным, но время на выполнение заданий сокращается. Тестируемые делятся на два образа II и I. Из схемы видно, что по второй выборке тестируемые получают возможность на улучшение результата, что реально защищает их права на объективность оценивания уровня знаний. Далее имеем скрещивание частных законов, которые проявляются в мутации специфического закона тестирования о переходе количества неправильных ответов в качество знаний, проявляющихся в лингвистической форме. Риски (ошибки) попадания в образы по второй выборке $b=0,03$, $v=0,16$. Предельный процент тестируемых с низким уровнем знаний, но получивших положительные оценки $QL=q_0=10\%$. Разработаны также методы выделения «образов уровня знаний» и для количественного признака, когда каждое выполненное задание имеет количественное значение в диапазоне $[0, 1000]$, однако рамки статьи не позволяют привести такие примеры.

Таким образом, использование информационно-генетических алгоритмов для выделения необходимых свойств образовательных тестов в форме законов тестирования и их реализация для выделения «образа уровня знаний» наглядно показывает необходимость дальнейших исследований по их применению для решения новых задач по оценке уровня знаний в экзаменах с целью повышения их объективности, достоверности и эффективности [14].

2.2.5 Модель Раши

Система тестирования на основе модели Раши обладает важными достоинствами, среди которых, прежде всего, необходимо отметить следующие.

Модель Раши превращает измерения, сделанные в дихотомических и порядковых шкалах в линейные измерения, в результате качественные данные

анализируются с помощью количественных методов. Это позволяет использовать широкий спектр статистических процедур.

Оценка трудности тестовых заданий не зависит от выборки испытуемых, на которых была получена и, аналогично, оценка уровня знаний испытуемых не зависит от используемого набора тестовых заданий.

Пропуск данных для некоторых комбинаций (испытуемый — тестовое задание) не является критическим.

Сама система тестирования достаточно проста, по сравнению с другими аналогичными системами она характеризуется наименьшим числом параметров — только один параметр уровня знаний для каждого испытуемого и только один параметр трудности для каждого задания.

Модель Раша опирается на четкие и конструктивные понятия «трудность задания» и «уровень знаний». Так, одно задание считается более трудным, чем другое, если вероятность правильного ответа на первое задание меньше, чем на второе, независимо от того, кто их выполняет. Аналогично, более подготовленный студент имеет большую вероятность правильно ответить на все задания, чем менее подготовленный.

Благодаря простой структуре модели существуют удобные вычислительные процедуры для многоаспектной проверки адекватности модели: для всего набора тестовых результатов, для каждого испытуемого, для каждого задания и для каждого конкретного ответа.

«Остатки», получаемые при аппроксимации результатов тестирования моделью можно использовать для выделения различных типов испытуемых.

Однако, несмотря на 40-летний опыт применения этой системы тестирования за рубежом во многих областях знания, прежде всего в образовании, медицине и психологии, до сих пор продолжаются дискуссии об истинной ценности и эффективности системы тестирования на основе модели Раша. До сих пор существуют две крайние точки зрения на эту модель тестирования.

Наиболее убежденные сторонники модели Раша утверждают следующее: «Можно ли собрать или построить или сформулировать данные так, чтобы они соответствовали определению измерения (модели Раша)? Если нет, — то такие данные бесполезны».

Их наиболее последовательные оппоненты утверждают следующее: «Данные — это данные, а модель — это конструкция исследователя, которая подвержена ошибкам». Например, при построении регрессии, выбрасывая те или иные данные, можно получить любую зависимость, но мы тем самым ограничиваем реальный мир данных. Таким образом, мы создаем искусственную переменную, о которой мало что знаем.

Для практики одним из наиболее важных критериев является точность оценивания. Поэтому выбор темы в значительной степени обусловлен

противоречивой информацией относительно точности системы тестирования на основе модели Раша. Кроме того, не удалось найти работы, в которых проведен всесторонний анализ точности модели Раша. В известных работах только даются те или иные общие рекомендации по использованию этих моделей.

Чем больше точность, тем лучше работает модель. В случае отсутствия ошибок измерения любая модель в смысле точности измерения работает идеально. Но на практике ошибки всегда есть и поэтому важно знать, насколько точные оценки позволяет получать та или иная модель.

На основе имитационного моделирования исследуются точность оценивания уровней знаний и трудностей заданий, а также число итераций, требуемых для вычисления этих оценок (методом наибольшего правдоподобия) в многофакторной ситуации в зависимости от:

- диапазона уровней знаний испытуемых;
- диапазона трудностей заданий;
- степени соответствия диапазонов уровней знаний испытуемых и трудностей заданий;
- числа испытуемых;
- числа заданий;
- степени соответствия данных модели;
- доли пропущенных данных.

Для статистической обработки результатов моделирования используется многофакторный дисперсионный анализ [15].

2.2.6 Абсолютная временная шкала измерения знаний

Знания являются абсолютной субстанцией: они либо есть, либо их нет. По крайней мере, так считается в любой форме традиционного оценивания знаний - как на выпускных экзаменах в школах, так и на вступительных экзаменах в вузы. Поэтому интересно проанализировать возможности абсолютных шкал оценки и при переходе к измерению знаний на основе тестов.

В данных исследованиях изучаются возможности так называемой «абсолютной временной шкалы оценивания знаний». Формулируются ее принципы. Формулируются этапы последовательного перехода от традиционной формы экзаменов к тестовой форме этого подхода, на их основе - требования к созданию тестовых материалов этого подхода.

Анализируется опыт использования данного подхода на вступительных экзаменах в Тверском государственном университете на протяжении 4-х лет.

Изучается диагностический потенциал данного подхода. Формулируется принцип «трехуровневого абстрагирования» для диагностических тестирований.

Тестирование с использованием абсолютной шкалы оценки имеет ценность диагностическую даже более, чем для итоговых экзаменов.

Изучается уровень достоверности результатов компьютерного тестирования в данном подходе и соотношение «случайного» и «достоверного» в итоговой оценке. В ходе тестирования была рассмотрена зависимость (в данном подходе) итоговой оценки по математике от времени тестирования: за все время тестирования (40 минут) в первые 15 минут (первые 4 задания по математике) оценка менялась наиболее заметно. За последние же 10 минут итоговая оценка изменялась не более чем на 10 баллов - доля «случайного» в итоговой оценке.

Таким образом, при использовании абсолютной шкалы данного подхода существует возможность ответить на вопросы:

- 1) существует ли предел, к которому сходится итоговая оценка с увеличением времени тестирования (или количества заданий теста);
- 2) какова погрешность «измерения знаний» если прервать тестирование в некоторый определенный момент, например через 40 минут.

Еще одно очевидное преимущество абсолютной шкалы оценивания — итоговая оценка появляется на экране компьютера сразу же после выполнения теста испытуемым [16].

2.2.7 Методика статистического анализа качества обучения

Предлагаемая методика основывается на том, что учебный процесс является частным случаем технологического процесса и ему должны быть свойственны такие же методы анализа, какие приняты для производственных процессов. Однако слепо перенести подобные методики нельзя, особенно это касается содержательного анализа процесса.

Для того чтобы проанализировать учебный процесс нужно иметь, во-первых, критерий качества обучения, а, во-вторых, проследить его изменение во времени. В качестве наиболее информативного критерия качества обучения следует использовать степень обученности учащихся - СОУ. Этот критерий основан на статистике полученных учащимися оценок за выполнение отдельных заданий или контрольных работ. Оценки входят в СОУ с «весом» равным интегралу вероятности получения данной оценки для некоторого «типového» распределения оценок.

В качестве такого «типového» распределения используется стандартное распределение Гаусса с параметрами: среднее значение оценки - 4 и стандартное отклонение - 1,39. Такое распределение обладает одним особым свойством: для этого распределения значения СОУ и качественной успеваемости совпадают и составляют 0,64. Это свойство выделяет «типové» распределение среди других распределений со средней оценкой 4.

Расчеты для «типového» распределения показывают, что если СОУ больше 0,76, то обученность «отличная», если СОУ от 0,5 до 0,76, то

обученность «хорошая», если СОУ от 0,24 до 0,5, то обученность «удовлетворительная», если менее 0,24, то «неудовлетворительная».

Для оценки изменения СОУ во времени используется известная в математической статистике методика, связанная с критерием «3s». Согласно этой методике, если какой либо процесс идет нормально, то отдельные значения должны укладываться в интервал «3s» относительно среднего значения (s - стандартное отклонение) с определенной точностью. Те значения, которые не укладываются в заданный интервал, являются отклонениями от стандартного распределения. Чем меньше таких отклонений, тем больше соответствие анализируемого распределения стандартному. Что касается применения этой методики для технологических процессов, то ее надо скорректировать - следует учитывать только те значения, которые выходят за нижнюю границу интервала.

Если взять отношение числа значений попадающих в интервал «3s» к общему количеству значений, то такую величину можно назвать коэффициентом стандартности распределения, а в случае рассмотрения учебного процесса - коэффициентом отлаженности учебного процесса (КОУП). Расчеты показывают, что если значение КОУП больше 0,94, то процесс можно считать «отлично отлаженным», если КОУП от 0,84 до 0,94 - «хорошо отлаженным», если КОУП от 0,69 до 0,84 - «почти отлаженным», если менее 0,69 - «не отлаженным».

Для общей оценки учебного процесса можно перемножить среднее значение СОУ по предмету за год на КОУП. Полученную величину можно трактовать как фактор качества учебного процесса (ФКУП). Этот фактор имеет большее число градаций, чем СОУ и КОУП. «Отличному» качеству соответствует ФКУП больше 0,71, «очень хорошему» от 0,64 до 0,71, «хорошему» от 0,41 до 0,64, «удовлетворительному» от 0,17 до 0,41 и «неудовлетворительному» менее 0,17 [17].

2.2.8 Технология рейтинговых исследований качества образования с применением нейронных сетей

Основными компонентами процесса оценки качества образования являются: сбор исходных данных и получение результирующей информации о состоянии системы образования.

В информационной системе мониторинга качества образования (ИСМО), создаваемой в Брянском государственном техническом университете, разрабатывается информационная технология рейтинговых исследований качества образования на основе применения нейронных сетей. Программный комплекс нейрокомпьютерной обработки применяется в качестве одного из модулей финишной обработки информации. Выходная информация такого модуля - вычисленные «рейтинги качества» для каждого вуза [9].

Исследования проводятся по следующим основным направлениям:

классификация - определение принадлежности каждого вуза к конкретной группе качества.

На первом этапе выявляются основные тенденции в образовании, такие, как уровень разрыва в качестве образования, плотность распределения объектов мониторинга по группам качества. При этом возможны следующие варианты реализации первого этапа:

- глобальная кластеризация — предполагает наличие в исходной выборке данных по отечественным и зарубежным вузам;
- федеральная кластеризация - наличие в выборке данных по вузам из подчинения Министерства образования;
- отраслевая кластеризация - выборка состоит из данных по однотипным вузам (техническим, педагогическим, медицинским, и т. д.);

Для адекватности кластерных исследований периодичность их проведения должна быть достаточно большой.

На этапе классификации исходные данные по каждому вузу подаются на входы обученной нейронной сети, после чего определяется «рейтинг качества» данного вуза. Многие параметры, по которым происходит оценка качества образования, подвержены динамическим изменениям. Поэтому определение рейтинга качества образования в информационной системе будет выполняться ежеквартально.

Ряд последовательных операций на каждом этапе определяет облик соответствующей информационной технологии. Информационная технология этапа кластерных исследований определяется следующими операциями:

- получение исходных данных для проведения исследований с сервера информационной системы (отдельно для каждого этапа реализации);
- проведение кластерных исследований посредством самоорганизующейся карты Кохонена;
- идентификация «групп качества образования», на которые было разбито исходное множество учебных заведений;
- присвоение «рейтингов качества» каждой «группе качества».

Завершающая операция на данном этапе может дать заключение о глобальных тенденциях в качестве образования и являться основанием для проведения мероприятий организационно-управленческого плана. Такой тенденцией может быть, например, резкое увеличение количества объектов в группах с низким рейтингом качества.

В ИСМО в основу технологии классификации положено использование многослойной нейронной сети, обучаемой по методу обратного распространения ошибки. В данном случае технологический облик этапа определяется особенностями процесса обучения нейронной сети и состоит из следующих операций:

- формирование исходной выборки для обучения нейронной сети, в которой на входы нейронной сети подаются параметры качества образования, а на выходы-рейтинги качества;
- подача на входы сети нового примера, соответствующего вузу, рейтинг которого определяется;
- выполнение предыдущей процедуры для всех вузов, включенных в систему мониторинга.

Выполнение первого этапа, на котором генерируются примеры для обучения нейронной сети, может основываться на методе экспертных оценок. Функционирование информационной системы происходит в полуавтоматическом режиме. На персонал информационной системы возлагается обязанность пополнения базы данных параметров оценки качества образования.

Текущая обработка по этапу классификации может изменять рейтинг качества конкретного вуза, но не меняет картину качества в целом. Этим определяется небольшой период проведения итоговых обработок по этому этапу. Итоговые обработки, проводимые в рамках кластерных исследований, способны изменить общую картину качества [9].

2.2.9 Модель адаптивного тестового контроля

Процедура тестирования предполагает анализ ответов на последовательность тестовых заданий определенной сложности. Проведем аналогию с поведением поискового алгоритма оптимизации для некоторой гипотетической функции Y , максимум которой необходимо найти. В задачах оценивания по тестированию - это максимум функции уровня знаний.

Реализация поискового алгоритма сводится к последовательному анализу локальной окрестности функционала Y , оценки градиента и выбора очередной области исследования. Если при оценке градиента имеют место помехи, то нельзя говорить о сходимости алгоритма. В обычном смысле он сходится вообще не будет, а будет «блуждать» вокруг области экстремума.

Аналогично можно поступить в случае тестового контроля. Если ответ правильный, то предполагается, что уровень подготовки студента выше сложности предъявленной задачи и он способен решать задачи заданной сложности, в противном случае - неспособен. Это подобно оценке градиента некоторой гипотетической функции регрессии, в которой градиент сам является случайной величиной.

2.3 Адаптивное тестирование

Адаптивное тестирование — это широкий класс методик тестирования, предусматривающих изменение последовательности предлагаемых заданий в зависимости от сложности в самом процессе тестирования с учетом ответов

испытуемого. К адаптивному тестированию обычно относят алгоритмы постановки заданий, построенные для пунктов теста, предварительно выбранных с помощью различных моделей и методов анализа пунктов. При таком обучении в процессе прохождения теста строится модель обучаемого, которая используется для генерации последующих заданий тестирования в зависимости от уровня обучаемого.

Как отмечает М.Б. Челышкова «в последние годы в практике образования складывается ситуация, под влиянием которой традиционное тестирование, осуществляемое с помощью стандартизированных тестов фиксированной длины, перерастает в современные эффективные формы адаптивного тестирования, базирующегося на отличных от традиционных теоретико-методологических основах и иных технологиях конструирования и предъявления тестов...

На протяжении ряда лет вопросы адаптивного тестирования неоднократно привлекали внимание зарубежных ученых. Об этом свидетельствуют многочисленные исследования фундаментального характера и публикации таких авторов как J.A. Arter, R.K. Hambleton, J.L. Horn, C.D. Jensema, G.G. Kingsbury, F.M. Lord, J. Millman, L. Nauels, R.J. Owen, K.J. Patience, M.D. Reckase, J. Spray, H. Swaminathan, V.W. Urry, M. Waters, D.J. Weiss, R.W. Wood, A.R. Zara и многих других. С начала 90-х годов компьютерное адаптивное тестирование получило за рубежом широкое признание в сфере практики. В значительной мере этому способствовали прикладные теоретические исследования таких ученых, как C.V. Bunderson, D.K. Inouye, G.G. Kingsbury, J.B. Olsen, H. Wainer, D.J. Weiss, работы которых были не только поддержаны педагогической общественностью, но и нашли широкое применение в практике профессионального отбора в ряде структур промышленного и военного комплексов» [18].

Выше была проведена аналогия с поведением поискового алгоритма оптимизации для некоторой гипотетической функции Y , максимум которой необходимо найти, где в задачах оценивания по тестированию это максимум функции уровня знаний.

Данная технология наглядно продемонстрирована на приведенном далее графике: по оси абсцисс откладывается номер вопроса, предъявляемого испытуемому, по оси ординат - уровень подготовленности испытуемого и сложность вопросов (в единой шкале логитов).

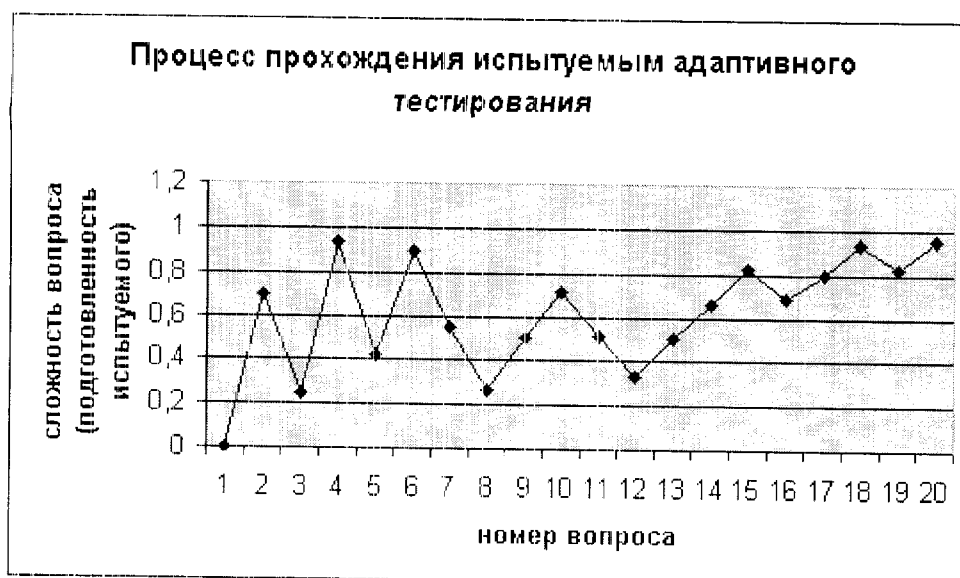


Рисунок 1 – Процесс прохождения испытуемым адаптивного тестирования

Предлагается использовать следующий подход. Считаем, что если тестируемый решил задание, то у него появляется желание решить более сложное задание. Если нет - то им будет сделана еще одна попытка решения задания той же сложности. Если оно также не решено, то предъявляется задача пониженной сложности. Если сразу не решено менее сложное задание, то к решению предлагается задача меньшей сложности. Аналогично происходит процесс повышения сложности заданий. В результате, если исключить этап обучения при решении задач, студент выберет для себя определенный уровень сложности, вокруг которого и будет размываться сложность заданий.

Сложность тестового задания задана некоторым числовым значением, таким образом, процесс тестирования можно представить в виде однородной марковской цепи, в которой состояниями цепи являются меры сложности заданий, а вероятности переходов по сложностям определяются лишь тем, в каком состоянии находилась система на предыдущем шаге.

Адаптивное обучение и адаптивный тестовый контроль, в своем единстве, представляют собой современный компьютерный вариант реализации принципа индивидуализации обучения.

Адаптивное обучение:

- помогает открыть возможности свободного выбора траектории учения;
- потребует определения перечня учебных задач и тестовых заданий для проверки подготовленности не только по минимальным, но и по более высоким требованиям;
- основано на использовании оценок, получаемых при входном адаптивном тестовом контроле. По итогам контроля выбирается первый учебный материал, с которого и начинается адаптивное обучение. Следующий за этим контроль выявляет меру усвоения предложенного материала, и т.д.;

- позволяет обеспечить выдачу учебного материала на оптимальном уровне трудности. Известно, что легкие материалы не обладают заметным развивающим потенциалом, а трудные задания снижают учебную мотивацию у большинства учащихся;
- обеспечивает знания без привычных, для массового образования, многочисленных пробелов в индивидуальной подготовленности студентов;
- достигает требуемой структуры и желаемого уровня знаний;
- позволяет регулировать трудоемкость и количество предъявляемых заданий в зависимости от ответа учащегося на текущее задание. В случае правильного ответа следующее задание он получит труднее, в случае неправильного – легче.

Целесообразность адаптивного контроля вытекает из соображений рационализации традиционного процесса тестирования, где всем претендентам дается одинаковый набор заданий. Знающему претенденту нет необходимости давать легкие задания, из-за высокой вероятности их правильного решения. Из-за высокой вероятности неправильного решения нет смысла давать трудные задания слабому претенденту. Использование заданий, соответствующих уровню подготовленности, существенно повышает точность измерений и минимизирует время индивидуального тестирования.

Не менее важным является вопрос секретности заданий: для каждого испытуемого набор заданий является уникальным, в результате чего минимизируется возможность подсказок, выучивания правильных ответов и т.д. При повторном прохождении теста испытуемый также вынужден решать новые задания, что уменьшает влияние эффекта тренированности.

Существует три варианта адаптивного тестирования:

Пирамидальное тестирование – всем претендентам дается задание средней трудности и уже затем, в зависимости от ответа, каждому дается задание легче или труднее. На каждом шаге применяется правило деления шкалы трудности заданий пополам. Например, если в тесте имеется двадцать одно упорядоченное по трудности задание, тестирование начинается с одиннадцатого. Если претендент действовал правильно, ему предъявляется шестнадцатое задание и, в случае успеха, девятнадцатое. Аналогично, слабо подготовленный претендент после неудачи на одиннадцатом задании пробует силы на шестом, и далее, по тому же принципу, до момента стабилизации на близких по трудности заданиях;

Flexilevel – когда контроль начинается с любого уровня трудности, а потом происходит постепенное приближение к реальному уровню подготовленности;

Stradaptive – когда тестирование проводится с базы тестовых заданий, где задания разделены по уровням трудности. После верного выполнения,

следующее задание берется из более высокого уровня трудности, после неверного -- наоборот. Это требует предварительного определения трудности всех заданий.

В адаптивном тесте заранее известны параметры трудности и дифференцирующей способности каждого тестового задания. Он создается в виде базы тестовых заданий, где задания упорядочены в соответствии с их характеристиками. Главная характеристика заданий адаптивного теста - это их сложность, полученная эмпирически. Это означает: каждое задание адаптивного теста прошло эмпирическую апробацию на заданной выборке, которая соответствует предполагаемым характеристикам претендента.

Адаптивный контроль позволяет рационализировать традиционное тестирование. В рамках адаптивного контроля эффективно выбираются и используются те задания базы тестовых заданий, которые соответствуют уровню подготовленности претендента. В итоге, адаптивное тестирование определяется как «совокупность процессов генерации, предъявления и оценки результатов выполнения адаптивных тестов, обеспечивающую прирост эффективности измерений по сравнению с традиционным тестированием благодаря оптимизации подбора характеристик заданий, их количества, последовательности и скорости предъявления применительно к особенностям подготовки тестируемых обучающихся» [18].

2.4 Обработка данных в тестологии

При разработке банка тестовых заданий неизменно возникает необходимость обработки больших объемов данных, сравнение испытуемых между собой, создание новых методик, их апробация на различных выборках, оценка правомерности использования тех или иных методик применительно к той выборке, с которой работает преподаватель, сведение воедино данных полученных при использовании различных методик, как на одной, так и на разных выборках.

Большинство из вышеперечисленных задач по традиции (или скорее по привычке) решаются при помощи классической теории тестирования (Classical Test Theory – СТТ). Классическая теория тестирования была разработана для анализа и построения тестов. Она рассматривает полученные при тестировании баллы как постоянные числа, и основывается на изучении статистик от начальных данных.

2.4.1 Классическая теория тестирования

Классическая теория тестирования предполагает следующую последовательность действий. Вначале первичный (исходный) тест предоставляется обучаемым для контроля знаний на этапе самопроверки: во время подготовки к практическим занятиям, лабораторным, экзаменам.

Результаты прохождения тестов фиксируются в базе данных, а затем обрабатываются с применением методов классической теории тестирования. На этом этапе обработки результатов определяются следующие величины:

- средний арифметический балл для всех испытуемых

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}, \quad (3)$$

где X_i - итоговый балл i -го испытуемого;

- сумма квадратов отклонений результатов испытуемых от $\bar{X}$

$$SS_X = \sum_{i=1}^N (X_i - \bar{X})^2 \quad (4)$$

- дисперсия результатов испытуемых

$$D = \frac{SS_X}{N-1} \quad (5)$$

- стандартное (среднеквадратическое) отклонение результатов испытуемых от среднего

$$S_X = \sqrt{D} \quad (6)$$

При нормальном распределении результатов тестирования $X = 3S_X$;

- коэффициент корреляции Пирсона:

$$r_{XY} = \frac{SP_{XY}}{SS_X * SS_Y}, \quad (7)$$

$$SP_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) \quad (8)$$

где X и Y - параметры, связи которых между собой рассчитываются; SS_X , SS_Y - суммы квадратов отклонений x и y от их средних значений.

Все те задания, для которых коэффициент связи с итоговым баллом r_{XY} меньший или равен нулю, непригодны для контроля знаний, поэтому их необходимо удалять из создаваемого теста и существенно переделывать и

улучшать. Задания, для которых $0 < r_{xy} < 0,3$, являются кандидатами на удаление из теста. Задания, прошедшие отбор, включаются в тест, который может использоваться на этапе текущего контроля той же выборки обучаемых, которая участвовала в отборе тестовых заданий на этапе самоконтроля. Таким образом, созданный с применением классической теории тест будет наиболее адекватно оценивать именно ту группу обучаемых, для которой этот тест и предназначен.

- легкость задания:

$$F_{ac} = \frac{X}{X_{MAX}}, \quad (9)$$

где X – средняя оценка всех отвечавших на это задание, X_{MAX} – максимальная возможная оценка за это задание;

- дискриминация задания: коэффициент корреляции Пирсона между заданием и итоговым баллом или между заданием и итоговым баллом минус это задание. Дискриминация равна +1, когда те, кто хорошо отвечают на это задание, хорошо пишут тест; и равна -1, когда те, кто хорошо отвечают на это задание, плохо пишут тест;

- надежность (стабильность) теста: корреляция между баллами обучаемых по тесту в первый раз и во второй раз (test-retest);

- надежность теста (R Бакхауза):

$$R = (\lambda + 1) \frac{1 - \sum n_j S_j^2}{n S_X^2} - \lambda \frac{\sum (n_{j,t} m_{j,t} m_{t,j}) - n M^2}{n S_X^2}, \quad (10)$$

$$\lambda = \frac{\sum n_j}{\sum n_{j,t}}, \quad (11)$$

где k – число заданий, n – число экзаменуемых, n_j – число людей, отвечавших на задание j , $n_{j,t}$ – число людей, отвечавших на задания j и t , S_j – стандартное отклонение задания j , M – средний балл теста, m_j – средний балл по заданию j , $m_{j,t}$ – средний балл по заданию j тех, кто ответил на вопрос t .

- стандартная ошибка измерений:

$$SE = \frac{S_X}{\sqrt{1-r}}, \quad (12)$$

где r – надежность теста. Баллы 95% экзаменуемых находятся внутри $\bar{X} \pm 2SE$;

- индекс выбора (если анализируются вопросы по выбору):

$$C = \frac{N_i}{N} \quad (13)$$

показывает, насколько эффективно вопросы привлекают экзаменуемых.

Построение теста при помощи классической теории тестирования:

1. Исходя из теоретических предположений и при хорошем понимании того, что надо измерить, пишутся задания теста. Необходимо написать как минимум в два раза больше заданий, чем предполагается включить в окончательный вариант теста.
2. Калибровка: полученные задания нужно протестировать при помощи людей, близких к той популяции, для которой предназначается тест.
3. Следует отбросить задания с низкой ($<0,2$) или отрицательной корреляцией задание-итоговый балл.
4. Если дискриминация отрицательна, то, возможно, нарушилось предположение об однородности теста.
5. Выбираем нужное число заданий с самой высокой корреляцией задание-итоговый балл, так как они повышают надежность теста, снижая стандартную ошибку измерения.
6. Чтобы подобрать необходимую сложность теста, можно заменить часть заданий на более или менее сложные. Не желательно иметь легкость задания более 0,85 или менее 0,15. Для того, чтобы обучающиеся всех уровней подготовки могли быть оценены, лучше иметь задания разного уровня легкости (сложности).
7. Таким образом, получается тест, который будет наиболее адекватным образом оценивать именно ту группу обучаемых, для которых он предназначался.

При всех своих достоинствах, таких как простота и привычность, классическая теория тестирования имеет, тем не менее, и ряд недостатков [19]:

1. Сложно сравнить сходные особенности обследованных, выявленные при помощи разных методик.
2. Сложно оценить уровень выраженности тех или иных качеств измеряемых диагностической методикой в независимости от уровня проявленности этих качеств в группе испытуемых. Это происходит в результате того, что тест, созданный и обработанный при помощи классической теории тестирования будет наиболее адекватным образом оценивать именно ту группу обследуемых, для которой он предназначался, и, применяя его для обследования другой группы, мы будем сравнивать испытуемых не друг с

другом или определенным объективным показателем, а с испытуемыми из базовой группы.

3. Тестовые баллы по-разному достоверны для разных испытуемых: в экстремальных случаях оценки менее надежны (достоверны), чем в середине распределения.
4. В целом классическая теория тестирования недостаточна надежна и универсальна.

2.4.2 IRT - теория тестирования

Для решения проблем, возникающих при использовании классической теории тестирования, была разработана IRT - теория тестирования (Item Response Theory), которая предполагает, что получаемые при тестировании баллы испытуемых определяются ненаблюдаемыми переменными – латентными параметрами и имеет следующие достоинства:

1. Характеристики заданий не зависят от группы испытуемых, при помощи которой они были получены.
2. Оценки проявленности того или иного диагностического параметра у испытуемого не зависят от используемой диагностической методики.
3. Благодаря итерационной процедуре нахождения оценок параметров IRT учитывает степень значимости («вес») ответа на каждый вопрос методики для получения конечного балла.

Большинство существующих тестовых и опросниковых методик позволяют провести дифференциальную диагностику различных качеств испытуемых относительно некоего «среднего» уровня, взятого зачастую совершенно безотносительно к актуальной для исследователя группе испытуемых. Разумеется, при таком подходе, мы лишь приблизительно можем оценить выраженность требуемого качества испытуемого как «низкую», «среднюю» или «высокую». Для оценки же того, насколько эта выраженности низкая или высокая, как отличаются друг от друга испытуемые получившие экстремальные баллы, необходимо создание новых методик, что требует больших трудозатрат и продолжительного времени, в то время как задания существующих методик уже подобраны так, чтобы работать на повышение надежности теста.

Для решения поставленной выше задачи с использованием уже существующих методик можно использовать положительные и уникальные свойства Item Response Theory.

Возможность Item Response Theory «адаптироваться» к конкретной выборке в рамках одной популяции, позволяет не только проводить диагностику, например, личностных качеств, в некой случайной выборке, но и проводить тонкую диагностику в группах, члены которых обладают одинаковой (с точки зрения большинства используемых ныне методик)

выраженностью какого-либо качества сравнивая их не с абстрактным «средним показателем», а друг с другом [20].

Решению данной задачи способствует и такое свойство Item Response Theory как точечный, а не интервальный характер оценки.

Одним из недостатков традиционных тестов является то, что ответы на все вопросы имеют одинаковый вес в получении конечного бала. Тонкая же диагностика требует, чтобы ответу на каждый вопрос методики соответствовал индивидуальный бал («сложность» в педагогических тестах), учитывающий его вклад в конечную оценку. Возможность оценить каждый вопрос методики и есть еще одно свойство Item Response Theory, которое кардинально отличает ее от классической теории тестирования и значительно расширяет возможности создателей и пользователей тестов.

Таким образом, возможность учитывать «вклад» ответа на каждое задание (его «сложность») позволяет создавать методики, где ответ на каждый вопрос имел бы свой вес в определении результата.

Следующее направление использования Item Response Theory связано с оценкой (апробацией) новых или уже существующих диагностических методик для работы с различными группами испытуемых.

Апробация новых диагностических методик путем сравнения их с уже используемыми и замеряющими те же показатели, на одной популяции возможна, конечно, и при помощи классической теории тестирования, но Item Response Theory имеет математический аппарат, позволяющий решить эту проблему с большей достоверностью и надежностью.

Теория IRT предполагает, что и индивидов, и задания можно расположить на одной оси «способность – трудность» или «интенсивность свойства – сила пункта». Каждому испытуемому ставится в соответствие только одно значение латентного параметра («способности»).

В общем виде вероятность ответа зависит от множества свойств испытуемого, но в моделях IRT рассматривается лишь одномерный случай.

Главное отличие IRT от классической теории теста в том, что в ней не ставятся и не решаются фундаментальные проблемы эмпирической валидности и надежности теста: задача априорно соотносится лишь с одним свойством, т. е. тест заранее считается валидным. Вся процедура сводится к получению оценок параметров трудности задания и к измерению «способностей» испытуемых (образованию «характеристических кривых»).

В классической теории теста индивидуальный балл (уровень свойства) считается некоторым постоянным значением. В IRT латентный параметр трактуется как непрерывная переменная.

Первичной моделью в IRT стала модель латентной дистанции, предложенная Г. Рашем: разность уровня способности и трудности теста $x_i - \beta_j$,

где x_i – положение i -го испытуемого на шкале, а β_j – положение j -го задания на той же шкале. Расстояние $(x_i - \beta_j)$ характеризует отставание способности испытуемого от уровня сложности задания. Если разница велика и отрицательна, то задание не может быть выполнено, так как для данного испытуемого оно слишком сложно. Если же разница велика и положительна, то задание также не информативно, ибо испытуемый заведомо легко и правильно его решит.

Вероятность правильного решения задания (или ответа «да») i -м испытуемым:

$$P_i(x_{ij}) = f(x_i - \beta_j) \quad (14)$$

Вероятность выполнения j -го задания группой испытуемых:

$$P_j = (x - \beta_j) \quad (15)$$

В IRT функции x и f называются функциями выбора пункта. Соответственно первая является характеристической функцией испытуемого, а вторая – характеристической функцией задания.

Считается, что латентные переменные x и β нормально распределены, поэтому для характеристических функций выбирают либо логистическую функцию, либо интегральную функцию нормированного нормального распределения (они мало отличаются друг от друга).

Поскольку логистическую функцию проще аналитически задавать, ее используют чаще, чем функцию нормального распределения.

Кроме «свойства» и «силы пункта» (она же – трудность задания) в аналитическую модель IRT могут включаться и другие переменные. Все варианты IRT классифицируются по числу используемых в них переменных.

Наиболее известны однопараметрическая модель Г. Раша, двухпараметрическая модель А. Бирнбаума и его же трехпараметрическая модель.

Вероятность того, что i -тый испытуемый выполнит j -тое задание в случае использования модели Раша имеет вид:

$$P_{ij} = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \quad (16)$$

Вероятность успеха зависит только от разницы между уровнем способности и сложностью задания.

Модель Раша не позволяет заданиям различаться по дискриминации: задания можно расположить только по уровню их сложности.

Достоинства модели Раша по сравнению с другими моделями IRT:

1. Простота.
2. Для оценивания параметров не требуется никаких предположений и достаточно только сырых данных. В других моделях IRT необходимы дополнительные ограничения для контроля взаимодействия параметров.
3. Минимальные полезные данные: 4 задания для 10 человек. Для других моделей IRT необходимо не менее 1000 человек.
4. Данные для стабильных оценок: 20 заданий для 200 человек. Для других моделей IRT: не существует.

Уравнение для двухпараметрической модели Бирбаума:

$$P_{ij} = \frac{\exp \alpha_j (\theta_i - \beta_j)}{1 + \exp \alpha_j (\theta_i - \beta_j)}, \quad (17)$$

где α_j – различающая способность (дискриминация задания).

Трехпараметрическая модель Бирбаума (Birbaum, 3PL) для дихотомических данных:

$$P_{ij} = C_j + (1 - C_j) \frac{\exp \alpha_j (\theta_i - \beta_j)}{1 + \exp \alpha_j (\theta_i - \beta_j)}, \quad (18)$$

где α_j – различающая способность (дискриминация задания), C_j – вероятность угадывания правильного ответа на j-е задание. Иногда также C_j называют уровнем псевдо-успеха, то есть вероятностью ответить на задание правильно для экзаменуемых с минимальной способностью.

Для измерения «силы пункта» (трудность задания) и «величины свойства» (способность испытуемого) в одной шкале Г.Рашем была введена единица измерения «логит».

Уровень «способности» испытуемого в «логитах» определяется на шкале интервалов с помощью формулы:

$$\theta_i^0 = \ln \frac{P_i}{q_i}, \quad (19)$$

где $i = 1 \div n$, n – число испытуемых, P_i – доля правильных ответов i-го испытуемого на задания теста, q_i – доля неправильных ответов, а также:

$$p_i + q_i = 1 \quad (20)$$

Для первичного определения трудности задания в логитах используют оценку

$$\beta_j^0 = \ln \frac{q_j}{p_j}, \quad (21)$$

где $j = 1 \div n$, n - число заданий, q_j - доля неправильных ответов, p_j - доля правильных ответов для испытуемых группы на j -е задание, а также:

$$p_j + q_j = 1 \quad (22)$$

Параметры β и θ изменяются от «плюса» до «минуса», при этом, если $\beta < -6$, это указывает на то, что на эти задания практически каждый испытуемый дает правильный («ключевой») ответ. При $\beta > 6$ с заданием не сможет справиться ни один испытуемый, точнее - вероятность дать «ключевой» ответ ничтожна.

Рекомендуется рассматривать лишь интервалы от -3 до +3 как для β (трудности), так и для θ (способности).

Второй этап шкалирования испытуемых и заданий сводится к тому, что шкалы преобразуются в единую шкалу путем «уничтожения» влияния трудности задания на результат индивидов. И к тому же элиминируется влияние индивидуальных способностей на решение заданий различной трудности.

Для шкалы испытуемых:

$$\theta_i = \beta + x\theta_i^0, \quad (23)$$

где $i = 1 \div n$,

$$x = \sqrt{1 + \frac{W^2}{2.89}}, \quad (24)$$

β – среднее значение логитов трудности заданий теста, W – стандартное отклонение распределения начальных значений параметра β , n – число испытуемых.

Для шкалы заданий:

$$\beta_i = \theta + y\beta_j^0, \quad (25)$$

где $j = 1 \div n$,

$$y = 1 + \frac{V^2}{2.89}, \quad (26)$$

θ – среднее значение логитов уровней способностей, V – стандартное отклонение распределения начальных значений «способности», n – число заданий в тесте.

Эти эмпирические оценки используются в качестве окончательных характеристик измеряемого свойства и самого измерительного инструмента (заданий теста).

Если перед исследователем стоит задача конструирования теста, то он приступает к получению характеристических кривых заданий теста. Характеристические кривые могут накладываться одна на другую. В этом случае избыточные задания выбраковываются. На определенных участках оси θ («способность») характеристические кривые заданий могут вовсе отсутствовать. Тогда разработчик теста должен добавить задания недостающей трудности, чтобы равномерно заполнить ими весь интервал шкалы логитов от -6 до +6. Заданий средней трудности должно быть больше, чем на «краях» распределения, чтобы тест обладал необходимой дифференцирующей (различающей) силой.

Вся процедура эмпирической проверки теста повторяется несколько раз, пока разработчик не останется доволен результатом работы. Естественно, чем больше заданий, различающихся по уровню трудности, предложил разработчик для первичного варианта теста, тем меньше итераций он будет проводить.

2.4.3 Сравнительный анализ IRT и классической теории тестирования

Другой, более гибкий подход основан на взаимодействии этих теорий. Основная цель состоит в разработке математической модели процесса тестирования, параметрами которой, подлежащими определению, служат различные характеристики участников тестирования и самого теста. В настоящее время Item Response Theory представляет собой довольно

обширную и, к сожалению, непростую теорию, активно использующую разнообразный арсенал математических методов. Поэтому зачастую целесообразно бывает использовать сочетание двух упомянутых теорий. С помощью статистических методов классической теории проводится первичный анализ качественных характеристик полученного варианта теста, а с помощью Item Response Theory, более углублённый анализ характеристик тестовых заданий. Наиболее эффективным использование IRT становится при сформированном банке тестовых заданий, что обеспечивает возможность автоматической генерации тестов, с определёнными наперёд заданными и научно-обоснованными характеристиками точности и надёжности.

Независимо от выбранной концепции, существенным моментом в тестологии является начальный этап конструирования, на котором происходит формирование целей создания тестов. Этап целеполагания является наиболее трудным и вместе с тем наиболее важным: от результатов его выполнения зависит в первую очередь качество содержания теста или содержательная валидность, являющаяся базовым качественным критерием тестов. Оценка тестовых заданий на пригодность к достижению поставленных целей всего теста, реализуется с помощью экспертов [1].

Таким образом, на данный момент создание качественных тестов реализуется в рамках двух базовых итерационных подходов: классический подход обеспечивает начальный этап создания, а также оценки качества тестов, и современная Item Response Theory, позволяющая получить устойчивые оценки латентных параметров испытуемых и трудности тестовых заданий и поддерживающий конструирование тестов на продвинутом этапе, причём применение математического аппарата обеих этих теорий целесообразно только при грамотно сформированной таксономии целей тестового контроля и удовлетворяющих ей тестовых заданиях [19]. Ниже кратко изложены основные возможности создания тестов в рамках классической и современной теории тестирования, см. таблицу 1.

Таблица 1 – Основные положения классической и IRT теории

Классическая теория	Item Response Theory
Основная задача теста: Получить значение истинного балла (T) испытуемого исходя из наблюдаемого результата (X), с учётом случайной ошибки измерения (E), откуда вытекает основной постулат классической теории тестирования:	Основная задача теста: Получить устойчивую объективную оценку латентного параметра уровня знаний испытуемого по исследуемому предмету, не зависящую от конкретного теста. Возможности: 1) Установление связи между

$X_i = T_i + E_i$ Возможности: Извлечь первичную информацию о тесте, на основании матрицы результатов апробационного тестирования, а именно: 1) Оценить статистическую сложность заданий. 2) Получить интеркорреляцию между заданиями теста и корреляцию баллов заданий и внешнего критерия (суммы баллов испытуемых), для определения валидности тестовых заданий. 3) Быстро оценить качества теста на основе графического вида кривой распределения тестовых баллов испытуемых. 4) Получить оценку надёжности результатов тестирования посредством корреляционного анализа баллов испытуемых по тесту, либо по нескольким его вариантам. 5) Построить доверительный интервал, в пределах которого находится истинный балл испытуемого, либо получить точечную регрессионную оценку. 6) Интерпретировать результаты в терминах Z-шкалы, либо процентильной шкалы, то есть определить место (рейтинг) испытуемого в выборке.	латентными параметрами испытуемых и наблюдаемыми результатами выполнения теста. Наблюдаемые результаты выполнения теста порождаются взаимодействием двух множеств латентных параметров теста: уровнем знаний испытуемых и трудности заданий. 2) Параметры уровня знаний испытуемого и трудности заданий теста отображаются в единую шкалу логитов, что позволяет реализовать идею адаптивного тестирования, когда для каждого испытуемого (с конкретным уровнем знаний) отбираются задания определённой сложности. 3) Существует возможность помимо стандартных критериев качества, ввести новый критерий - эффективность теста, путём введения особого класса информационных функций, посредством которых происходит оценка количества информации, обеспечиваемое j-м заданием при уровне знаний θ_i. 4) Реально только в рамках Item Response Theory реализована возможность применения не только рейтинговой, но и интервальной шкалы, а это значит, уровень подготовки можно оценить количественно.
---	--

3 РЕАЛИЗАЦИЯ СИСТЕМЫ ТЕСТИРОВАНИЯ УРОВНЯ ЗНАНИЙ СТУДЕНТОВ

3.1 Модель системы тестирования уровня знаний студентов

Процесс организации и проведения тестирования знаний студентов ИнеУ кратко можно представить в виде следующей последовательности действий (рисунок 2):

- 1) заполнение базы тестовых заданий по разделу указанного предмета;
- 2) формирование базового теста на установление сложности заданий;
- 3) проведение базового тестирования с испытуемой группой;
- 4) анализ полученных результатов, расчет сложностей тестовых заданий с помощью IRT-теории и присвоение этих значений тестовым заданиям;
- 5) формирование адаптивного теста для определения уровня знаний студентов;
- 6) проведение адаптивного тестирования;
- 7) установление нижних границ уровня сложности задания для каждой отметки на шкале оценок;
- 8) расчет среднего уровня сложности тестовых заданий, с которыми справился студент;
- 9) проецирование полученных средних результатов на шкалу оценок.

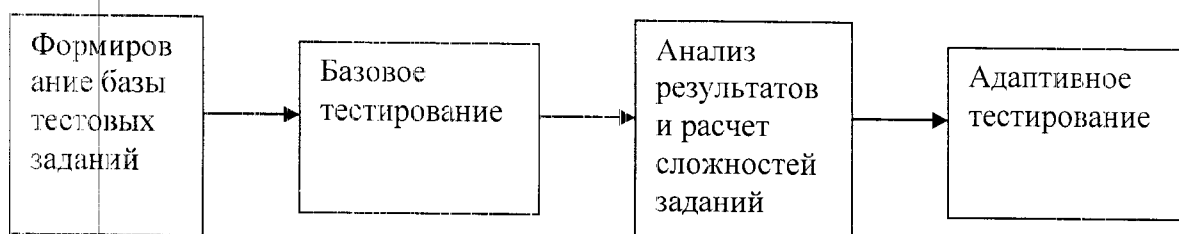


Рисунок 2 - Обобщенная схема процесса организации и проведения тестирования знаний студентов ИнеУ.

Особый интерес представляет модель адаптивного тестирования (рисунок3).

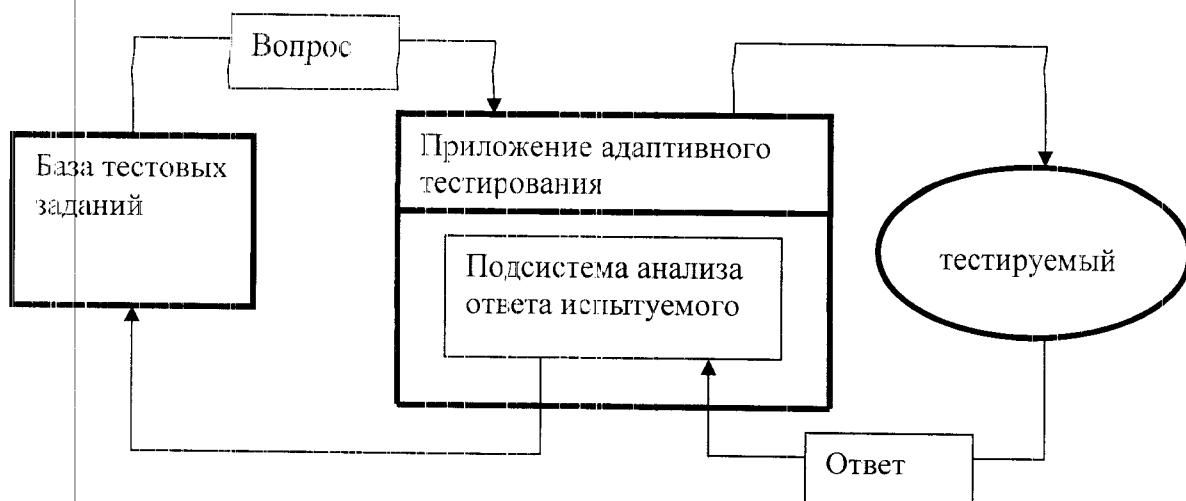


Рисунок 3 – Модель адаптивного тестирования

Испытуемому предоставляется вопрос среднего уровня сложности. Ответ поступает в подсистему анализа ответов, которая рассчитывает сложность следующего задания и посылает запрос СУБД для выборки вопроса необходимой сложности. В случае успешного ответа подсистема анализа ответов увеличивает уровень сложности последующего вопроса, в ином случае - уменьшает. Таким образом, студент выберет для себя определенный уровень сложности, вокруг которого и будет размываться сложность заданий.

3.2 Структура базы данных разрабатываемого тестового комплекса

В современном мире сложно найти такую сферу деятельности человека, где бы ни использовались базы данных, которые позволяют организовать эффективный и удобный сбор, хранение и обработку информации различного рода.

База данных — это набор сведений, относящихся к определенной теме или задаче. Существуют различные модели баз данных. Наибольшее распространение получила реляционная модель. Она лежит в основе таких баз данных, как Paradox, FoxPro, Access, Microsoft SQL Server, Oracle, InterBase и многих других. Достоинствами реляционной модели данных являются простота, гибкость структуры и удобство реализации. Реляционная база данных представляет собой совокупность таблиц (отношений), содержащих данные.

На рисунке 4 представлена схема базы данных автоматизировано системы контроля знаний с применением методов искусственного интеллекта.

Имеющийся в составе Delphi компилятор с языка ObjectPascal является одним из самых производительных в мире. В Delphi компиляция производится непосредственно в родной машинный код, в то время как существуют компиляторы, превращающие программу в так называемый р-код, который затем интерпретируется виртуальной р-машиной. Это не может не сказаться на фактическом быстродействии готового приложения.

Визуальная технология разработки программ – позволяет быстро создавать приложения путём размещения в форме стандартных компонентов. При этом соответствующий код программы автоматически генерируется Delphi. Такая технология освобождает разработчика от рутинной работы по созданию пользовательского интерфейса и позволяет уделить больше внимания внутренней организации данных и обработке данных.

Одним из основных критериев, при выборе инструмента разработки приложений баз данных является масштабируемость возможность работать с данными в различных платформах. Масштабируемость на практике - одно и то же приложение можно использовать как для локального, так и для более серьезного клиент-серверного вариантов. Поддержка баз данных в Delphi осуществляется с помощью набора драйверов соединений с SQL-северами Borland SQL Links for Windows, которые позволяют интегрированному в Delphi ядру процессора баз данных Borland, (BDE) Borland Database Engine, получать доступ к локальным базам данных Paradox, dBASE, Access, FoxPro, а также SQL-серверам InterBase, Informix, Oracle, Sybase, DB2, Microsoft SQL Server.

3.4 Язык SQL как стандартный язык баз данных

Определение реляционной системы требует, чтобы весь диалог с базой данных велся на едином языке - иногда его называют общим подязыком данных. На сегодняшний день единственным стандартным языком для работы с реляционными базами данных является структурированный язык запросов (Structured Query Language – SQL). На сегодняшний день SQL поддерживают свыше ста СУБД, работающих как на персональных компьютерах, так и на больших ЭВМ. Как следует из названия, SQL является языком программирования, который применяется для организации взаимодействия пользователя с базой данных. Если пользователю необходимо прочитать данные из базы данных, он запрашивает их у СУБД с помощью SQL. Как показано на рисунке 5 СУБД обрабатывает запрос, находит требуемые данные и посылает их пользователю. Процесс запрашивания данных и получения результата называется запросом к базе данных: отсюда и название - структурированный язык запросов.

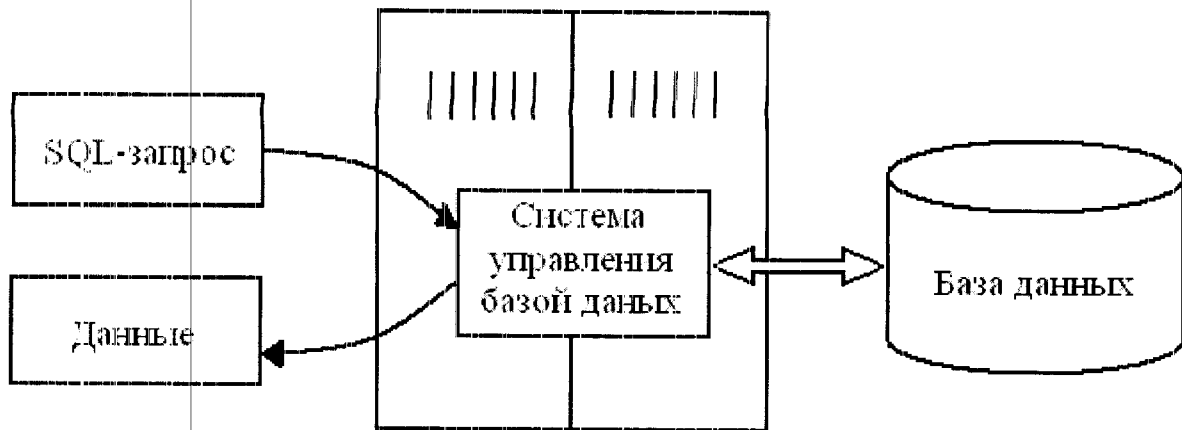


Рисунок 5 – Применение SQL-запроса для доступа к базе данных.

Запросы являются наилучшим способом отбора данных, необходимых для решения определённой задачи. Одно из преимуществ запросов является то, что они позволяют достаточно быстро отобрать необходимые данные из нескольких связанных таблиц. Имеется две разновидности операций по манипуляции с данными - выборка данных и модификация данных. Выборка - это поиск необходимых данных, а модификация включает такие операции как добавление, удаление или изменение данных.

Причиной того, что SQL стал стандартным инструментом для управления данными на персональных компьютерах, мини-компьютерах и больших ЭВМ, можно назвать совокупность следующих особенностей этого языка:

- независимость от конкретных СУБД;
- переносимость с одной вычислительной системы на другую;
- протокол ODBC - программный интерфейс, основанный на SQL;
- реляционная основа;
- высокоуровневая структура, напоминающая английский язык;
- обеспечение программного доступа к базам данных;
- возможность различного представления данных;
- полноценность как языка, предназначенного для работы с базами данных;
- возможность динамического определения данных;
- поддержка архитектуры клиент/сервер.

3.5 Программная реализация системы тестирования уровня знаний студентов

Разработанная система контроля знаний состоит из трех приложений:

- Организация и обработка результатов тестирования
- Базовое тестирование
- Адаптивное тестирование

3.5.1 Приложение «Организация и обработка результатов тестирования»

Приложение «Организация и обработка результатов тестирования» является рабочим местом преподавателя.

Вход в автоматизированную систему контроля знаний осуществляется после авторизации пользователя, имеющего право на изменение данных (рисунок 6).

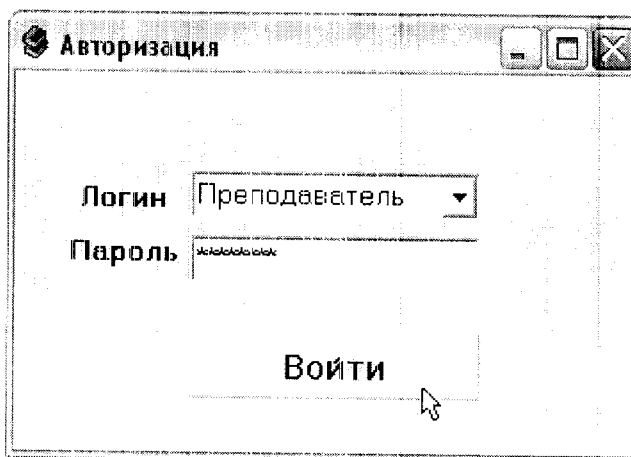


Рисунок 6 – Авторизация пользователя

При успешной авторизации открывается главное окно, содержащее основные функции приложения (рисунок 7), к которым относятся:

- редактирование базы тестовых заданий;
- создание базового теста;
- обработка результатов базового тестирования;
- создание адаптивного теста;
- обработка результатов адаптивного тестирования.

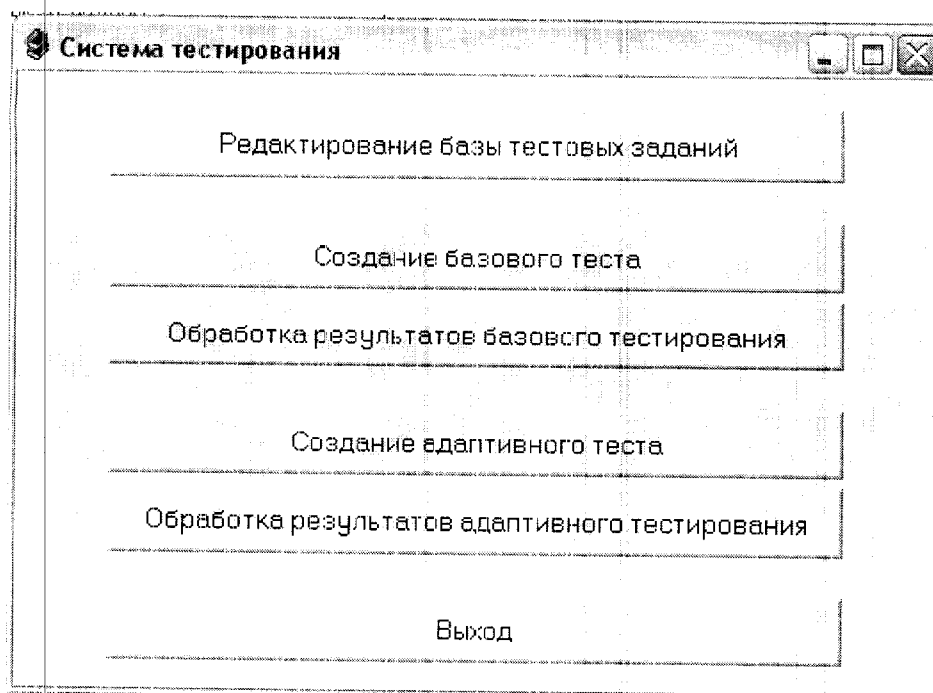


Рисунок 7 – Основные функции приложения «Организация и обработка результатов тестирования»

При выборе функции «Редактирование базы тестовых заданий» открывается окно, где отображаются имеющиеся в базе предметы и разделы, по которым составлены или будут составлены тесты (рисунок 8).

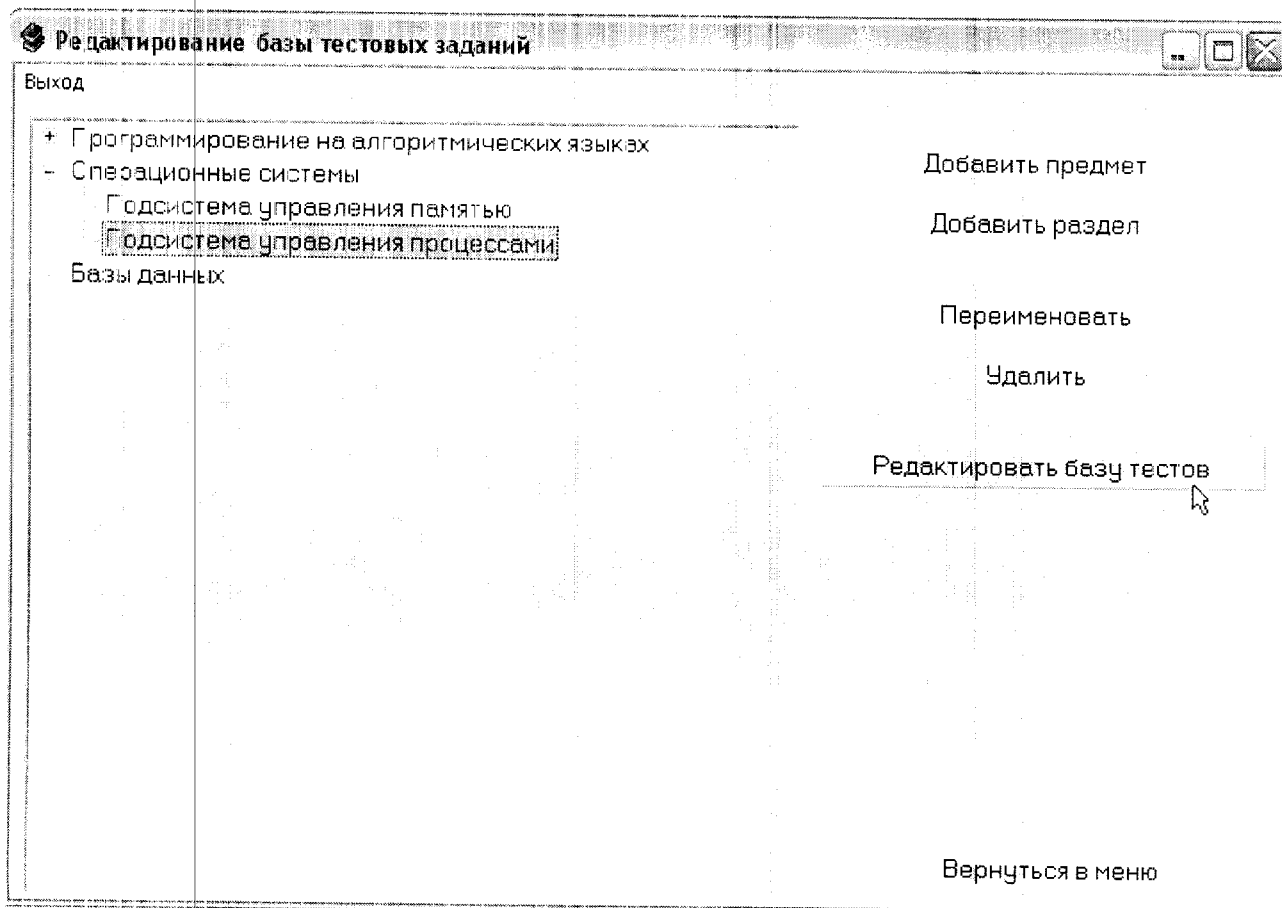


Рисунок 8 – Окно «Редактирование базы тестовых заданий»

В данном окне преподавателю предоставляется возможность следующих действий:

- добавление нового предмета;
- добавление нового раздела в текущем предмете;
- редактирование текущего предмета (или раздела);
- удаление текущего предмета (или раздела);
- редактирование базы тестов.

Для добавления нового предмета (раздела) необходимо кликнуть мышью на кнопку с названием «Добавить предмет» («Добавить раздел»). После чего появится диалоговое окно с полем ввода, где необходимо указать название предмета (раздела) (рисунок 9). Новый элемент автоматически появится в дереве списка предметов и разделов.

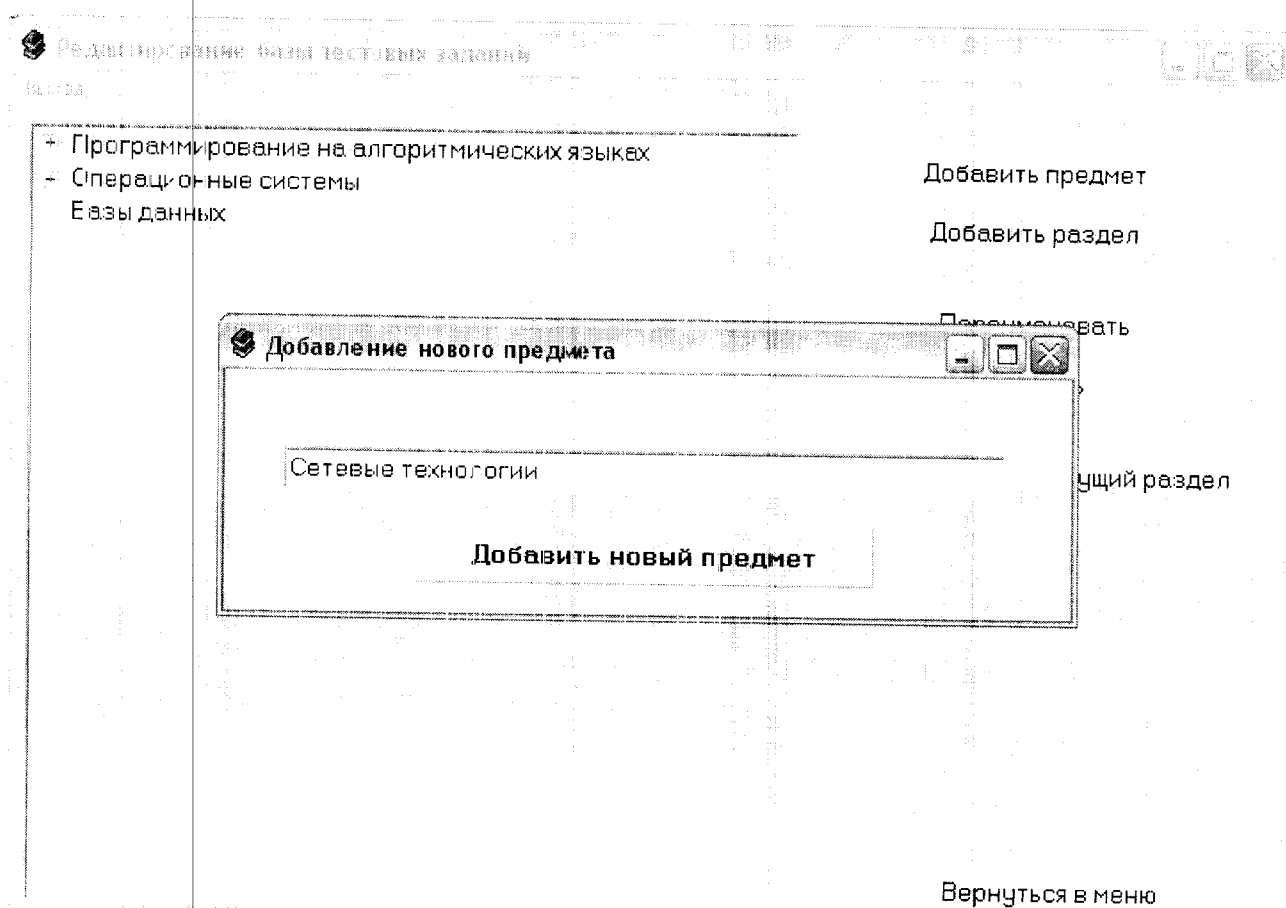


Рисунок 9 - Добавление нового предмета

Переименование и удаление текущего предмета (раздела) осуществляется аналогичным способом – нажатием на одноименную кнопку (рисунок 10).

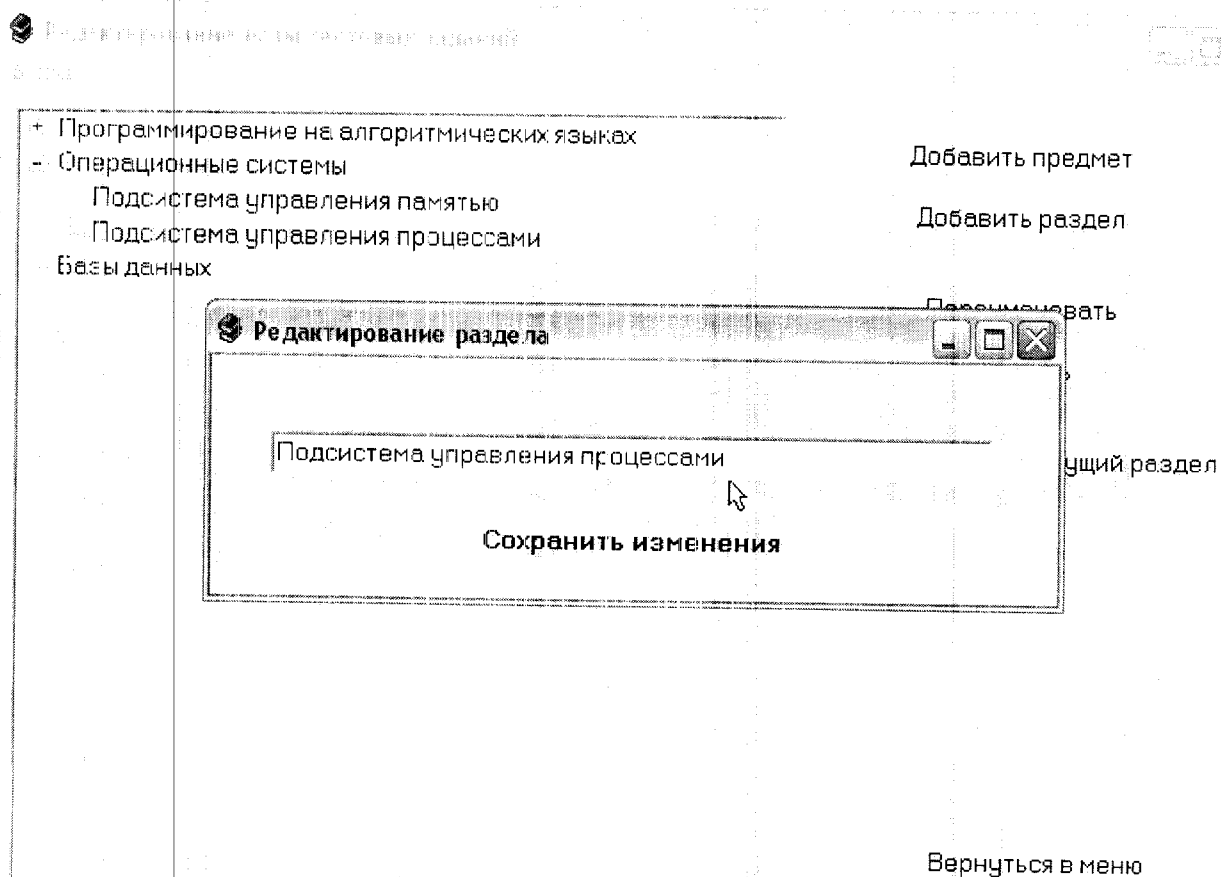


Рисунок 10 - Редактирование предмета (раздела)

Форма для редактирования базы тестовых заданий открывается нажатием кнопки «Редактировать базу тестов». В появившемся окне отобразится таблица со списком вопросов, относящихся к данному разделу, и соответствующих им заданными и расчетными сложностями, а также кнопки, выполняющие основные операции с базой тестовых вопросов (рисунок 11):

- добавление вопроса;
- редактирование текущего вопроса;
- удаление текущего вопроса.

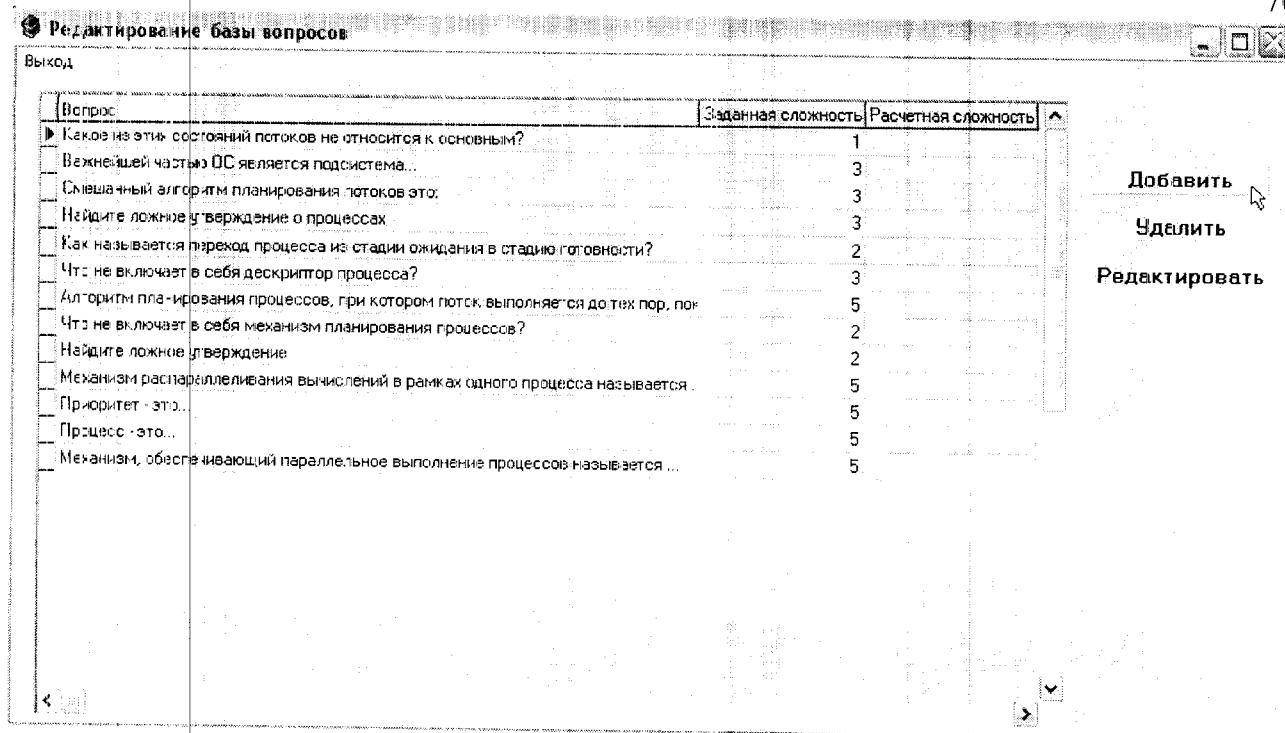


Рисунок 11 - Окно «Редактирование базы вопросов»

Интерфейс программы при создании нового тестового задания представлен на рисунке 12. После ввода вопроса тестового задания необходимо указать возможные варианты ответов, при этом правильным принимается всегда ответ №1. Количество ответов может варьироваться от 2 до 10. В поле ввода «Сложность» необходимо ввести значение сложности задания по умолчанию. После нажатия кнопки «Сохранить» отредактированный вопрос будет сохранен в базу тестовых заданий.

Редактирование вопроса

Вопрос:
Приоритет - это...

Кол-во ответов Сложность **Сохранить** **Отмена**

Ответ №1 (правильный)
число, характеризующее степень привилегированности потоков, при использовании ресурсов вычислительной машины

Ответ №2
динамический объект, возникающий в ОС после того, как пользователь или ОС решает запустить программу на выполнение

Ответ №3
статистический объект, представляющий собой файл с кодом и данными

Ответ №4
максимальное количество потоков, обрабатываемых процессором

Ответ №5
прерывание процесса при возникновении какой-либо ситуации

Рисунок 12 – Редактирование вопроса

Для создания адаптивного теста, преподавателю необходимо (рисунок 13):

- заполнить поле «Тема адаптивного теста»;
- выбрать из выпадающих списков предмет и раздел контролируемых знаний;
- указать количество вопросов, предлагаемых испытуемому;
- наложить ограничение по времени (в минутах) на прохождение студентом тестирования.

Создание адаптивного теста

Тема адаптивного теста: Тест 1 "Подсистема управления процессами"

Предмет: Операционные системы

Раздел: Подсистема управления процессами

Количество вопросов: 30

Время на тест (мин): 60

Сохранить Отмена

Рисунок 13 – Создание адаптивного теста

Выбор теста для анализа осуществляется в окне «Выбор темы базового теста» выбором текущей записи, а затем двойным кликом мыши или нажатием кнопки «Выбрать» (рисунок 14).

Выбор темы базового теста

Тема теста	Раздел	Время(мин)
Тест "Программирование на языке Pascal"	Pascal	40
Тест "Подсистема управления процессами"	Подсистема управления процессами	60

Выбрать

Рисунок 14 - Выбор темы базового теста

Результаты проведенного тестирования отображаются в окне «Сводная таблица результатов» (рисунок 15). Преподавателю предоставляется информация на двух вкладках.

Вкладка «Сложность тестовых заданий» выводит следующие данные:

- задание теста;
- доля студентов, не ответивших на данный вопрос;
- рассчитанная с помощью IRT-теории сложность тестового задания.

Список тестовых заданий для удобства представлен в порядке убывания их трудностей. При нажатии кнопки «Записать расчетную сложность» уровень сложностей заданий будет зафиксирован в базе для дальнейшего использования в адаптивном тестировании.

Сводная таблица результатов			
Сложность тестовых заданий		Результаты студентов	
№	Сложность	Доля	Задания теста
1	6	0/12	Механизм распараллеливания вычислений в рамках одного процесса назыв
2	3.08	1/12	Найдите ложное утверждение о процессах
3	1.92	2/12	Механизм, обеспечивающий параллельное выполнение процессов называет
4	1.17	3/12	Приоритет - это...
5	1.17	3/12	Что не включает в себя дескриптор процесса?
6	0.58	4/12	Механизм распараллеливания вычислений в рамках одного процесса назыв
7	0.58	4/12	Важнейшей частью ОС является подсистема...
8	0.05	5/12	Как называется переход процесса из стадии ожидания в стадию готовности?
9	0.05	5/12	Смешанный алгоритм планирования потоков это:
10	-0.44	6/12	Что не включает в себя механизм планирования процессов?
11	-0.93	7/12	Процесс - это...
12	-1.46	8/12	Найдите ложное утверждение
13	-2.05	9/12	Алгоритм планирования процессов, при котором поток выполняется до тех п
14	-2.8	10/12	Каков из этих состояний потоков не относится к основным?

Записать расчетную сложность Отобразить график

Рисунок 15 – Сводная таблица результатов базового тестирования

На данном этапе преподавателю необходимо проанализировать полученные результаты.

Необходимо исключить слишком трудные или слишком легкие задания, то есть те тестовые вопросы, сложность которых составляет $\geq +6$ (очень трудные) и ≤ -6 (очень легкие).

Далее равномерно наполнить базу заданиями таким образом, чтобы сложности равномерно увеличивались от -6 до +6. Наглядную информацию о том, равномерно ли распределены трудности заданий, можно получить с помощью графика, на котором в порядке убывания по оси абсцисс отложены номера вопросов, а по оси ординат уровни трудностей заданий.

Как видно на рисунке 16, трудности тестовых заданий по данному разделу достаточно равномерно распределены, за исключением первого

вопроса (его сложность составляет +6 логит). Данное тестовое задание следует исключить из базы.

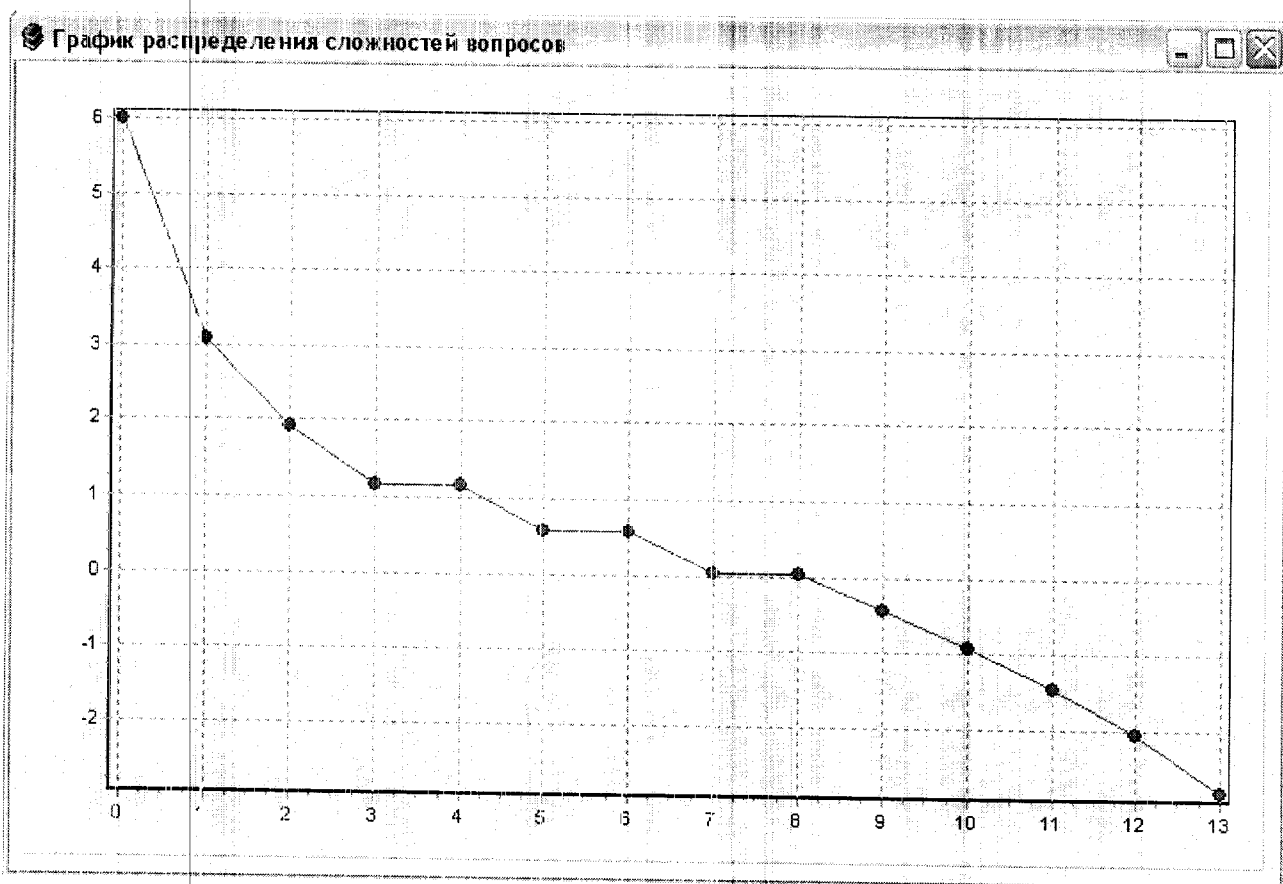


Рисунок 16 - График распределения сложностей вопросов

3.5.2 Приложение «Базовое тестирование»

Приложение «Базовое тестирование» проводит опрос отобранной группы по тестовым заданиям определенного раздела для расчета сложностей этих заданий. Как и в предыдущем приложении, вход в систему осуществляется после авторизации пользователя (рисунок 17).

Базовое тестирование

Логин
Студент

Пароль
[маска]

Войти

Рисунок 17 – Авторизация пользователя

После входа в систему, студенту необходимо выбрать тему теста, по которой будет производиться опрос, ввести свою фамилию, имя и отчество, а также указать группу и нажать кнопку «Начать тестирование».

Рисунок 18 – Базовое тестирование

Основное окно базового теста (рисунок 19), содержит следующую информацию:

- вопрос тестового задания;
- варианты ответов;
- номер текущего вопроса из общего количества;
- время, оставшееся на прохождение теста.

Тестовые вопросы предоставляются тестируемому в разном порядке, последовательность вариантов ответов также является случайной. С помощью кнопок «Предыдущий» и «Следующий» опрашиваемый может осуществлять прямую навигацию по вопросам теста, для быстрого перемещения по тестовым заданиям необходимо указать номер требуемого вопроса в соответствующем поле. В течение тестирования испытуемый может менять вариант правильного ответа. Выбранные студентом варианты ответов будут зафиксированы в базе после нажатия кнопки «Завершить тест». Все неотвеченные тестовые задания считаются как неправильные варианты ответов.

Базовое тестирование на установление сложности задания

Выход

Завершить тест Осталось 0:0:26 Вопрос № 12 из 14 Предыдущий Следующий

Приоритет - это...

☐ Ответ №1
прекращение процесса при возникновении какой-либо ситуации

☒ Ответ №2
число, характеризующее степень привилегированности потоков, при использовании ресурсов вычислительной машины

☐ Ответ №3
максимальное количество потоков, обрабатываемых процессором

☐ Ответ №4
динамический объект, возникающий в ОС после того, как пользователь или ОС решает запустить программу на выполнение

☐ Ответ №5
статистический объект, представляющий собой файл с кодом и данными

Рисунок 19 – Основное окно тестирования на установление сложностей заданий

3.5.3 Приложение «Адаптивное тестирование»

Приложение «Адаптивное тестирование» проводит опрос студентов по тестовым заданиям, сложность которых была установлена после анализа результатов базового тестирования. Вход в систему осуществляется после авторизации пользователя. Как и в случае базового тестирования, студенту необходимо выбрать тему теста и указать свои данные – фамилию, имя, отчество и группу.

Прохождение адаптивного теста в корне отличается от базового тестирования. Здесь нет заранее заданной последовательности вопросов. После нажатия кнопки «принять ответ», студенту будет предоставлен следующий вопрос в зависимости от того, правильно ли ответил он на предыдущее тестовое задание (рисунок 20). Таким образом, студенту будут предоставляться вопросы примерно той сложности, на которую он способен ответить.

Адаптивное тестирование

Завершить тест Вопрос №2 из 30 Осталось времени: 0:59:11 **Принять ответ**

Найдите ложное утверждение:

☐ Ответ №1
Поток это расщепление процесса на отдельные параллельные ветви

☐ Ответ №2
Диспетчеризация это реализация результатов планирования процессов

☐ Ответ №3
Диспетчеризация включает в себя сохранение контекста текущего потока

☐ Ответ №4
Диспетчеризация включает в себя переключение потоков

☒ Ответ №5
Ресурсами владеют потоки, а процессы лишь их наследуют

Рисунок 20 – Адаптивное тестирование

ЗАКЛЮЧЕНИЕ

Объектом исследований в рамках диссертационной работы были выделены тестовые системы контроля знаний студентов ВУЗов.

Следует отметить, что применение тестирования существенно повышает требования к уровню квалификации преподавателя. Умение в форме тестов ставить сложные вопросы - проблема трудоемкая, требующая от преподавателя более глубоких знаний предметной области преподаваемой дисциплины, оно вынуждает его постоянно работать над собой.

Тем не менее, следует отметить целый ряд его педагогических, организационных, экономических и прочих достоинств.

Во-первых, компьютерное тестирование открывает возможность объективного, систематического и подлинно индивидуального контроля знаний студентов, позволяя решать многие педагогические задачи, о которых при традиционной методике не может быть даже и речи. Так, например, оно позволяет иметь надежную обратную связь на всех этапах обучения, выявлять лидеров по признаку успеваемости, обнаруживать массовые и индивидуальные пробелы в знаниях, своевременно корректировать содержание и форму учебного процесса и т.д.

В-вторых, оно отвечает требованию информатизации учебного процесса, без которой немислимо дальнейшее существование высшей школы.

Однако, традиционное тестирование не вполне отвечает современным требованиям к оценке уровня знаний студентов, что требует применения интеллектуальных методов тестирования, позволяющих приблизиться тестовому опросу к традиционному экзамену.

В ходе диссертационной работы были проанализированы современные модели и метода интеллектуального тестирования, а также теории обработки данных тестирования.

Результатом данной работы явилась разработка новой методики тестирования знаний учащихся, состоящая из следующих этапов:

- проведение базового тестирования для расчета сложностей вопросов;
- расчет уровней трудностей заданий с помощью IRT-теории;
- проведение адаптивного тестового контроля знаний.

Предложенная методика была реализована в виде программного тестового комплекса. Разработанное программное обеспечение может быть использовано в качестве системы определения уровня знаний студентов ИНЕУ на различных этапах обучения, что повлечет за собой появление возможности выявить слабые стороны процесса обучения и повышение качества контроля усвоения знаний учащимися ИНЕУ.

Новые информационные технологии позволяют в корне изменить и контроль усвоения учебных материалов студентами. Речь идет о дополнении

традиционных форм контроля знаний методами компьютерного тестирования. Наша точка зрения состоит в том, что компьютерное тестирование не предполагает отказа от традиционных форм контроля, а является серьезным дополнением к ним.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Морев И.А. Образовательные информационные технологии. Часть 2. Педагогические измерения. Учебное пособие. Владивосток: Изд. Дальневост. ун-та, 2004 - 174 с.
2. Переверзев В.Ю. Критериально - ориентированные педагогические тесты для итоговой аттестации студентов. М.: Изд. НМЦ СПО Минобразования РФ, 1998 -152 с.
3. Моисеев В.Б., Усманов В.В., Таранцева К.Р., Пятирублевый Л.Г. Оценивание результатов тестирования на основе экспертно-аналитических методов // Открытое образование. 2001. №3.- с.32-36.
4. Васильев В.И., Глухов В.В., Тягунова Т.Н. Культура компьютерного тестирования. Ч. 5. Оптимальная оценка уровня учебных достижений тестируемых. М.: МГУП, 2002 - 75 с.
5. Овчинников В.В. Оценивание учебных достижений учащихся при проведении централизованного тестирования. – М.: Изд. Век книги, 2001 -27 с.
6. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе, Учебное пособие, М.: Исследовательский центр, 2002 - 167 С.
7. Пак Н.И., Симонова А.Л. Методика составления тестовых заданий // ИНФО, 1998. № 5.
8. Люсин Д.В. Основы разработки и применения критериально-ориентированных педагогических тестов. М.: Исследовательский центр, 1993 - 51 С.
9. Шалимов П.Ю., Попоков В.И. Технология рейтинговых исследований качества образования с применением нейронных сетей // ИНФО, 2001, № 5
10. Пак Н.И., Филиппов В.В. О технологии создания компьютерных тестов // ИНФО, 1999, № 5.
11. Моисеев В.Б., Пятирублевый Л.Г., Таранцева К.Р. Распознавание образа обучаемых по уровням их знаний в компьютерном тестировании. Сборник материалов Интернет-конференции «Проблемы перехода классических университетов в систему открытого образования». Москва: МЭСИ, 2001- с. 131-137.
12. Сысоева Л.А., Толстоусова В.Г. Предметно-критериальная методика составления тестов. Конгресс конференция ИТО-2003 «Информационные технологии в образовании», 2003.
13. Пятирублевый Л.Г., Моисеев В.Б., Таранцева К.Р. Применение информационно-генетических алгоритмов в процедурах

- образовательного тестирования. Конгресс конференция ИТО-2003 «Информационные технологии в образовании», 2003.
14. Моисеев В.Б., Пятирублевый Л.Г., Таранцева К.Р. Информационный подход к выбору решений в системах адаптивного тестирования. Материалы конференции «Анализ качества образования и тестирование». Москва, МО РФ, МЭСИ, 2001.
 15. Маслак А.А., Анисимова Т.С., Осипов С.А. Исследование точности модели Раша на основе имитационного моделирования. Конгресс конференция ИТО-2003 «Информационные технологии в образовании», 2003.
 16. Крылов Ю.Н. Абсолютная временная шкала измерения знаний. динамика результатов тестирования во времени. Конгресс конференция ИТО-2003 «Информационные технологии в образовании», 2003.
 17. Журавлев В.Б. Методика статистического анализа учебного процесса. Конгресс конференция ИТО-2003 «Информационные технологии в образовании», 2003.
 18. Челышкова М.Б. Теория и практика конструирования педагогических тестов. Учебное пособие. М.: Изд. Логос, 2004 - 432 с.
 19. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. М.: Изд. Прометей, 2000 - 168 с.
 20. Linden W.J., Hambleton R.K. Handbook of modern item response theory, New York, 2001.