

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Саратовский государственный университет имени Н.Г. Чернышевского

Факультет компьютерных наук и информационных технологий

Кафедра информационных систем и технологий в обучении

ПЕДАГОГИЧЕСКОЕ ТЕСТИРОВАНИЕ

*Учебно-методическое пособие для студентов,
обучающихся по направлению подготовки
44.33.01 «Педагогическое образование» профиль «Информатика»,
очной формы обучения*

Саратов 2015

*Печатается по решению
кафедры информационных систем
и технологий в обучении*

Векслер В.А. Педагогическое тестирование для студентов, обучающихся по направлению подготовки 44.33.01 «Педагогическое образование» профиль «Информатика», очной формы обучения: учебно-методическое пособие / Векслер В.А. – Саратов: СГУ. 2015. - 53 с.

*Пособие предназначено для студентов, изучающих дисциплину
«Современные средства оценивания результатов обучения».*

Содержание

Введение.....	4
1. Возникновение тестологии	5
2. Этапы конструирования и классификации педагогических тестов	9
3. Тестовые задания закрытого типа	13
4. Особенности проектирования нестандартных тестовых заданий.....	17
5. Психолого-педагогические аспекты тестирования	22
6. Валидность теста	27
7. Стандартизация теста и интерпретация результатов тестирования	31
8. Матрица результатов тестирования	37
9. Использование программных средств	43
Заключение	48
Практикум	49
Вопросы для самостоятельного изучения	52
Литература:	54

Введение

Важным элементом образовательного процесса является контроль знаний и умений учащихся. Постепенный переход от традиционных форм контроля и оценивания знаний к инновационным методам, одним из которых является тестирование, отвечает общей концепции модернизации российской системы образования.

Тестирование в педагогике выполняет три основные взаимосвязанные функции: диагностическую, обучающую и воспитательную:

1. Диагностическая функция заключается в выявлении уровней знаний, умений, навыков обучаемого. Это главная задача тестирования.

2. Обучающая функция тестирования выражена в мотивировании обучаемого к активизации деятельности по усвоению учебного материала.

3. Воспитательная функция проявляется в периодичности и неизбежности тестового контроля. Это дисциплинирует, организует и направляет деятельность учащихся, помогает выявить и устранить пробелы в знаниях, формирует стремление развить свои способности.

В практической деятельности используются разнообразные формы тестовых заданий:

1. Задания закрытых форм:

- а) задания с выбором одного правильного ответа.
- б) задания с выбором нескольких правильных ответов
- с) задания с выбором одного неправильного ответа.

2. Задания открытых форм.

а) задания на дополнение.

б) задания со свободно конструированным ответом.

3. Задания на установление соответствия.

4. Задания на установление правильной последовательности

Пособие предназначено для студентов, изучающих курс «Современные средства оценивания результатов обучения». Цель курса, его содержание направлено на углубленное изучение проблемы, анализ практического опыта осуществления функций контроля и педагогической оценки, освещение и анализ типичных дидактических затруднений, встречающихся в контрольно-оценочной деятельности учителя.

В пособии рассматривается понятие тестирования как метода и инструмента педагогического измерения, выделяются формы тестовых заданий, даются рекомендации по их проектированию, рассматриваются статистические методы определения качества теста, приводится практикум для самостоятельного выполнения.

1. Возникновение тестологии

Педагогический тест представляет собой систему дидактических заданий специфической формы, позволяющей оценить уровень обученности учащихся: совокупность их представлений об объекте познания, знания, умения и навыки в той или иной области содержания. Изучает возможности педагогического тестирования специальная наука – тестология.

Понятие «тестология» имеет несколько значений: в переводе с английского языка это «наука об испытании», другое понимание данного термина определяет ее как метод педагогических измерений, и их интерпретации. Тестология стала междисциплинарной наукой, целью которой стало создание действенных и строго научно определенных методик и инструментов измерения и диагностирования испытуемых. Как наука, она появилась и сформировалась в конце 19 – начале 20 веков на стыке нескольких поведенческих наук: психологии, социологии, педагогики [1].

Сегодня считается, что родоначальником тестологии, ученым сделавшим первые попытки измерить различия между людьми в области элементарных психических процессов, стал английский исследователь Френсис Гальтон (1882–1911). В 1884 году им была организована антропометрическая лаборатория на проводившейся тогда Всемирной выставке в Лондоне. В данной лаборатории любой посетитель мог измерить свои физические возможности по разнообразным показателям: рост, сила кисти, острота зрения и др. Таким образом, можно было собрать и проанализировать первые данные по индивидуальным особенностям простых психофизических функций. Ф. Гальтон считал, что тесты сенсорного различения необходимо использовать как средство оценивающее интеллект человека.

Он определил три фундаментальных принципа тестирования:

- 1) использование серии совершенного одинаковых испытаний к большому количеству испытуемых;
- 2) необходимость накопления и обработки статистических результатов для анализа полученных результатов;
- 3) установление эталонов оценки.

Понятие «умственный тест» стал популярен после публикации статьи Джеймса Мак-Кина Кеттелла (1860–1944) «Умственные тесты и измерения», в 1890 году. Дж. Кеттелл посчитал тест уникальным средством для проведения научного эксперимента.

Он так же сформулировал ряд необходимых требований:

- необходимо создавать совершенно одинаковые условия для всех испытуемых;
- время тестирования ограничено (1 час);
- для всех диагностируемых используются одинаковые инструкции

- испытуемые должны четко понимать предлагаемые задачи для выполнения;
- испытание должно проводиться без зрителей;
- итоги проведенной диагностики должны обрабатываться с помощью статистических математических методов анализа (определяются минимальный, максимальный и средний показатель, вычисляется математическое ожидание и дисперсия).

Создатели самого первого педагогического теста, для измерения интеллектуальных способностей детей, А. Бине и Т. Симон, в период с 1905 по 1911 гг. провели эмпирическую проверку заданий, которые должны были быть включены в их тест. Критериями оценки заданий были эмпирическая мера трудности каждого задания, которую они определяли по доле правильных ответов в группах детей разного возраста, и степень совпадения результатов теста с мнением преподавателей. Таким образом, были заложены основные характеристики теста: норма трудности и валидность. Ими была разработана уникальная для того времени шкала Бине — Симона. Показатель интеллекта Бине определял как разность между «умственным» и биологическим возрастом.

Необходимо подчеркнуть, что разработки по созданию, совершенствованию и использованию тестов при решении практических задач, сопровождалась исследованиями в областях создания наборов статистических методов обработки и анализа полученных результатов. В этом направлении работали многие ученые, среди них был и Карл Пирсон, создавший в 1896 г. фундамент основ теории корреляции. Под влиянием его трудов формируется новая наука - прикладная статистика.

В начале XX века Чарльз Эдвард Спирмен (1863 – 1945), английский ученый, развивает теорию корреляционного анализа для стандартизации тестов. В 1904 г. он публикует свои классические исследования: «Общий интеллект, объективно детерминированный и измеренный» и «Доказательность и измерение связи между двумя предметами». Он приходит к следующему выводу: «корреляция между двумя переменными свидетельствует о наличии общего фактора, некоторой причины, определяющей эти переменные, и специфических факторов, характерных для каждой переменной». Тем самым им был определен путь целенаправленного отбора тестов для измерения различных сторон интеллекта человека. В ходе дальнейших исследований обнаружилось, что корреляции, явно или неявно существующие между тестами, не могут быть объяснены только наличием генерального общего фактора. В своем исследовании «Способности человека», изданной в 1927 г. была предложена двухфакторная теория интеллекта, которая выделяла уже оговоренный ранее: общий (генеральный) фактор, образующий фундамент любых успешных умственных действий и им был предложен ряд специальных факторов, нужных для решения задач в отдельных областях.

С начала XX века определилось и педагогическое направление в развитии тестологии [3]. Американский исследователь В.А. Макколл решил разделить тесты на два вида:

- Educatinal Test (педагогические или образовательные),
- Intelligence Test (психологические или определяющие уровень развития интеллекта).

Основной задачей педагогических тестов, по его мнению, стало диагностическое измерение успешности проведенного обучения по тем или иным предметам за определенный период, а также эффективность использования в образовательном процессе инновационных методов и технологий. Целью применения педагогических тестов, как считал В. Макколл, стало создание рекомендаций по объединению в группы учащихся, по признаку усваивания одинакового по объему материала и усваивания материала примерно с одинаковой скоростью.

Разработка первого педагогического теста принадлежит американскому психологу и педагогу Эдуарду Ли Торндайку (1874 – 1949). Его, не без основания, считают основоположателем современной системы педагогических тестовых измерений. Итогом проведенной исследовательской работы в области диагностирования обучаемых и применения метода тестов в педагогике стала книга «Введение в теорию психологии и социальных измерений» (издана в 1904 г.). Первый же унифицированный тест, вышедший под руководством Э. Торндайка, был тест на решение задач по счету, впервые снабженный «показателем».

В 1915 году Р.М. Йеркс создал свою серию тестов, главное отличие которой – полное изменение подхода к подсчету результатов проведенного тестирования. Вместо существовавших ранее возрастных долей, которые предложил в своих исследованиях А. Бине, диагностируемый получает за каждый правильно решенный тест заранее строго определенное количество баллов. Таким образом, проведение и подсчет результатов тестирования стало более эффективным [10].

Первые годы зарождения тестологии были годами поиска, позже тестирование выходит за пределы строго научного эксперимента и обращается в массовое явление, приводящее к полному итоговому осмыслению его роли и места в системе образования. Разрабатываются и внедряются в учебный процесс тесты О. Стоуна по счету, Б. Зекингема для диагностирования правописания, Э. Торндайка тестирующих знания в большинстве школьных общеобразовательных дисциплин. Т. Келли разрабатывает диагностический аппарат позволяющий измерять интересы и наклонности школьников.

Для более эффективной работы по разработке и использованию тестов стали создаваться частные и государственные организации и службы. В 1926 г. Советом колледжей (США) был принят к использованию тест SAT («Scholastic Aptitude Test» - полностью унифицированный оценочный тест для проведения вступительных экзаменов в высшие учебные заведения США). Он так же определил структуру испытаний (в виде заданий для

тестирования) для оценки профессиональной деятельности работников образовательных учреждений. Позже, в 1947 г. в США была создана «Educational Testing Service» (перев. с англ. – «Служба тестирования в образовании»), которая стала как научным центром педагогических исследований, так и базовой организацией, которая стала контролировать использование инструментов педагогического тестирования на всех уровнях образования.

Развиваться тестология в России начала после революции 1917 г. Было создано большое количество научно-исследовательских центров по разработке советских стандартов тестов (на основе существующих американских) и проверки уже получивших мировую известность измерительных диагностик. Но вскоре, отношение к тестологии резко изменилось. Хорошие результаты показывали испытуемые с «неблагополучным» для советского строя происхождением и плохие показатели показывали дети с пролетарскими корнями. Постановлением Наркомпроса 1936 г. работы были свернуты, в связи с «буржуазным» оттенком данного педагогического инструментария. Тем не менее тесты не исчезли полностью, их продолжали использовать при контроле выполнения самостоятельных работ, практических заданий и пр. Тестологические исследования, на научном уровне, начали возвращаться после 60-х годов 20-го века, на фоне дидактических исследований в области программированного обучения. К концу 20-го века, в России, снова были созданы центры, работающие в области тестологических исследований.

Сегодня роль тестирования как важнейшего, универсального инструмента контроля знаний обучающихся стремительно возрастает, происходит массовое, активное внедрение тестов в учебный процесс и создаются прямые предпосылки для создания и развития конструктивной и эффективной системы управления качеством образования.

Контрольные вопросы:

1. Дайте определение понятию «педагогический тест».
2. Определите цель тестологии.
3. Какой вклад в развитие тестологии внес Френсис Гальтон?
4. Требования к тесту Дж. Кеттелла.
5. Опишите назначение шкалы Бине — Симона.
6. Какой вклад в развитие тестологии внес Чарльз Эдвард Спирмен?
7. Какие в начале XX века определились педагогические направления в развитии тестологии за рубежом?
8. Какой вклад в развитие тестологии внес Эдуард Ли Торндайк?
9. Какие виды тестов предложил американский исследователь В.А. Макколл?
10. В чем отличие тестов Р.М. Йеркса от ранее существовавших тестов?
11. Какое назначение теста SAT?
12. Как развивалась тестология в России?

2. Этапы конструирования и классификации педагогических тестов

На основе исследований Федерального института педагогических измерений (США) и ряда отечественных тестологов выделим основные этапы конструирования педагогических тестов:

1. Этап целеположения (определяем вид, ресурсные возможности, источники финансирования, бюджет, круг привлекаемых специалистов).
2. Подготовительный этап (анализ и отбор содержания проверяемого материала, определение типов и количество используемых знаний, установка требований времени проверки, определение системы оценивания, выработка условий проведения).
3. Этап составления теста.
 - 3.1. Разработка и апробация тестовых заданий (в репрезентативных группах).
 - 3.2. Определение и расчет показателей качества созданных заданий (валидность, надежность),.
 - 3.3. Анализ результатов проведенной апробации.
4. Этап проведения теста и анализа полученных результатов
 - 4.1. Стандартизация теста (единообразие процедуры проведения и оценки выполнения тестовых заданий).
 - 4.2. Подготовительные работы по техническому оснащению.
 - 4.3. Отслеживание проведения теста.
 - 4.4. Обработка результатов, получение статистических данных для анализа. Подготовка итогового отчета.

Педагогические тесты имеют множество классификаций [2,11,14,15].

Классификация по целям использования:

1. Входное тестирования (проверка знаний и умений в начале обучения, выявляет готовность к усвоению знаний, определяет степень владения материалом до его изучения, помогает выбрать правильную стратегию).
2. Формирующее и диагностирующее тестирование (проводится в процессе обучения):
 - 2.1. Формирующее тестирование определяет качество усвоения материала, выполняет функцию обратной связи: если большинство не справилось, то нужно повторить; если

меньшинство не справилось, то с ними необходимо индивидуально заниматься.

- 2.2. Диагностический тест выявляет причины допущенных ошибок, задания предельно детализированы.

Пример: если в формирующем тесте не справились с заданием

типа: а) $\frac{\sqrt{15+7^2}}{2}$, то в диагностическом тесте проводится

проверка действий: а) 7^2 , б) $15+7^2$, в) $\sqrt{64}$, д) .

Классификация по формам предъявления информации:

1. Бланочные тесты – испытуемые отмечают или вписывают правильные ответы в бланки.
2. Компьютерные – задания высвечиваются на экране монитора.
3. Предметные – испытуемый манипулирует реальными физическими объектами.

Классификация по степени однородности задач:

1. Гомогенные – контроль знаний и умений проводится по одному предмету.
2. Гетерогенные – измерение уровня подготовленности по нескольким учебным предметам. При этом отметим два подвида:
 - 2.1. Полидисциплинарные – тест по нескольким разделам, каждый раздел является субтестом, при этом каждый субтест является тестом по отдельной дисциплине.
 - 2.2. Междисциплинарный – тест состоит из обобщенных заданий, проверяются интерактивные знания и умения путем сложного конструирования вопросов по смежным дисциплинам (математика и физика, химия и биология, история и культурология).

Классификация по характеру действий:

1. Вербальные – умственные действия (словесно-логические, вопросы).

2. Невербальные – практическое манипулирование.

Классификация по подходам к разработке тестов:

1. Нормативно-ориентированные – в данных тестах происходит упорядочивание всех испытуемых по уровню подготовленности: если все плохо ответили, то все равно проводится ранжирование среди низких результатов. В соответствии с эмпирическими данными возникают нормы, от показателей которого и происходит ранжирование. Выделяются требования: если все отвечают или все не отвечают на некоторое задание, то данное задание должно быть удалено; если все выполнили тест полностью или все не выполнили тест полностью, то никакого ранжирования не происходит – тест не подходит. Распределение заданий по трудности в нормативно-ориентированном тесте приведено на рисунке 2.1.



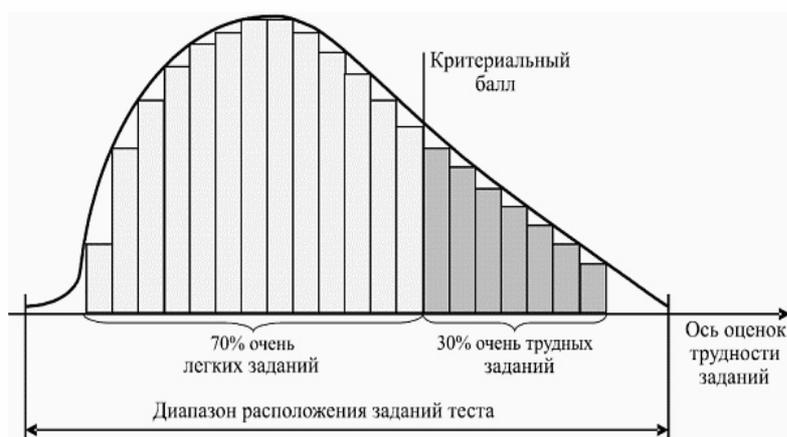
Рисунок.2.1. Распределение заданий по трудности в нормативно-ориентированном тесте.

Характеристики нормативно-ориентированного теста:

- а) В среднем выполняют все задания – 5-10% от общего числа испытуемых.
- б) Результаты сравниваются с результатами других испытуемых.

- с) Подбор заданий по трудности – распределение оценок трудности близкое к нормальному, основная часть заданий имеет трудность 40%.
- д) Разброс результатов (вариативность баллов) – высокий, поскольку основная цель дифференцирование испытуемых по уровню подготовки.

2. Критериально-ориентированный – в данном виде тестов проверяется степень усвоения испытуемыми определенного раздела, измеряется уровень относительно полных знаний. Данные тесты бывают ориентированы на предметную область и квалификационные. Цель проведения: «Знает ли материал?» Определены правила: все выполнил – значит все знает, все не выполнил – значит все не знает. Распределение заданий по трудности в критериально-ориентированном тесте приведено на рисунке 2.2.



Распределение заданий по трудности в критериально-ориентированном тесте

Рисунок 2.2 Распределение заданий по трудности в критериально-ориентированном тесте.

Характеристики критериально-ориентированного теста:

- а) В среднем выполняют все задания – 90% от общего числа испытуемых.
- б) Результаты сравниваются с содержательной областью дисциплины.
- с) Подбор заданий по трудности – распределение скошенное.

d) Разброс результатов (вариативность баллов) – низкий, внутри результатов группы испытуемых, превысивших по своим результатам критериальный балл.

Контрольные вопросы:

1. Какие существуют основные этапы конструирования педагогических тестов?
2. Опишите классификацию по целям использования
3. В чем различие между формирующим и диагностическим тестированием?
4. Опишите классификацию по формам предъявления информации.
5. Какие тесты называются гомогенными?
6. Какие тесты называются гетерогенными? Какие существуют их подвиды?
7. Опишите отличия между нормативно-ориентированным и критериально-ориентированными тестами.

3. Тестовые задания закрытого типа

Тестовые задания бывают нескольких типов. Самый популярный из них – тип тестовых заданий закрытого типа, они получили широкое распространение в основном из-за легкости автоматизации. Задания закрытого типа предполагают выбор одного или нескольких правильных ответов из нескольких предложенных вариантов. Основным недостатком таких заданий – возможность угадывания [2].

Выделим основные правила составления тестовых заданий [4, 5,6,8,15]::

1. Начинайте формулировать вопрос с правильного ответа.
2. Содержание задания должно отвечать программным требованиям и отражает содержание обучения.
3. Вопрос должен содержать одну законченную мысль.
4. При составлении вопросов следует особенно внимательно использовать слова «иногда», «часто», «всегда», «все» и «никогда».
5. Вопрос должен быть четко сформулирован, избегая слова «большой», «небольшой», «малый», «много», «мало», «меньше», «больше».

6. Избегайте вводных фраз и предложений имеющих мало связи с основной мыслью.

7. Неправильные ответы должны быть разумными, не должно быть явных неточностей.

8. Реже используйте отрицание в основной части. Избегайте двойных отрицаний («Почему нельзя не делать?»).

9. Лучше использовать длинный вопрос и короткий ответ.

10. Из текста заданий необходимо исключить все вербальные ассоциации, способствующие выбору правильного ответа путем угадывания.

11. Из числа тестовых заданий закрытого типа исключайте задания требующие громоздких вычислений.

12. Используемая терминология не должна выходить за рамки учебного курса.

13. Текст заданий и варианты ответов должны быть записаны разными шрифтами, для быстрого их отличения друг от друга при выполнении.

14. В начале каждого теста должна быть предусмотрена инструкция по его применению.

Дистрактором в тестовых заданиях закрытого типа назовем вариант неверного ответа. Слово «дистрактор» произошло от английского слова «distract», которое переводится как «отвлекать». Таким образом, дистракторы – это неправильные ответы, которые отвлекают и сбивают с толка испытуемого.

Выделим требования к дистракторам:

1. В основной части задания должно быть больше ключевых слов, в ответах не более трех наиболее ключевых слов.

2. Ответы к одному заданию должны приблизительно одной длины.

3. Все дистракторы к каждому заданию должны быть равнопривлекательными.

4. Дистракторы не должны быть неполными правильными ответами, превращающимися в правильные при определенных условиях.

5. Не должно быть неработающих дистракторов (их обязательно должны выбирать).

6. Частота выбора дистракторов должна быть примерно одинакова.

7. Из ответов необходимо исключить повторяющиеся слова путем ввода их в основной текст задания.

8. В ответах не рекомендуется использовать слова: «все», «ни одного», «никогда», «всегда», «ни один из перечисленных», «все вышеперечисленные». В отдельных случаях они способствуют угадыванию правильного ответа.

9. Из числа дистракторов исключают ответы вытекающие один из другого.

10. Все ответы должны быть параллельными по конструкции и грамматически согласованными с основной частью задания.

11. Желательно не использовать в ответах следующие слова: «да/нет», «верно/неверно». Рекомендуется создавать более содержательные ответы.

12. Если номер ответа совпадает с его числовым значением, то они обязательно пишутся в одной строке (1. 1 Бит, 2. 2 Бит, 3. 8 Бит, 4. 4 Бит). Если дистракторы это слова, то их желательно расположить в алфавитном порядке.

Заметим, что большое количество дистракторов уменьшает вероятность угадывания, но требует большего времени при разработке. Плохо знающие материал учащиеся пытаются найти правильный ответ путем угадывания по «формальным» признакам: самый длинный ответ, самый умный ответ, хорошо обоснованный ответ).

Под фасетом понимается форма, обеспечивающая представление нескольких вариантов одного и того же элемента содержания теста. Фасетный тест представляет собой большой массив элементов – фасеток, из которых путем различных сочетаний конструируются тестовые задания.

Приведем небольшие примеры фасетного теста:

1. В одном килобайте {мегабайте, гигабайте} содержится байт:

1) 1024, 2) 1048576, 3) 1073741824.

2. Год начала Советско-финской войны {Второй Мировой войны, Великой Отечественной войны}

1) 1941, 2) 1914, 3) 1939.

Таким образом, для снижения вероятности угадывания мы можем порекомендовать:

1. Увеличение количества дистракторов в задании до 4-х.
2. Использование фасетов заданий.
3. Жесткое временное ограничение на выполнение задание.
4. Использовать формулы для коррекции баллов с учетом возможности угадывания.

Оценивать результаты проведенного тестирования можно по-разному.

Алгоритмы оценки выполнения тестового задания

- Дихотомическая оценка выполнения задания. Термин «дихотомически оцениваемые тестовые задания» означает, что ответ экзаменуемого находится только в одной из двух возможных категорий, правильный (за него выставляют 1 балл), или неправильный (0 баллов).

- Политомическая оценка выполнения задания. Термин «политомически оцениваемые тестовые задания» означает, что за разные ответы на задание проставляют несколько различных чисел.

Рассмотрим несколько методов оценивания заданий закрытой формы.

1. Традиционный метод. Данный метод подсчета баллов заключается в следующем: суммируются все баллов по каждому заданию.

1.1. Оценка с весовым коэффициентом. При условии выбора испытуемым только правильных ответов начисляется полный балл, в противном случае начисляется ноль баллов. Недостаток данного способа заключается в снижении дифференцирующей способности задания, т.к.

испытуемые с «частичным набором знаний» и испытуемые с «полным отсутствием знаний» практически не различимы.

1.2. Метод частичного балла. Суть данного метода заключается в следующем: частично баллы начисляются в случае выбора даже не полного ответа, но в случае выбора в сочетании с правильными ответами некоего дистрактора – баллы обнуляются полностью или частично уменьшаются.

2. Коррекция баллов с учетом возможного угадывания. По данному методу производятся вычисления по следующей формуле: $Z_i = X_i - W_i / (k - 1)$, где Z_i - скорректированный тестовый балл i -го испытуемого, X_i – тестовый балл i -го испытуемого без коррекции, W_i - число ошибочных ответов у i -го испытуемого, k – количество ответов в заданиях (должна быть постоянная величина). Например: В тесте было 30 заданий, на каждое давалось по 4 ответа. Испытуемый на 20 заданий ответил правильно, 10 заданий выполнено с ошибками. Вычисляем балл по форме оценки с весовым коэффициентом, но с учетом коррекции по угадыванию: $Z_i = 20 - 10 / (4 - 1) = 16,7$.

3. Метод коррекции «Правильный + пропуск». Формула, основанная на модели случайного угадывания, которая корректирует "эффект угадывания", имеет вид: $Z_i = X_i + N_i / k$, где X_i – скорректированный индивидуальный балл i -го испытуемого, X_i - число правильных ответов i -го испытуемого, N_i - число пропущенных заданий у i -го испытуемого, k - число вариантов ответов в каждом задании (все задания должны иметь k вариантов). Предположим, что студент А и студент Б сдают некоторый тест. Тест состоит из 150 заданий закрытой формы с четырьмя вариантами ответов. Студент А правильно ответил на 90 заданий, а оставшиеся 60 заданий он пропустил (так как не был уверен в выборе ответа). Студент Б также правильно ответил на 90 заданий, однако на остальные 60 заданий он ответил путем простого угадывания. Так как студент Б имеет шанс в 25% угадать правильный ответ, то он может набрать 15 дополнительных баллов только путем угадывания. При традиционном подсчете баллов студент А наберет 90 баллов, в то время как студенту Б будет начислено 105 баллов. Таким образом, индивидуальные баллы у этих двух студентов имеют значительную разницу, которая не отражает, уровень их подготовки, а является лишь функцией склонности к угадыванию правильного ответа. Эта разница должна учитываться, так как тестирование проводится, как правило, однократно, а его результаты существенно влияют на аттестацию. Если же мы применим к подсчету баллов метод коррекции «Правильный+пропуск», то мы получим одинаковое количество баллов у данных студентов (105). Таким образом, данный способ позволит правильно дифференцировать студентов по их уровню знаний.

Резюмирую все вышесказанное, отметим, что задания закрытого типа в педагогических тестах широко распространены. Основная причина такого широкого распространения это легкость, как компьютерной автоматизации, так и ручного анализа. Для данного типа заданий выработаны методы конструирования содержания и весь необходимый инструментарий

правильного оценивания испытуемых с учетом всех возможных погрешностей, в том числе фактора угадывания.

Контрольные вопросы:

1. Преимущества тестовых заданий закрытого типа.
2. Основные правила составления тестовых заданий.
3. Дистрактор. Требования к дистрактору.
4. Назначение фасетного теста.
5. Какие вы можете дать рекомендации по снижению вероятности угадывания правильного ответа в тесте?
6. Приведите примеры алгоритмов оценки выполнения тестового задания.

4. Особенности проектирования нестандартных тестовых заданий

Тестирование, как элемент контроля знаний, уже давно вошло в практику. Развивается наука тестология, описывающая правила создания тестов. Тестовые задания мы можем условно разбить на две группы: стандартные (часто применимые в виду легкости автоматизации) и нестандартные (редко применимые, в виду трудностей автоматизированного применения). В частности, к стандартным тестовым заданиям мы можем отнести задания закрытой формы, т.е. задания предполагающие выбор одного и нескольких ответов из предложенных вариантов. К нестандартным мы отнесем задания открытой формы, задания на установление соответствия и задания на установление правильной последовательности.

Отличительной особенностью заданий открытой формы является, то, что правильный ответ записывает сам испытуемый в процессе их выполнения. К заданиям открытой формы мы можем отнести задания на дополнение и задания со свободно конструируемыми ответами.

Задания на дополнение предполагают, что ответ формируется самим тестируемым и конечно предполагает наиболее краткий и как правило лаконичный ответ. Задания на дополнение должны сопровождать инструкции начинающиеся словом «Дополните». Далее будет следовать основная часть задания, где сформирована проблема и делается пропуск на месте элемента, знание которого проверяется.

Сформируем ряд требований которые необходимо выполнять при создании заданий такого типа [4, 5,6,13,15]:

1. Необходимо использовать стандартные формулировки.
2. Необходимо соблюсти точность задания.
3. Вариант ответов должен быть один.

4. Каждое задание должно быть нацелено на одно дополнение, место дополнения обозначается подчеркиком.

5. Дополнение лучше спроектировать в конце задания (предложения).

6. Все подчеркики для дополнений в одном тесте должны быть одной длины.

7. После подчеркика, в случае необходимости, укажите единицу измерения.

8. В тексте задания желательно исключить повторы и двойное отрицание.

К достоинствам таких тестовых заданий можно отнести:

1. Исключение возможности угадать правильное слово.

2. Удобно использовать при проверки номенклатурных знаний (определение, свойства, характеристики, этапы), учебной эрудиции обучаемых.

3. Легко разрабатываются.

4. Краткость и однозначность ответов.

5. Необходимость воспроизведение ответов по памяти.

6. Простота и удобство проверки теста экспертом.

К существенным недостаткам мы можем отнести:

1. Сложность формализации правильного ответа.

2. Возможность неоднозначности формулировки.

3. Трудности связанные с использованием автоматических средств для обработки результатов тестирования.

Выделим общие логические принципы, которых нужно придерживаться, при формулировке заданий:

1. Принцип логической определенности содержания задания – задания являются логически определенными, если они способны продуцировать у знающих учащихся правильный ответ и если содержание и форма задания помогают организовать мыслительный процесс нахождения такого ответа. Пример нарушения: Первая социал-демократическая партия была создана в _____ (Где? или Когда?). Любое задание должно пониматься с первого прочтения.

2. Принцип вариативности (фасетности) заданий. Идея фасета ассоциируется с практикой создания нескольких вариантов заданий одинаковой сложности.

3. Принцип обратимости. Если обратимые задания находятся в одном тесте, возникает тавтология (ответ на одно задание можно найти из обратимого задания).

4. Принцип краткости – формулировка заданий кратка и лаконична.

5. Принцип неотрицательности – в основе задания не рекомендуется использовать отрицательные частицы, а так же определение понятия через перечисление элементов не входящих в него. Пример нарушения принципа: Банки не ставятся на _____ (позвоночник и много еще куда).

6. Принцип логической соразмерности объема определяющего понятия объему определяемого понятия. Пример нарушения: Первым космонавтом был _____ (Ю. Гагарин или мужчина).

К заданиям открытого типа относят так же задания со свободно конструируемыми ответами. Такая форма заданий предполагает представление ответа в развернутом виде, т.е. отражение полного решения поставленной задачи со всеми необходимыми пояснениями или представление ответа в виде небольшого сочинения. В свободно конструируемом ответе испытуемый излагает свои мысли в свободной форме. Данная форма внешне схожа с традиционными формами проверки знаний: письменная контрольная работа, сочинение. Задания такого типа преследуют цель проверки и оценки различных уровней познавательной деятельности испытуемых: проверка знаний фактического материала, умение четко и ясно выражать свои мысли, логичность составления ответа на вопрос, оригинальность, демонстрация способов решения поставленной задачи, отображение всего процесса выполнения, повышение объективности оценивания.

Отметим ряд требований необходимых для проведения тестирования со свободно конструируемыми ответами:

1. Создание четких инструкций в которых оговаривается полнота ответа для получения максимального балла, отражение рекомендаций касающихся длины ответа.

2. Персональный компьютер может выполнять лишь роль сбора ответов, а не их проверки.

3. Для обеспечения объективности оценивания ответа необходимо привлечь двух-трех экспертов.

4. Рекомендуются использовать такие задания только в тех случаях, где оценка другими способами не применима.

5. Убедитесь, что задания соотносятся с целями обучения.

6. Установите лимит времени выполнения.

7. Данный тип ответов наиболее подходит для отображения хода вычислений.

8. Для анализа ответа разработчик должен установить критерии оценивания ответа – т.е. перечисление некоторых образовательных условий при выполнении которых ответ можно засчитать как правильный.

Задания со свободно конструируемыми ответами на практике обладают рядом преимуществ:

1. Они интересны и разнообразны в содержательном плане.

2. Оценке подаются умения характеризующие познавательную деятельность учащихся, выявляются способы решения задач, анализируются типичные ошибки.

3. Задания такого типа легко разрабатываются.

4. Фактор угадывания отсутствует.

5. Проверяют сложные структурные знания по предмету.

К недостаткам заданий такого тип отнесем:

1. Трудность автоматизации проверки.
2. Необходимы значительные временные и умственные затраты эксперта при проверке.
3. Очень субъективны в проверке, поэтому потенциально менее надежны.
4. Требуют большого времени у испытуемого на выполнение.

К нестандартным тестовым заданиям относятся задания на установление соответствия. Эти задания, по мнению Н.Н. Самылкиной, позволяют проверить ассоциативные знания, знания взаимосвязи определений и фактов, сущности явлений, соотношения между объектами и их свойствами, законами и формулами. При выполнении таких заданий испытуемый определяет соответствие между элементами двух множеств [15].

Определим требования к разработке заданий на установление соответствия:

1. Рекомендуются начинать задание с фразы «Установите соответствие». Далее вами должны быть предложены два множества в виде списка элементов. Слева элементы «задающего» множества (постановка проблемы), которые как правило кодируются с помощью арабских цифр, справа элементы «должного» множества (раскрытие проблемы), именно их нужно выбирать по принципу соответствия, они кодируются буквами русского алфавита.

2. Рекомендуются список второго множества сделать длиннее (т.е. включать дистракторы – неверные отвлекающие внимание ответы, которые равновероятно подобраны), для снижения процента угадывания.

3. Элементы двух множеств должны быть строго выбраны по одному основанию.

4. Рекомендуются помещать в текст задания названия столбцов (обобщающее определение множества).

5. Если задания не предполагают четкого соответствия (т.е. одному элементу левого списка соответствует только один элемент правого списка), то в инструкции к заданию должно обязательно указываться о наличии неоднозначного соответствия, о наличии дистракторов, и о том, сколько раз могут использоваться элементы второго списка.

6. В связи с особенностью восприятия, число элементов одного списка не должно превышать десяти, количество элементов левой группы не менее двух элементов.

7. Весь текст одного задания необходимо расположить на одной странице, перенос отдельных элементов не допускается.

8. Испытуемый может рисовать линии соответствия выполняя задания, но рекомендовано выписывать ответ, например: «1б, 2а, 3с». Для этого, в самой нижней части задания отводится место для ответов.

9. При оценивании таких ответов, за каждое правильно установленное соответствие начисляется один балл.

К принципам составления заданий на установление соответствия относятся: фасетность, краткость и точность формулировок, понятность для испытуемых, однородность элементов.

Выделим три вида заданий на установление соответствия:

1. Соответствия взаимно-однозначные: любому элементу из левого списка соответствует только один элемент из правого списка и наоборот.
2. Соответствие не взаимно-однозначное: различным элементам из левого списка может соответствовать один и тот же элемент из правого списка.
3. Соответствие с многократным выбором: каждому элементу левого столбца могут соответствовать несколько элементов правого столбца (по мнению В. Аванесова, данный вид является громоздким и порождает на практике множество ошибок [4]).

Задания на установление правильной последовательности, так же относят к нестандартным тестовым заданиям. По определению Н.Н. Самылкиной, в этих заданиях, проверяют знания определенной последовательности действий для получения конечного результата, знание хронологий событий, расстановка правильного порядка элементов, построение логических цепочек. Заметим, то, что данные задания являются частными случаем заданий на установление соответствия, с точки зрения логики их конструирования [15].

Выделим требования к заданиям на установление правильной последовательности:

1. Основные элементы задания: инструкция, название задания, содержание задания, место для ответов.
2. Инструкция для испытуемого начинается со слов «Установить правильную последовательность».
3. В содержании задания ранжируемые элементы ставятся в случайном порядке.
4. Для оценивания, чаще всего используют дихотомическую оценку (если все правильно – получаем балл, иначе получаем ноль баллов). Применим более сложный вариант оценивания: идеальный ответ – ноль баллов, ошибка в конце задания – один балл, ошибка в середине – два балла, ошибка в начале – ноль баллов).

Отметим особенности заданий на установление правильной последовательности:

1. Задания довольно специфичны, содержание многих дисциплин не поддаются трансформации в эту форму.
2. Проверке подвержены определенные знания: алгоритмические, процессуальные, процедурные, технологические.
3. Более трудны для учащихся по сравнению с заданиями других форм.
4. Имеют низкий коэффициент угадывания.
5. Трудны в разработке.

Таким образом, все многообразие нестандартных тестовых заданий мы можем порекомендовать использовать при создании тестов, что позволит максимально снизить коэффициент угадывания и более глубоко провести контроль знаний, умений и навыков испытуемого.

Контрольные вопросы:

1. Выделите группы тестовых заданий.
2. Определите особенности заданий открытой формы.
3. Отличия заданий на дополнения от других форм тестовых заданий.
4. Какие требования рекомендуется соблюдать при создании заданий на дополнение?
5. Какие существенные недостатки заданий на дополнение?
6. Логические принципы формулировки тестовых заданий.
7. Особенности заданий со свободно конструируемыми ответами.
8. Требования необходимые для проведения тестирования со свободно конструируемыми ответами.
9. Какие существенные недостатки заданий со свободно конструируемыми ответами?
10. Определите назначение заданий на установление соответствия.
11. Перечислите принципы составления заданий на установление соответствия.
12. Требования к заданиям на установление правильной последовательности.

5. Психолого-педагогические аспекты тестирования

Сегодня тест – это инструмент, состоящий из квалиметрически (с использованием различных методов определения качества) тщательно экспериментально проверенной и апробированной системы тестовых заданий, стандартизированной процедуры проведения и заранее спроектированной технологии анализа результатов для измерения качеств и свойств индивида, достижений учебно-познавательной деятельности, измерение которых можно провести только в условиях систематического обучения.

Тесты можно разделить по целям и содержанию [3, 7]:

1. Тесты личности – служат для оценивания эмоционально-волевых качеств индивида.
2. Тесты интеллекта – анализируют уровень развития познавательных процессов и функций мышления.

3. Тесты способностей – предназначены для оценивания индивидуальных возможностей индивида в овладении различных видов деятельности.

4. Тесты достижений – служат для оценки развития комплекса знаний, умений и навыков по окончании обучения.

Тест, отличен от других форм контроля. Выделим существующие различия:

1. Содержательное наполнение теста подвержено четкому планированию. При разработке заданий для теста проводится отбор содержательной базы, происходит планирование подходящей формы заданий, определяется их необходимое количество и способы визуального расположения. По окончании выявления содержательного наполнения проводится анализ полученного результата экспертами.

2. Существует своя неповторимая форма заданий, отличающая тест от других форм контроля. В тестах форма заданий обязательно становится стандартизированной – по формам предъявления и по формам записи ответов испытуемыми.

3. Тестовые задания характеризуются таким параметром как трудность. На основе данных параметров можно получить статистические характеристики.

4. Присутствие специализированных шкал, которые соотносятся с существующими стандартизированными нормами для подведения итоговых результатов проведенного тестирования.

5. Существуют дополнительные оценки точности для проведенных измерений результатов проведенного тестирования.

Для тестов существует ряд характеристик, наиболее существенными из которых являются:

1. Валидность. Показывает насколько соответствует проверяемый материал поставленным педагогом целям контроля.

2. Надежность. Отражает насколько результаты тестирования стали устойчивыми при многократном использовании данного материала, так же показывает постоянство при влиянии посторонних случайных факторов. Можно отметить 3 метода при помощи которых можно проверить надежность: проведение повторного тестирования через две недели, проведение параллельного тестирования и деления теста на две части – четные и нечетные группы вопросов.

3. Репрезентативность – отражает на сколько в тесте был полностью охвачен изученных учащимися учебный материал.

4. Стандартизованность – выработка единого набора действий и правил для проведения тестирования и последующего подведения итогов и получения статистических характеристик.

5. Адаптивность – параметр указывает возможно ли применять тест в изменяющихся условиях.

6. Надежность от угадывания – насколько тест защищен от такого случайного фактора, как угадывания правильного ответа испытуемым.

Появление в контексте обучения такой инновационной формы контроля как тестирования приводит, прежде всего, к изменению как восприятия так и существующих установок контроля. Происходит ряд изменений, в частности, меняется психологическая ситуация тесно связанная с принятием (или одобрением) или неприятием (или отторжением) тестов как новой формы контроля.

Н.Ф. Ефремовой было проведено диагностическое исследование поставившее своей целью выявить оценку методов традиционного и тестового контроля (в нем приняли участие около 500 человек) [11]. Было выявлено следующее:

1. Наибольший балл при этом получила объективность процедуры тестирования и независимость оценки от мнения преподавателя, а также постулат неразглашение оценок;

2. Среди преимуществ традиционного контроля были оценены: возможность помощи, низкий уровень трудностей задания.

Многие исследования отметили важность психологической подготовки к тестированию:

1. Степень улучшения результатов тестирования зависит от способностей и знаний учащихся, количества и вида предварительных занятий, особенностей тестов.

2. Подготовка к тестированию наиболее полезна для слабых учащихся, на результаты сильных учеников, она влияет незначительно.

3. Степень положительного влияния предварительной подготовки на изменение результатов тестирования находится в прямой зависимости от тесноты связи между содержанием тренировочных заданий и содержанием теста.

4. Предварительная подготовка снижает тревожность в поведении испытуемых во время тестирования и мотивирует к выполнению теста.

5. Предтестовые тренинги помогают механизмам памяти работать в жестко лимитированном по времени контрольном мероприятии.

Под мотивацией в педагогике, сегодня, понимается единая система целей, потребностей мотивов (динамический процесс управления поведением человека), побуждающих индивида как сознательно относиться к учению и к контролю, как составной части так и быть активным в учебно-познавательной работе. Мотивация выполняет в образовании следующие функции: побуждающую, направляющую, организующую и смысловую. При проведении тестирования мотивация служит одним из основных психологических факторов успешности контроля.

На уровень мотивации учебно-познавательной деятельности всегда влияют комплексы объективных факторов (так их как история изучаемого предмета, авторитет преподавателя) и субъективных (жизненные планы человека, стремление к самореализации, желание иметь хорошие оценки). Процесс побуждения человека к тестированию нередко оказывается очень сложным социально-психологическим, интеллектуальным, эмоциональным и волевым актом. Человек имеющий опыт выполнения тестов обладает

значимым преимуществом перед теми, кто впервые участвует в тестировании: ведь им было преодолено чувство неизвестности, развита уверенность в себе и своих возможностях. Американский психолог Анна Анастаси считает, что у человека опытного в работе заданиями тестового формата, показатели на одиннадцать процентов выше, чем у неискушенных испытуемых [5].

Существует два важных компонента тестовой тревожности – эмоциональная и озабоченная тревожность. Эмоциональная составляющая тестовой тревожности охватывает чувства, физиологические реакции, такие как напряжение. Озабоченность связана с негативными мыслями, с ожиданием неудачи при выполнении теста и с озабоченностью последствий неуспеха. В связи с этим психологами разрабатывается технология «поведенческой терапии».

Исследования американского психолога Анны Анастаси и ряда других психологов показывают, тестовая тревожность способна убывать без «терапевтического вмешательства». Тестовая тревожность – комплексный феномен, вызываемый множеством разнообразных по сути и содержанию причин, специфических для каждого конкретного человека. К числу способов позволяющих преодолевать данный вид тревожности относят совершенствование языковых коммуникативно-речевых навыков и умений, правильная разработка инструкций и руководств к тестам [5].

Методика проведения тестирования должна базироваться на личности педагогов, но не только тех которые участвуют как эксперты в создании теста, но и тех, кто связан с самим процессом проведения данной формы контроля. Именно поэтому необходимо сказать и об этических нормах и принципах тестирования в образовании. Стоит отметить, что они прежде всего должны быть продиктованы кодексом профессиональной этики педагога. Являясь частью совокупности общечеловеческих норм, регламентирующих поведение педагога на тестировании в соответствии с требованиями долга, профессиональной честности, принципами гуманизма, они становятся базовыми, фундаментальными и основополагающими принципами при стандартизации теста.

К данной, достаточно большой группе принципов мы можем отнести:

1. Принцип конфиденциальности – определяет правила неразглашения сведений о результатах тестирования без согласия испытуемого.

2. Принцип доступности – связан с правом испытуемых на получение доступа к содержательному описанию и полной интерпретации тестовых результатов, получение информации выявленной в результате анализа проблем и неудач выявленных при выполнении отдельных заданий проведенного теста.

3. Принцип обоснованности и динамического отражения развития – обуславливает систематическое обновление данных о подготовленности обучаемых полученных с помощью тестов (мониторинг).

4. Принцип осведомленного согласия – предполагает согласие испытуемых на проведение тестирования и анкетирование.

5. Принцип отбора содержания – содержание теста определено как полностью оптимальное отображение содержания материала учебной дисциплины в системе созданных тестовых заданий.

6. Принцип научной достоверности – в тест включается только то содержание проверяемого материала, которое является объективно научно истинным и подвергается некой рациональной аргументации.

Психологические и педагогические тесты в образовании, между собой имеют глубокие различия. Анна Анастаси пишет, что педагогические тесты оценивают результаты усвоения общего для всех обучаемых программного курса, а психологические тесты предназначены для отражения индивидуального восприятия жизненного опыта, различающегося по социальным и экономическим условиям его приобретения [5].

Результаты проведенного тестирования в психологии используются главным образом для прогнозирования успешности какой-либо деятельности в зависимости от полученных качественных и количественных оценок личностных характеристик, и склонностей или носят рекомендательный характер при выявлении личностных проблем, для определения путей их преодоления.

Педагогические тесты, в частности, в итоговом контроле, дают педагогу информацию о количественной оценке существующего уровня подготовленности обучаемого, которая предназначена, в основном, для принятия административных и управленческих решений.

Определим ряд внешних различий между психологическими и педагогическими тестами в образовании:

1. Формулировка заданий (в психологических тестах носит как правило косвенный характер, в педагогических тестах – прямой характер)

2. Диагностика (в психологических тестах не существует «правильных» ответов, происходит определение характеристик личности, при этом педагогические тесты проверяют знания, а значит требуют строго соответствия ответа испытуемого правильному ответу).

3. Цели интерпретации (в психологических тестах определяют характеристики личности, в педагогических выявляется уровень усвоения знаний у испытуемого).

Заметим, что иногда границы между рассматриваемыми видами тестирования могут стираться, особенно в тех случаях, когда тесты нацелены на проверку логического мышления и творческих способностей учащихся. Резюмируя выше сказанное отметим, что, таким образом так как тестирование является уникальным методом контроля и оценивания знаний, позволяющий провести качественное и количественное измерения уровня знаний, и психологических особенностей, и предпочтений испытуемого, необходимо сформировать у участника тестирования такую психологическую атмосферу, которая позволит ему полностью раскрыться. Именно поэтому, сегодня, разработка методик, требований и правил

проведения тестирования является важной задачей повышения качества образовательного процесса.

Контрольные вопросы:

1. Определите тест как метод педагогического измерения.
2. Определите тест как педагогический инструмент.
3. Приведите классификацию тестов по целям и содержанию.
4. Существенные различия тестов от других форм контроля.
5. Перечислите основные характеристики тестов.
6. В чем выражается важность психологической подготовки к тестированию?
7. Какие существуют компоненты тестовой тревожности?
8. Принципы тестирования в образовании.
9. Выделите различия между психологическими и педагогическими тестами в образовании.

6. Валидность теста

Понятие «валидность» произошло от англ. слова «valid», что в переводе означает «годный». Валидность — один из базовых показателей качества созданного теста, характеризует пригодность тестовых результатов для той цели, ради чего проводилось тестирование. Из двух тестов, предназначенных для одной и той же цели, более эффективен тот, который быстрее, дешевле и качественнее измеряет знания данной группы испытуемых.

Аванесов В.С. в своей статье «Тесты: история и теория» пишет: «Валидность характеризует пригодность теста для измерения определенной величины, но нельзя говорить о валидности не указав его применения» [3]. Приведем пример для понимания смысла валидности. Два лучника стреляют по мишеням. Первый лучник набрал 70 очков, а второй 80 из 100. На первый взгляд кажется, что второй лучник лучше первого. Но при уточнении результатов оказалось, что второй лучник поразил чужую мишень. Поэтому, второй стрелок является «не валидным», он не может достигнуть цели, которая перед ним ставилась.

В научной работе Майорова А. «Теория и практика создания тестов для системы образования» говорится «Валидность – определяет насколько тест отражает то, что он должен оценивать» [13].

Валидизация теста, это ни что иное как процесс накопления подтверждений для доказательства его валидности.

В теории тестирования существует несколько классификаций валидности. Майоров А. выделяет три вида валидности: конструктивная, критериальная и содержательная [13].

Конструктивная валидность – определяется в тех случаях, когда представление об измеряемом свойстве существует в форме абстрактного образа. Например, для объяснения существования определенных качеств личности создается концептуальная модель (обоснованная теоретически), которая с помощью тестов подтверждается или опровергается. Опасности нарушения конструктивной валидности заключаются в неправильном обозначении причины и эффекта с помощью абстрактных терминов, терминов, взятых из обыденного языка или формальной теории.

В критериальной или эмпирической валидности оценивается величина связи теста с независимыми критериями. Для оценки степени связи используются средства математической статистики (корреляция двух рядов значений: баллов по тесту и показателей по внешним параметрам). При этом можно выделить два подвида: текущая и прогностическая валидность. Текущая валидность рассматривает корреляцию с критерием существующим в настоящее время (например, тест по предмету сравнивается со школьными оценками). Прогностическая валидность определяется на основе сравнения результатов с критериями которые предположительно появятся в будущем (например, результаты прошлого года предполагаемо переносим на следующий год).

Содержательная валидность характеризует тест по степени его соответствия предметной области, глубина и проработка соответствия. Анна Анастаси пишет «Содержательная валидность означает систематическую проверку содержания теста, с тем чтобы установить соответствует ли оно репрезентативной выборке измеряемой области» [5].

Согласно Клайну П., тест с высокой содержательной валидностью можно определить следующим образом [12]:

1. Указать категорию лиц, для которых предназначен тест.
2. Составить список знаний, умений и навыков, подлежащих тестированию.
3. Выполнить внешнюю экспертизу полученного списка на предмет его полноты и обоснованности.
4. На основе списка составить перечень заданий.
5. Выполнить внешнюю экспертизу составленных заданий.
6. На данной основе создать тестовые задания образующие тест, при этом тест будет содержательно валидным.

Определить коэффициент валидности теста – значит определить, как выполнение теста соотносится с другими независимо сделанными оценками знаний испытуемых. Для определения валидности требуется независимый внешний критерий, то есть оценка эксперта (преподавателя). За коэффициент валидности принимают коэффициент корреляции результатов тестовых измерений и критерия. Если экспертная оценка знаний испытуемых, полученная независимо от процедуры тестирования, представлена числовой последовательностью Y_1, Y_2, \dots, Y_n , то коэффициент валидности теста может

быть рассчитан по формуле, предложенной в своей работе Анастаси А., Аванесовым В.С [1, 4, 5, 15]:

$$V = \frac{\frac{\sum_{i=1}^n (Y_i \cdot y_i)}{n} - \bar{Y} \cdot \bar{y}}{S_Y \cdot S_y} \cdot \frac{n}{n-1},$$

где n – количество испытуемых, Y_i – экспертная оценка i -го учащегося, \bar{Y} – средняя арифметическая экспертных оценок, S_Y – стандартное отклонение экспертных оценок, y_i – количество правильных ответов у i -го учащегося, \bar{y} – средняя арифметическая количества правильных ответов, S_y – стандартное отклонение количества правильных ответов:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n},$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}},$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n},$$

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}.$$

Интерпретация результатов:

- а) если V принимает значение в диапазоне от 0,6 до 1 – валидность считается высокой,
- б) если V принимает значение в диапазоне от 0,3 до 0,6 – валидность считается средней,
- с) если V принимает значение меньше 0,3 – валидность считается низкой,

Рассмотрим пример расчета валидности теста.

Тест «Информационные процессы» должна выполнить группа студентов из 7 человек (в нашем примере мы рассмотрим небольшую группу, на практике рекомендуется проводить проверку теста на группе не менее 30 человек). Тест состоит из 10 заданий.

Экспертом, предварительно оцениваются результаты (экспертом может выступить преподаватель информационных технологий, хорошо знающий

данную группу студентов, и имеющий возможность предсказать результата выполнения ими данного теста).

Результаты экспертного оценивания приведены в таблице 6.1.

Таблица 6.1. Результаты экспертного оценивания.

Номер студента	1	2	3	4	5	6	7
Y_i	5	10	3	4	7	6	9

Расчеты можно произвести в электронных таблицах Microsoft Excel используя встроенные функции для вычислений.

Рассчитаем \bar{Y} - среднюю арифметическую экспертных оценок:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{5 + 10 + 3 + 4 + 7 + 6 + 9}{7} = 6,3.$$

Функция СРЗНАЧ в Microsoft Office Excel возвращает среднее значение (среднее арифметическое) аргументов. Например, если диапазон A1:A20 содержит числа, формула =СРЗНАЧ(A1:A20) возвращает среднее значение этих чисел.

Найдем S_Y - стандартное отклонение экспертных оценок:

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}} = \sqrt{\frac{(5 - 6,3)^2 + (10 - 6,3)^2 + (3 - 6,3)^2 + (4 - 6,3)^2 + (7 - 6,3)^2 + (6 - 6,3)^2 + (9 - 6,3)^2}{6}}$$

Функция СТАНДОТКЛОН в Microsoft Office Excel оценивает стандартное отклонение по выборке с использованием «n-1» метода. Стандартное отклонение — это мера того, насколько широко разбросаны точки данных относительно их среднего.

После проведения теста, были получены результаты, приведенные в таблице 6.2.

Таблица 6.2. Результаты тестирования.

Номер студента	1	2	3	4	5	6	7
y_i	4	9	3	3	8	6	9

Рассчитаем \bar{y} - среднюю арифметическую количества правильных оценок:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{4 + 9 + 3 + 3 + 8 + 6 + 9}{7} = 6.$$

Найдем S_y - стандартное отклонение количества правильных оценок:

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{(4-6)^2 \cdot (9-6)^2 \cdot (3-6)^2 \cdot (3-6)^2 \cdot (8-6)^2 \cdot (6-6)^2 \cdot (9-6)^2}{6}} \approx 2,7.$$

Рассчитаем коэффициент валидности:

$$V = \frac{\frac{\sum_{i=1}^n (Y_i \cdot y_i)}{n} - Y \cdot \bar{y}}{S_Y \cdot S_y} \cdot \frac{n}{n-1} \approx \frac{(5 \cdot 4) + (10 \cdot 9) + (3 \cdot 3) + (4 \cdot 3) + (7 \cdot 8) + (6 \cdot 6) + (9 \cdot 9)}{7} - \frac{6 \cdot 6}{2,6 \cdot 2,7}$$

Вывод: Валидность, рассмотренного в примере теста «Информационные процессы», высокая.

Контрольные вопросы:

1. Определите понятие «Валидность теста».
2. Классификация валидности.
3. Формула нахождения коэффициента валидности.
4. Роль эксперта в определении валидности теста.

7. Стандартизация теста и интерпретация результатов тестирования

Стандартизацией теста называется комплекс процедур, позволяющий создать для всех испытуемых равные условия, и подразумевает единообразный подход к процедуре проведения и оценивания результатов выполнения испытуемыми заданий [15].

Выделим ряд требований по организации тестирования:

1. Ни одному испытуемому не дается никаких преимуществ перед другими.
2. Группа тестируемых должна быть «выровнена по мотивации».
3. Тестирование различных групп испытуемых должно занимать равные временные промежутки и проводится при одинаковых внешних условиях. Если тест повторно используется в условиях, которые существенно отличаются от тех, где он был первоначально создан, то почти всегда требуется соотнесение этих условий друг с другом или адаптация теста к новым условиям.

4. Содержание теста должно соответствовать требованиям стандартов образования.

5. Все испытуемые выполняют одни и те же задания (параллельные, подобные формы заданий).

6. В тест включаются задания одной формы либо различных форм с соответствующими весовыми коэффициентами, значение которых получены статистическим путем.

7. Установление норм. Норма теста – это средний уровень развития большой совокупности людей, похожих на данного испытуемого по ряду социально-демографических характеристик. В большинстве случаев тестовый показатель индивидуума выявляется на основе сравнения с оценками, полученными по данному тесту другими людьми. Норма теста обычно определяется в результате тестирования большой выборки испытуемых определённого возраста и пола и усреднения полученных оценок с их последующим разделением по возрасту, полу и ряду других показателей. Эта группа людей называется выборкой стандартизации и служит для установления норм. Всякая норма обычно со временем изменяется вместе с естественными изменениями, поэтому есть правило, согласно которому нормы теста, особенно интеллектуального, должны пересматриваться, как минимум один раз в пять лет. Система подсчета баллов, на основе норм, должна быть предварительно разработана и применена ко всем ответам испытуемых без исключения. Стандартизация теста заключается в приведении процедуры оценок к общепринятым нормативам. Стандартизация предполагает преобразование нормальной или искусственно нормализованной шкалы первичных оценок в шкальные оценки. Тестовые нормы, полученные в ходе стандартизации, представляют собой систему шкал с характеристиками распределения тестового балла для различных выборок. Они не являются «внутренними» свойствами теста, а лишь облегчают его практическое применение.

8. Проверка тестовых работ должна быть строго регламентирована, а именно проверяющим даются эталоны правильных ответов и стандартизованные критерии оценок.

9. Тест должен быть социокультурно адаптирован т.е необходимо соблюдать соответствие тестовых заданий и оценок, которые испытуемый получает по этим заданиям, особенностям культуры, сложившимся в том или ином обществе, где данный тест используется, если он заимствован в другой стране.

10. Для обеспечения единообразия условий проведения теста, разработчик описывает подробные указания по проведению каждого вновь разработанного теста. Формулирование таких указаний – важная часть стандартизации нового теста. Она включает точные указания, относительно используемого стимульного материала, временных ограничений, устных инструкций испытуемому, пробных образцов заданий, допустимых ответов на вопросы обследуемого и других тонкостей проведения теста.

11. Учет «внешних факторов». Например, зачитывая вслух инструкцию или задание, нужно принимать во внимание тон голоса, скорость речи, интонацию, паузы и выражение лица.

12. Тест должен сопровождаться руководством для пользователя, в котором описываются:

- a) назначение теста и его описание;
- b) показания для применения;
- c) состав теста;
- d) информация об апробации теста (цели, объем и состав выборки, основные статистические характеристики);
- e) инструкция по процедуре проведения тестирования;
- f) ключи;
- g) трудность и дискриминативность теста («дискриминативность» - «тонкость измерения», т.е. способность дифференцировать тестируемых относительно «минимальных» и «максимальных» результатов теста, дифференцирующая способность);
- h) данные о надежности и валидности теста;
- l) другие статистические материалы;
- ж) правила и инструкции для обработки данных;
- к) устройство шкалы, правила и особенности интерпретации данных.

Таким образом, если тесты стандартизированы, то мы можем сказать что возможно провести сравнение показателей, полученные одним испытуемым, с таковыми в генеральной совокупности или соответствующих группах. Тем самым достигается адекватная интерпретация показателя отдельного испытуемого. Таким образом, стандартизация теста наиболее важна в тех случаях, когда осуществляется сравнение показателей обследуемых. При этом важна выработка нормы, или нормативных показателей. Для получения стандартных норм нужно тщательно отобрать большее количество испытуемых в соответствии с ясно обозначенным критерием.

Дискриминативность измеряется показателем дельта Фергюсона и принимает максимальное значение при равномерном (прямоугольном) распределении показателей ($d=1$).

При разработке теста необходимо стремиться к тому, чтобы его задания как можно тоньше измеряли тестируемое свойство. Например, если в результате обследования почти все испытуемые получают примерно одинаковые результаты, то это означает, что тест измеряет очень грубо. Чем большее количество градаций результатов можно получить при помощи теста, тем выше его разрешающая способность. Мера тонкости измерения (или степень дифференцируемости результатов) теста называется дискриминативностью. Дискриминативность теста измеряется показателем дельта Фергюсона:

$$\delta = \frac{n + 1(N^2 - \sum [f_i^2])}{nN^2},$$

где N – количество испытуемых, n – количество заданий, f_i - частота встречаемости каждого показателя. Наименьшая дискриминативность теста при $\delta = 0$, наибольшая при $\delta = 1$.

Рассмотрим простейший пример расчета индекса дискриминативности заданий.

Проводится тестирование по теме «Информационные процессы». В тестировании принимают участие 7 человек, количество заданий - 4.

Цель задания: овладение навыком расчета индекса дискриминативности.

Алгоритм вычислений:

1. Составьте таблицу частот встречаемости каждого показателя.
2. Подсчитайте, как часто встречаются значения показателей для данного теста.
3. Возведите эти числа в квадрат и просуммируйте их: Σf^2 .
4. Прибавьте единицу к количеству заданий.
5. Возведите в квадрат количество испытуемых.
6. Помножьте количество заданий на результат шага 4.
7. Теперь у нас есть все элементы формулы. Подставьте их и рассчитайте коэффициент.
8. Сделайте вывод о дискриминативности теста «Информационные процессы».

Оснащение: микрокалькулятор или компьютер, таблица первичных данных (Таблица 7.1).

Таблица 7.1. Первичные данные.

Испытуемый	Результат (в баллах)
Испытуемый 1	2
Испытуемый 2	1
Испытуемый 3	0
Испытуемый 4	3
Испытуемый 5	2
Испытуемый 6	1
Испытуемый 7	4

На основании таблицы первичных данных получим таблицу частот встречаемости каждого показателя (таблица 7.2).

Таблица 7.2. Таблица частот встречаемости показателей теста.

Баллы	0	1	2	3	4
f_i	1	2	2	1	1

Вычислим квадраты частот (таблица 7.3):

Таблица 7.3. Таблица квадратов частот.

Баллы	0	1	2	3	4	$\sum f_i^2$
f_i	1	2	2	1	1	
f_i^2	1	4	4	1	1	11

Произведем вычисления показателем дельта Фергюсона, где $n=4$ (количество заданий), $N=7$ (количество испытуемых), $N^2=49$:

$$\delta = \frac{n + 1(N^2 - \sum [f_i^2])}{nN^2} = \frac{(4 + 1)(49 - 11)}{4 \cdot 49} = \frac{190}{196} \approx 0,97$$

Вывод: $\delta = 0,97$ данный показатель указывает на высокую дискриминативность теста «Информационные процессы», так как наибольшая дискриминативность при $\delta = 1$. Показатель $\delta = 0,97$ приближается к единице.

Дискриминативность или дифференцирующая способность отражает способность отделять испытуемых с высоким общим баллом по тесту от тех кто получает низкий балл и позволяет проверить качество созданных заданий в тесте.

Для проверки качества отдельных заданий будем использовать метод крайних групп. Данный метод позволяет определить может ли задание дифференцировать учащихся. Задания на которое одинаково хорошо могут ответить испытуемые как с высокими, так и с низкими способностями, не обладает хорошей дифференцирующей способностью. Таким образом, цель использования на практике данного метода выражается в удалении некачественных заданий.

Для определения дифференцирующей способности задания будем использовать формулу метода крайних групп:

$$D = \frac{N_{nmax} - N_{nmin}}{N_{nmin}}$$

$$N_{nmax} = N_{nmin} \cdot 0,27$$

где $0,27$ — 27% от количества испытуемых,

N_{nmin} — количество испытуемых в группе худших верно выполнивших задание,

N_{nmax} — количество студентов в группе лучших верно выполнивших задание.

Интерпретация результата:

- а) если D в промежутке от 0,3 до 1 – задание эффективно;
- б) если D в промежутке от 0,1 до 0,3 – задание обладает низкой степенью дифференциации;
- с) если D меньше 0,1 – задание не качественное.

Рассмотрим простейший пример расчета показателя дифференцирующей способности по методу крайних групп.

Было проведено тестирование в группе из 30 испытуемых. Количество заданий- 10. Полученные результаты приведены в таблице частот встречаемости показателей (таблица 7.4). Цель: проверить качество задания № 1.

Таблица 7.4. Таблица частот встречаемости показателей теста.

Баллы	0	1	2	3	4	5	6	7	8	9	10
f_i	2	1	3	2	1	2	4	4	3	5	3

Определим 27% от количества испытуемых для создания крайних групп. Примерно получим 8 испытуемых. Таким образом, возьмем 8 человек набравших малое количество баллов (в нашем примере 0,1,2,3 балла) – слабая группа и 8 человек набравших наибольшее количество баллов (в нашем примере – 9,10 баллов) – сильная группа. Проверим, как эти испытуемые отвечали на задание № 1 (таблица 7.5).

Таблица 7.5. Результаты выполнения задания № 1 испытуемыми слабой и сильной групп.

Испытуемый	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Результат	0	1	0	0	0	0	0	0	1	1	1	0	1	1	1	1
Группа	Слабая								Сильная							

Приведем данные для подстановки в формулу метода крайних групп:

$$N_{max} = N_{min} \approx 8,$$

$$N_{nmin} = 1,$$

$$N_{nmax} = 7.$$

Подставим данные и получим следующий результат:

$$D = \frac{N_{nmax}}{N_{max} - N_{nmin}} = \frac{7}{8 - 1} = \frac{7}{7} = 1 = 0,75$$

Вывод: задание № 1 в тесте является эффективным.

Контрольные вопросы:

1. Определите понятие «стандартизация теста».
2. Перечислите требования по организации процедуры тестирования.
3. Норма теста.
4. Определите понятие «дискриминативность».
5. Алгоритм вычисления дискриминативности.

8. Матрица результатов тестирования

Базисом для вычислений характеристик качества теста является матрица результатов тестирования. В ней по вертикали располагаются профили ответов тестируемых на каждое задание теста, по горизонтали – результаты выполнения каждым тестируемым заданий теста в дихотомическом способе оценивания (таблица 8.1.). Согласно таблице 8.1 в тестировании состоящем из 10 заданий, приняло участие 11 испытуемых. Перед вычислением характеристик качества теста необходимо удалить крайние значения: профиль в котором правильно выполнены все задания (тестируемый № 10) и профиль в котором неправильно выполнены все задания (тестируемый № 11)

Таблица 8.1. Матрица результатов тестирования.

Номер тестируемого (i)	Номер задания (j)										Индивидуальный балл X_i
	1	2	3	4	5	6	7	8	9	10	
1	0	0	0	1	0	0	0	0	0	0	1
2	1	1	0	1	1	1	0	1	1	1	8
3	1	1	1	0	0	0	1	0	0	0	4
4	1	1	1	1	0	1	1	1	1	1	9
5	0	1	1	0	0	0	0	0	0	0	2
6	1	1	1	0	1	0	0	0	0	0	4
7	1	1	0	1	1	0	0	0	0	0	4
8	1	0	1	1	0	1	1	0	0	0	5
9	1	1	0	0	1	1	1	1	0	0	6
10	1	1	1	1	1	1	1	1	1	1	10 (удалить, ряд не считать)
11	0	0	0	0	0	0	0	0	0	0	0 (удалить, ряд не считать)
Количество правильных ответов (R_j) по каждому заданию (исключаем удаленные)	7	7	5	5	4	4	4	3	2	2	

ряды)											
-------	--	--	--	--	--	--	--	--	--	--	--

В таблице 8.1: i - количество тестируемых, j – количество заданий, X_i – индивидуальный балл i -го тестируемого, R_j – количество правильных ответов на j -е задание, x_{ij} – результат выполнения i -м тестируемым j -го задания:

$$x_{ij} = \begin{cases} 1, & \text{при правильном ответе } i \text{ – го тестируемого на } j \text{ – е задание} \\ 0, & \text{при неправильном ответе } i \text{ – го тестируемого на } j \text{ – е задание} \end{cases}$$

Для графической интерпретации необходимо упорядочить матрицу по индивидуальному баллу в порядке возрастания (см. таблицу 8.2).

Таблица 8.2. Отсортированная матрица результатов тестирования.

Номер тестируемого (i)	Номер задания (j)										Индивидуальный балл X_i
	1	2	3	4	5	6	7	8	9	10	
1	0	0	0	1	0	0	0	0	0	0	1
5	0	1	1	0	0	0	0	0	0	0	2
3	1	1	1	0	0	0	1	0	0	0	4
6	1	1	1	0	1	0	0	0	0	0	4
7	1	1	0	1	1	0	0	0	0	0	4
8	1	0	1	1	0	1	1	0	0	0	5
9	1	1	0	0	1	1	1	1	0	0	6
2	1	1	0	1	1	1	0	1	1	1	8
4	1	1	1	1	0	1	1	1	1	1	9
Количество правильных ответов (R_j) по каждому заданию	7	7	5	5	4	4	4	3	2	2	

Строим частотное распределение: представляем результаты в виде ранжированного ряда (см. таблицу 8.3), и частотное распределение (см. таблицу 8.4), при этом сумма всех частот равна количеству испытуемых (исключаем испытуемых с крайними значениями).

Таблица 8.3. Ранжированный ряд.

Ранг	1	2	3	3	3	4	5	6	7
Номер	1	5	3	6	7	8	9	2	4
Балл	1	2	4	4	4	5	6	8	9

Таблица 8.4. Частотное распределение.

Балл	1	2	4	5	6	8	9
Частота	1	1	3	1	1	1	1

Строим гистограмму по частотному распределению (см. рисунок 8.1).

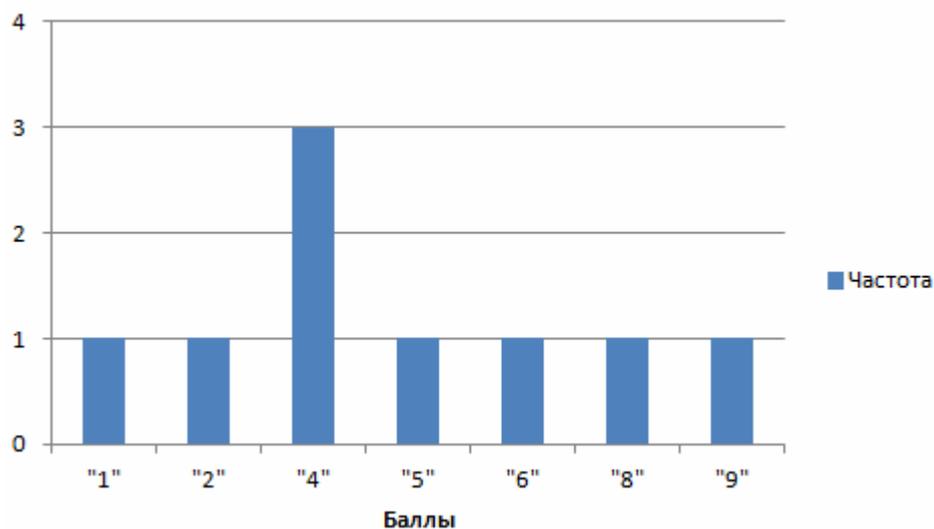


Рисунок 8.1. Гистограмма частотного распределения.

Строим гистограмму по сгруппированному частотному распределению (см. таблицу 8.5., рисунок 8.2).

Таблица 8.5. Сгруппированное частотное распределение.

Интервал баллов	Частота
1-3	2
4-6	5
7-9	2

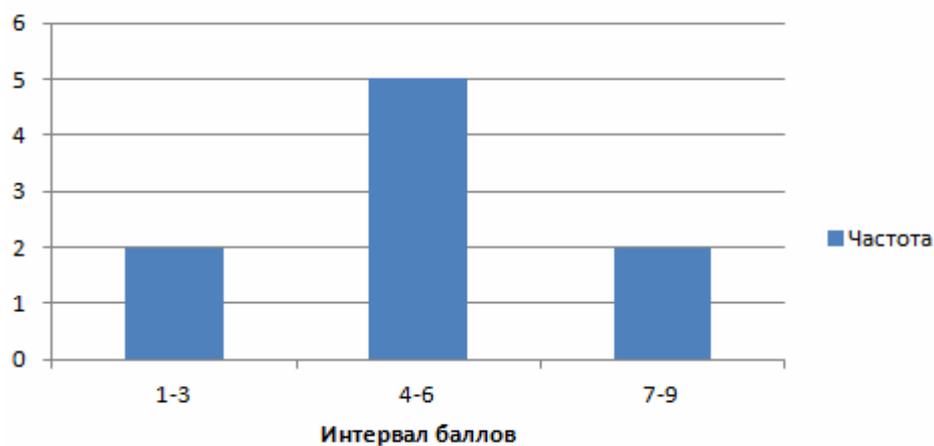


Рисунок 8.2. Гистограмма сгруппированного частотного распределения.

Показатель качества теста – нормальное распределение результатов.. Нормальное распределение, также называемое распределением Гаусса — распределение вероятностей, которое в одномерном случае задается функцией плотности вероятности, совпадающей с функцией Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

где параметр μ — математическое ожидание (среднее значение, константа, задающая сдвиг распределения по оси X), медиана и мода распределения, а параметр σ — среднеквадратическое отклонение (σ^2 — дисперсия) распределения (см. рисунок 8.3.).

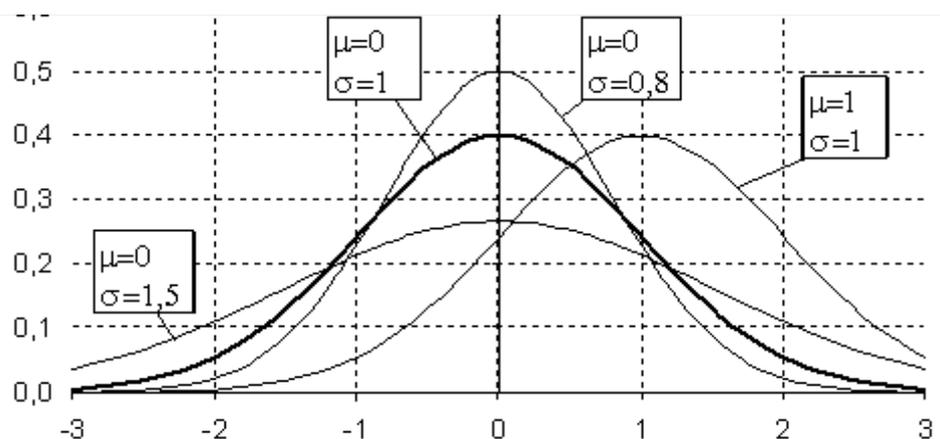


Рисунок 8.3. Нормальное распределение.

Мода – это наиболее часто встречающееся значение среди результатов выполненного теста. В рассматриваемом нами примере мода равна 4, так как этот результат встречается чаще других – 3 раза.

Если одинаково часто встречаются два значения (т.е. имеются две моды), то такое распределение называют бимодальным.

Нормальное распределение результатов всегда унимодальное (т.е. с одним значением моды) и симметричное. Бимодальное распределение говорит о неудачно построенном тесте.

Наш пример унимодальный, распределение симметричное относительно моды, мы можем сделать на основании этих данных о удачно построенном тесте.

Хороший нормативно-ориентированный тест (который может ранжировать испытуемых между собой) обеспечивает нормальное распределение индивидуальных баллов репрезентативной выборки тестируемых, если среднее значение баллов \bar{X} находится в центре

распределения, остальные баллы концентрируются вокруг примерно по следующему соотношению: 68% - частот баллов образуют центр, по 16% частот баллов сходятся по краям (рис.). Если происходит смещение \bar{X} влево то можно сделать вывод о том, что задания слишком сложны, смещение вправо говорит о легкости заданий.

Для анализа разброса эмпирических данных вокруг среднего используют дисперсию и стандартное отклонение – меры разброса.

Дисперсия отражает меру неоднородности результатов по тесту:

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}, \text{ где } X_i - \text{балл } i\text{-го испытуемого, } \bar{X} - \text{среднее}$$

значение баллов, n – количество испытуемых.

Вычислим для рассматриваемого нами примера:

Баллы испытуемых – 1,2,4,4,4,5,6,8,9.

$$n=9, \bar{X} = \frac{1 + 2 + 4 + 4 + 4 + 5 + 6 + 8 + 9}{9} \approx 4,8.$$

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{(1 - 4,8)^2 + (2 - 4,8)^2 + (4 - 4,8)^2 + (4 - 4,8)^2 + (4 - 4,8)^2 + (5 - 4,8)^2 + (6 - 4,8)^2 + (8 - 4,8)^2 + (9 - 4,8)^2}{8}$$

Низкая дисперсия говорит о слабом разделении испытуемых по уровню знаний, не позволяет с приемлемой точностью ранжировать их. Излишне высокая дисперсия указывает на сильную неоднородность группы испытуемых, на возможные нарушения процедуры тестирования, на недостаточно ясные формулировки заданий и т.п.

Дисперсия имеет среднее значение (наш диапазон от 1 до 9, со средней величиной 2,6), что указывает о качестве теста.

Чаще для анализа используют стандартное отклонение, равное корню из дисперсии. М.Б.Челышкова пишет, что если среднее арифметическое примерно равно утроенному стандартному отклонению $\bar{X} = 3 \cdot \sigma$, то можно считать дисперсию оптимальной, а распределение тестовых баллов близким к нормальному [11].

$$\sigma = \sqrt{2,6} \approx 1,61, \\ 3 \cdot \sigma = 3 \cdot 1,61 = 4,83 \approx \bar{X}.$$

На основании произведенных расчетов, можно сделать вывод о том, что распределение тестовых баллов близко к нормальному.

На основании матрицы результатов тестирования можно вычислить коэффициент точно-бисериальной корреляции, являющийся показателем валидности задания:

$$r_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_x} \sqrt{\frac{(N_1)_j \cdot (N_0)_j}{N(N - 1)}}$$

где, r_j - коэффициент точечно-бисериальной корреляции определяющий валидность j-го задания теста, $(\bar{X}_1)_j$ – среднее значение индивидуальных баллов тестируемых, верно выполнивших j-е задания теста, $(\bar{X}_0)_j$ - среднее значение индивидуальных баллов тестируемых, неверно выполнивших j-е задания теста, S_x - стандартное отклонение по множеству значений индивидуальных баллов, N – общее количество тестируемых, $(N_1)_j$ - количество тестируемых, верно выполнивших j-е задания теста, $(N_0)_j$ - количество тестируемых, неверно выполнивших j-е задания теста.

Замечание:

1. Расчетное значение находится в диапазоне от -1 до +1.
2. Результат вычисления определит, насколько задание пригодно для измерения в соответствии с общей целью построения темы.
3. О валидности задания символизирует приближения коэффициента к 0,5.
4. Рекомендуется оценить $(\bar{X}_1)_j - (\bar{X}_0)_j$, чем выше значение этой разности, тем лучше задание работает по дифференциации испытуемых.

Рассмотрим пример, нахождения коэффициента точечно-бисериальной корреляции. Результаты тестирования приведены в таблице 8.6. Установим валидность задания № 3.

Таблица 8.6. Матрица результатов тестирования.

Номер тестируемого (i)	Номер задания (j)										Индивидуальный балл X_i
	1	2	3	4	5	6	7	8	9	10	
1	0	0	0	1	0	0	0	0	0	0	1
5	0	1	1	0	0	0	0	0	0	0	2
3	1	1	1	0	0	0	1	0	0	0	4
6	1	1	1	0	1	0	0	0	0	0	4
7	1	1	0	1	1	0	0	0	0	0	4
8	1	0	1	1	0	1	1	0	0	0	5
9	1	1	0	0	1	1	1	1	0	0	6
2	1	1	0	1	1	1	0	1	1	1	8
4	1	1	1	1	0	1	1	1	1	1	9
Количество правильных ответов (R_j) по каждому заданию	7	7	5	5	4	4	4	3	2	2	

Решение: при $j=3$, $N=9$, $S_x=2,6$, $(N_0)_3 = 3$, $(N_1)_3 = 5$

$$(\bar{X}_1)_3 = \frac{2 + 4 + 4 + 5 + 9}{5} = 4,8$$

$$(\bar{X}_0)_3 = \frac{1 + 4 + 6 + 8}{4} = 4,75$$

$$r_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_x} \sqrt{\frac{(N_1)_j \cdot (N_0)_j}{N(N-1)}} = \frac{4,8 - 4,75}{2,6} \sqrt{\frac{5 \cdot 4}{9 \cdot 8}} \approx 0,02 \cdot 0,5 \approx 0,1$$

Вывод: Валидность задания № 3 равно 0,1, что говорит о низкой валидности данного задания, задание в тесте необходимо изменить, оно не отвечает поставленным целям.

Контрольные вопросы:

1. Для каких целей используется матрица результатов тестирования?
2. Опишите алгоритм построения гистограммы частотного распределения.
3. Мода. Умодальное и бимодальное распределение.
4. Смысл дисперсии как показателя качества теста.
5. Вычисление коэффициента точечно-бисериальной корреляции. Его назначение?

9. Использование программных средств

Использование элементов автоматизации значительно упрощает процесс проведения тестирования и обработки результатов: тестирование становится максимально быстрым по времени, уменьшаются накладные расходы, появляется возможность увеличения количества тестируемых за единицу времени, уменьшается время анализа результатов.

В форме компьютерных программ тесты впервые стали применяться в рамках первых систем автоматизированного обучения, разрабатываем в соответствии с концепцией программированного обучения [2, 4,5, 15].

Программированное обучение – метод обучения, выдвинутый профессором Б.Ф. Скиннером в 1954 г. Цель данного метода заключается в

том, что на базе кибернетического подхода можно повысить эффективность управления процессом обучения. В своей основе, программированное обучение, подразумевает работу слушателя по некой программе, в процессе выполнения которой, он овладевает знаниями. Роль преподавателя сводится к отслеживанию и регулированию программных действий.

Отметим преимущества компьютерного тестирования:

1. Обеспечение стандартизации.
2. Обеспечение индивидуальности процедуры контроля. Принцип индивидуальности лежит в основе адаптивного тестирования. Адаптивное тестирование это контроль, который позволяет регулировать трудность и число предъявляемых заданий каждому обучаемому в зависимости от его ответа на предыдущее задание: в случае правильного ответа он получит более сложное задание, в случае неправильного более легкое.
3. Повышение объективности контроля и исключение субъективных факторов (усталость преподавателя, его эмоциональность или плохое настроение, негативное или излишне позитивное отношение к личности испытуемого).
4. Оперативность статистической обработки результатов контроля.
5. Прозрачность процесса контрольного мероприятия.
6. Доступ испытуемых к полной информации и результатом проведенного контрольного мероприятия.
7. Освобождение преподавателя от выполнения повторяющейся трудоемкой и рутинной работы по организации контрольного мероприятия (заготовка бланков, инструктаж тестируемого, выдача заданий, ведение протокола) и проверки его результатов.
8. Обеспечение всесторонней и полной проверки.
9. Доступность и равноправие всех участников процедуры тестирования.
10. Тестируемому, некого стесняться – компьютер не может ни оценочно, ни эмоционально реагировать на не самые удачные ответы.

Выделим недостатки компьютерного тестирования:

1. Исключение из процедуры контроля речевого компонента.
2. Монотонность работы.
3. Отсутствие полноценной обратной связи.
4. Негативные реакции обычно вызывают различные ограничения которые иногда накладываются при выполнении заданий (время, возможность возврата к заданию, возможность исправления уже пройденного задания)
5. Не всегда учитывается уровень компьютерной грамотности испытуемых.
6. Нет полноценной возможности контролировать выполнение творческих заданий.

7. При анализе результатов компьютерного тестирования эксперты имеют дело только с полученными ответами испытуемого. Они не видят тестируемого, не общаются с ним, а значит, не владеют о нем дополнительной информацией, не могут выяснить его действительный объем знаний.

8. Возрастает вероятность ошибочных ответов (можно неверно понять инструкцию, перепутать клавиши для ответов и др.).

Основные задачи программных средств компьютерного тестирования:

1. Хранение тестовых заданий и создание на их основе тестов.
2. Регистрация и инструктаж испытуемых.
3. Проведение сеанса тестирования.
4. Обработка ответов и подсчет итоговых баллов. При этом обработка может быть:
 - 4.1. Локальной, т.е. выполняемой на месте тестирования.
 - 4.2. Удаленной, т.е. осуществляемой за пределами места проведения сеанса тестирования.
 - 4.3. Формальной, если возможно простое сравнение с ключом.
 - 4.4. Экспертной, если для проверки требуется привлечение экспертов.
 - 4.5. Оперативной, т.е. позволяющей продемонстрировать результаты сразу же после тестирования.
 - 4.6. Отложенной, т.е. получение результата удалено по времени по причине сложного алгоритма расчета баллов или необходимости получения заключения экспертов.

Программные средства для контроля и измерения уровня знаний, умений и навыков испытуемых выделены в отдельный вид образовательных электронных изданий.

При проектировании систем компьютерного тестирования придерживаются следующих правил:

1. Система желательно должна обеспечивать выполнение полного цикла разработки, апробации и практического использования теста.
2. Система должна обеспечивать одновременную работу нескольких (более 10) испытуемых.
3. Система должна делать выводы, т.е. проводить анализ статистических характеристик, обобщенный анализ по группе, индивидуальный анализ по испытуемым на основе тестирования разнесенного по времени.
4. Система должна обеспечивать комплектование различных тестов с минимальными трудозатратами со стороны педагогов.
5. В системе должны быть предусмотрены средства, уменьшающие вероятность намеренного искажения результатов (списывание, подбор правильного ответа, угадывание, подмена данных).

6. Желательно максимально снизить количество операций по обеспечению функционирования системы на рабочих местах (минимальный алгоритм действий для испытуемого).

7. Целесообразно спроектировать систему разработки и проведения тестирования в виде веб-приложений на основе серверного языка сценариев и базы данных. В этом случае клиентской программой для работы с оболочкой может служить стандартный браузер (отпадает необходимость установки дополнительных клиентских приложений).

8. Для обеспечения работы целесообразно включить в систему следующие функциональные модули:

8.1. Модуль тестолога (преподавателя): ввод и коррекция вопросов, создание и коррекция тестов, анализ результатов прохождения тестов, средства интерактивного взаимодействия с испытуемым (пересылка сообщений во время работы испытуемого с клиентской частью: предупреждения, комментарии, инструкции, объявление результатов).

8.2. Модуль испытуемого (тестируемого): выбор и прохождение тестов, просмотр результатов, средства интерактивного общения с преподавателем: вопросы о содержании и полученном результате.

8.3. Модуль административных операций (управление регистрационными данными тестологов и тестируемых, выполнение служебных и технических операций, контроль корректности использования системы).

8.4. В системе должны различаться вопросы (задания), тесты (сформированные комплекты), результаты и служебные данные.

9. Желательно предусмотреть назначение теста тестируемому или группе тестируемых (в случае итогового, рубежного или тематического теста). Диагностические и обучающие тесты могут быть доступны для использования испытуемыми в любое время.

10. Желательно экспорт данных о результатах тестирования во внешние программные приложения, например в электронные таблицы или подобные редакторы.

Эргономические требования к программам компьютерного тестирования строятся, учитывая в первую очередь возрастные различия испытуемых, устанавливают требования к изображению и подаче информации, режимам работы с ней, и по должны обеспечить повышение учебной мотивации к обучению [9]

Основным эргономическим требованием становится задача обеспечение в высшей степени гуманного отношения к испытуемому, а значит, организация дружественного интерфейса, обеспечение возможностей использования испытуемым всех необходимых в процессе взаимодействия с интерфейсом системы подсказок и методических указаний. Так же сюда мы можем отнести соответствие цветового колорита, упорядоченность и

выразительность элементов интерфейса. К требованиям отнесем соблюдение принципов адаптивности (приспособления к индивидуальным особенностям) и интерактивности (соблюдение режима двусторонней связи).

Эргономические требования очень тесно связаны с двумя группами требований: требования здоровьесберегающего характера и эстетическими требованиями. Говоря о первой группе, отметим, что программы компьютерного тестирования должны разрабатываться таким образом, чтобы их использование не противоречило санитарным нормам работы с вычислительной техникой. Вторая группа связана с тем, что необходимо установить соответствие эстетического оформления функциональному назначению; определить порядок, выразительность графических объектов учебно-познавательной среды; определить отсутствие орфографических, пунктуационных ошибок и неэтичных компонентов. Эстетические свойства должны характеризовать способность программного продукта через чувственно воспринимаемые образы удовлетворить потребность в эстетическом идеале. Должны присутствовать: информационная выразительность, рациональность форм, целостность композиции.

Контрольные вопросы:

1. Цель метода программированного обучения.
2. Перечислите преимущества компьютерного тестирования.
3. Выделите недостатки компьютерного тестирования.
4. Основные задачи программных средств компьютерного тестирования.
5. Правила проектирования систем компьютерного тестирования.
6. Эргономические требования к программам компьютерного тестирования.

Заключение

В современном образовании широко применяется тестирование, представляющее собой особую форму стандартизированных, кратких, ограниченных во временном диапазоне испытаний, предназначенных для установления количественных и качественных индивидуальных различий.

Относя педагогическое тестирование к инновационным методам оценивания, мы выделили следующие его преимущества:

1. Более высокая объективность контроля.
2. Оценка знаний становится более дифференцированной (больше градаций оценки, возможность переводить полученную оценку в любую шкалу оценивания).
3. Более высокая эффективность (можно проводить контроль одновременно на больших группах учащихся).
4. Обработка результатов проводится быстрее (особенно в случае компьютерного тестирования).

Таким образом, тест — это метод и инструмент педагогического измерения. При этом, если говорить о тесте, как о методе, заметим, что он предполагает особую технологию измерения, которая включает в свой состав правила выполнения операций по разработке системы тестовых заданий с заранее заданным набором качественных и количественных характеристик, с целью объективного и надежного оценивания учебных достижений испытуемых. Если же говорить о тесте как о педагогическом инструменте, то необходимо заметить, что здесь он выступает уже в форме системы заданий специфической формы, которая позволяет качественно и полно оценить структуру и эффективно измерить существующий уровень знаний, умений и навыков учащегося.

Практикум

Задание 1: Создайте тест по дисциплине «Информатика и ИКТ» согласно требованиям по созданию тестов закрытого типа.

1. Создайте тест закрытого типа в текстовом редакторе (не менее 10 вопросов)
2. Создайте автоматизированный тест закрытого типа в электронных таблицах (не менее 10 вопросов, вычисление итоговой оценки).

Задание 2: Создание теста по дисциплине «Информатика и ИКТ» согласно требованиям по созданию тестов открытого типа.

Задания:

1. Создайте тест открытого типа (в текстовом редакторе, не менее 10 вопросов, задание свободного конструирования с критериями оценки для эксперта)
2. Создайте автоматизированный тест открытого типа (в электронных таблицах, не менее 10 вопросов, задания с дополнением, вычисление итоговой оценки)

Задание 3: Расчет валидности теста. Тест «Системы счисления» должна выполнить группа студентов из 10 человек. Тест состоит из 8 заданий.

Экспертом, предварительно оцениваются результаты. Результаты экспертного оценивания приведены в таблице 1.

Таблица 1. Результаты экспертного оценивания.

Номер студента	1	2	3	4	5	6	7	8	9	10
Y_i	8	7	3	4	1	3	2	4	8	2

После проведения теста, были получены результаты, приведенные в таблице 2.

Таблица 2. Результаты тестирования.

Номер студента	1	2	3	4	5	6	7	8	9	10
Y_i	6	8	3	2	2	3	2	5	8	1

Задание 4: Расчет индекса дискриминативности заданий. Проводится тестирование по теме «Системы счисления». В тестировании принимают участие 9 человек, количество заданий - 5. Первичные данные приведены в таблице 3.

Таблица 3. Первичные данные.

Испытуемый	Результат (в баллах)
Испытуемый 1	5
Испытуемый 2	3
Испытуемый 3	0
Испытуемый 4	4
Испытуемый 5	2
Испытуемый 6	1
Испытуемый 7	3
Испытуемый 8	5
Испытуемый 9	4

Задание 5: Расчета показателя дифференцирующей способности по методу крайних групп.

Было проведено тестирование в группе из 30 испытуемых. Количество заданий- 10. Полученные результаты приведены в таблице частот встречаемости показателей (см. таблицу 4) и таблице результатов выполнения задания № 1 испытуемыми слабой и сильной групп (см. таблицу 5). Цель задания: проверить качество тестового задания № 1.

Таблица 4. Таблица частот встречаемости показателей теста.

Баллы	0	1	2	3	4	5	6	7	8	9	10
f_i	3	2	3	2	2	1	4	5	1	4	3

Таблица 5. Результаты выполнения задания № 1 испытуемыми слабой и сильной групп.

Испытуемый	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Результат	0	1	0	0	0	1	1	0	1	1	1	0	0	0	1	1
Группа	<i>Слабая</i>								<i>Сильная</i>							

Задание 6: Создайте тест (любая тема) по дисциплине «Информатика и ИКТ» на сайте <http://www.banktestov.ru/>. Требования: количество вопросов не менее пяти, использовать различные формы тестовых заданий).

Задание 7: Зарегистрируйтесь в сервисе <http://master-test.net/>. Создайте тест (любая тема) по дисциплине «Информатика и ИКТ». Активируйте тест. Опишите возможности сервиса по созданию класса студентов с отслеживанием их работы по вашему тесту.

Задание 8: Создайте бланк теста (5 вопросов разных видов) и бланк с ответами в следующем сервисе: <https://www.easymaker.com/>. В бесплатной версии создает бланочный вариант теста с бланком ответов. Примечание: После создания теста, выберите Print & Download.

Задание 9: Опишите следующие среды (назначение, технологические особенности, возможности)

1. «Система тестирования студентов» предназначена для проверки уровня знаний студентов.

http://vlsi.ru/Rus/Project_groups/Education_Innovations.htm

2. Система тестирования INDIGO

<http://indigotech.ru/>

3. Программный комплекс «Студия визуального тестирования»

<http://www.mmis.ru/Default.aspx?tabid=157>

4. Программа x-TLS

<http://xtls.org.ua/>

Задание 10: Дана матрица тестовых результатов (см. таблицу 6). Постройте гистограммы частотного, распределения, сгруппированного частотного распределения. Найдите моду, определите является ли распределение бимодальным или унимодальным. Найдите среднее значение, дисперсию, стандартное отклонение, коэффициенты точечно-бисериальной корреляции. Сделайте выводы.

Таблица 6. Матрица результатов тестирования.

Номер тестируемого (i)	Номер задания (j)									
	1	2	3	4	5	6	7	8	9	10
1	0	1	1	0	0	0	0	0	0	0
2	1	1	0	0	0	1	0	0	0	0
3	1	1	1	0	0	1	1	1	0	0
4	1	1	1	1	0	1	0	0	0	0
5	0	1	1	1	1	1	1	0	0	0
6	1	1	1	0	1	1	1	1	1	0
7	1	1	0	1	1	1	0	1	1	1
8	1	0	1	1	0	1	1	0	0	0
9	1	1	0	0	1	1	1	1	0	0
10	1	1	1	0	1	0	1	1	1	1
11	1	1	0	0	1	0	0	0	0	0

Вопросы для самостоятельного изучения

1. Сравните традиционные и новые средства оценки результатов обучения. Опишите их достоинства и недостатки.
2. Как оценивается качество российского образования отечественными и зарубежными экспертами?
3. Назовите современные центры тестирования.
4. Назовите функции контроля в современном учебном процессе.
5. Какие современные средства контроля выделяются в учебном процессе? В чем их преимущество по сравнению с традиционными формами контроля?
6. Раскройте возможности адаптивного компьютерного тестирования.
7. Какими способами определяется надежность теста?
8. Как производится экспертиза качества содержания теста?
9. Какие погрешности возможны при педагогических измерениях. Как определяются и оцениваются погрешности при измерении?
10. Назовите и охарактеризуйте основные модели педагогических измерений.
11. Расскажите о видах оценочных шкал. Дайте характеристику одному из пакетов прикладных программ по обработке или конструированию тестов.
12. Расскажите о порядке создания контрольно-измерительных материалов (КИМ-ов) для ЕГЭ.

Рекомендуемые интернет-ресурсы:

1. Базы нормативных документов
 - каталог образовательных ресурсов – www.edu.ru
 - министерство образования РФ – www.ed.gov.ru
 - ГосНИИ информационных технологий и телекоммуникации – www.informika.ru
 - портал Единого Экзамена – ege.edu.ru
 - Августовский педсовет – www.pedsovet.alledu.ru
 - Справочные правовые системы <http://www.consultant.ru/>
2. Образовательные ресурсы
 - портал «Учеба» - www.uroki.ru, www.posobie.ru, www.metodiki.ru
 - сайт энциклопедии Кирилла и Мефодия – www.km.ru
 - «Школьный сектор» - www.school-sector.relarn.ru
 - виртуальная школа Кирилла и Мефодия – vschool.km.ru
 - научная лаборатория школьников - www.nsu.ru/materials/ssl

- «Школьный мир» - school.holm.ru
 - «Школы в Интернет» - schools.techno.ru
 - Материалы для слушателей и выпускников программы Intel® «Обучение для будущего», посвященные современным технологиям и средствам оценивания результатов обучения
http://www.iteach.ru/met/index_student.php
 - Московский центр непрерывного математического образования - www.mcsme.ru
 - Рефераты, тесты, новости образования - www.5ballov.ru
 - Сайт «Все образование Интернета» - www.alledu.ru
 - Сетевое объединение методистов СОМ – сетевое объединение методистов: планы, программы, разработки уроков по предметам - <http://som.fio.ru/>
 - Сайт журнала «Вопросы Интернет Образования» - <http://center.fio.ru/vio>
 - Информационно-образовательный портал «Гуманитарные науки» - www.auditorium.ru
 - Конкурс образовательных ресурсов - <http://konkurs.auditorium.ru/konkurs/>
3. Печатные издания
- «Учительская газета» - www.ug.ru
 - «Первое сентября» - www.1september.ru
 - «Домашний компьютер» - www.homepc.ru
 - «Компьютер-Пресс» - www.compress.ru
 - «Мир ПК» - www.osp.ru/pcworld

Литература:

1. Аванесов В. С. Основные понятия педагогической тестологии//Научные проблемы тестового контроля знаний: Тез. докл. участников школы-семинара. М., 2005
2. Аванесов В.С. Научные основы тестового контроля знаний. М.: Исследовательский центр, 1994. – 135 с.
3. Аванесов В.С. Тесты: история и теория // Управление школой, 1999, №12.
4. Аванесов В.С. Формы тестовых заданий: учебное пособие для учителей школ, лицеев, преподавателей вузов и колледжей. 2-е изд. перераб. и расширен. — М.: Центр тестирования, 2005. — 156 с.
5. Анастаси А., Урбина С., Алексеев А.А. Психологическое тестирование - Санкт-Петербург, 2007. Сер. Мастера психологии (7-е международное издание)
6. Векслер В.А. Особенности проектирования нестандартных тестовых заданий// В.А. Векслер, Л.Б. Рейдель - NovaInfo.Ru. 2015. Т. 1. № 35. С. 204-209.
7. Векслер В.А. Психолого-педагогические аспекты тестирования// В.А. Векслер, О.Л. Коноваленко - NovaInfo.Ru. 2015. Т. 1. № 35. С. 199-204.
8. Векслер В.А. Тестовые задания закрытого типа //NovaInfo.Ru. 2015. Т. 2. № 34. С. 169-173.
9. Векслер В.А. Эргономические требования к электронным образовательным ресурсам // Психология, социология и педагогика. 2015. № 5 (44). С. 37-39.
10. Векслер В.А. Возникновение тестологии //Современные научные исследования и инновации. 2015. № 5-4 (49). С. 113-116.
11. Ефремова Н.Ф., Звонников В.И., Челышкова М.Б. Педагогические измерения в системе образования //Педагогика. 2006. - № 2. - С. 14-22.
12. Клайн П. Введение в психометрическое проектирование. Справочное руководство по конструированию тестов. –Киев: ПАН Лтд, 1994. -184 с.
13. Майоров А.Н. Теория и практика создания тестов для системы образования. – М.: «Интеллект-центр», 2001. -296 с.
14. Равен Джон Педагогическое тестирование: Проблемы, заблуждения, перспективы / Пер. с англ. - М.: "Когито-Центр", 1999.-144 с.
15. Самылкина Н.Н. Современные средства оценивания результатов обучения: учебное пособие - М.:Бином. Лаборатория знаний, 2012. - 197 с.