

СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

УЧЕБНОЕ ПОСОБИЕ
ДЛЯ СТУДЕНТОВ
ПЕДАГОГИЧЕСКИХ
НАПРАВЛЕНИЙ
ПОДГОТОВКИ



КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

**СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ
РЕЗУЛЬТАТОВ ОБУЧЕНИЯ**

**Учебное пособие для студентов
педагогических направлений подготовки**

Казань – 2020

УДК 378.4
ББК 74.202.8
С56

*Принято на заседании Учебно-методической комиссии Института
психологии и образования КФУ
Протокол №6 от 26 февраля 2020 года*

Рецензенты:

доктор педагогических наук, профессор кафедры педагогики Института
психологии и образования КФУ

Р. А. Валеева;

доктор педагогических наук, профессор кафедры педагогики высшей школы
Института психологии и образования КФУ

А. Р. Масалимова

С56 Современные средства оценивания результатов обучения: Учебное пособие для студентов педагогических направлений подготовки / В.К.Власова, В.Г.Закирова, С.Г.Григорьева, Л.Р.Каюмова, Э.Г.Сабирова. – Казань: Вестфалика, 2020. – 136 с.

Пособие по курсу «Современные средства оценивания результатов обучения» направлено на углубление знаний о методах и средствах педагогического контроля. Знакомит с теоретическими и методическими подходами конструирования педагогических тестов. Каждая тема лекционного материала закрепляется вопросами для самопроверки и заданиями к теме. Изучение дисциплины позволяет студентам в дальнейшем успешно осваивать содержание других дисциплин психолого-педагогического цикла в процессе обучения. Пособие предназначено для студентов педагогических направлений подготовки.

© Власова В.К., Закирова В. Г., Григорьева С.Г.,
Каюмова Л. Р., Сабирова Э.Г., 2020

Содержание

Тема 1. Оценка и качество образования	6
1.1. Понятие о качестве образования	6
1.2. Оценка как элемент управления качеством учебного процесса	8
1.3. Мониторинг качества образования	11
1.4. Педагогический контроль. Функции и принципы педагогического контроля.	15
1.5. Традиционные виды, методы и формы контроля	17
1.6. Современные методы и формы педагогического контроля (тестовый, модульно-рейтинговый)	21
1.7. Папка индивидуальных учебных достижений «портфолио»	25
1.8. Оценка: понятие, функции, критерии	29
Вопросы для самоконтроля	32
Практические задания и упражнения по теме	33
Тема 2. Теория тестирования: общие положения	34
2.1. Наука о тестах: pro & contra	34
2.2. Исторические предпосылки современного тестирования в отечественном образовании	39
Вопросы для самоконтроля	46
Практические задания и упражнения по теме	47
Тема 3. Педагогический тест как система заданий	48
3.1. Основные положения классической теории тестов	48
3.2. Обзор основных понятий общей теории педагогических измерений	53
3.3. Этапы формирования теста	56
3.4. Классификация теорий педагогических измерений по В.С. Аванесову	59
3.5. Классификация тестов и тестовых заданий	61
Вопросы для самоконтроля	75
Практические задания и упражнения по теме	75

Тема 4. Характеристики педагогического теста и методика их расчета..	77
4.1. Принципы отбора и критерии оценки содержания теста	77
4.2. Принципы отбора и критерии оценки содержания теста	95
4.3. Алгоритм расчета характеристик педагогического теста.....	105
4.3. Алгоритм расчета характеристик педагогического теста.....	107
4.3.1. Расчет надежности теста	107
4.3.2. Расчет валидности теста.....	108
Вопросы для самоконтроля.....	112
Практические задания и упражнения по теме	112
Вопросы для итоговой аттестации	114
Информационные источники.....	115
Приложения	118

ТЕМА 1. ОЦЕНКА И КАЧЕСТВО ОБРАЗОВАНИЯ

Перечень вопросов, рассматриваемых в теме:

- 1. Понятие о качестве образования*
- 2. Оценка как элемент управления качеством учебного процесса*
- 3. Мониторинг качества образования*
- 4. Педагогический контроль. Функции и принципы педагогического контроля*
- 5. Традиционные виды, методы и формы контроля*
- 6. Современные методы и формы педагогического контроля: тестовый, модульно-рейтинговый*
- 7. Папка индивидуальных учебных достижений «портфолио»*
- 8. Оценка: понятие, функции, критерии*

1.1. Понятие о качестве образования

В любой сфере деятельности человек стремится достичь максимально возможного качества этой деятельности. В настоящее время особо остро стоит проблема качества образования. Однако до сих пор не существует однозначного определения понятия «качество образования». Обсуждению этого понятия Международным институтом планирования образования были посвящены две международные конференции в 1969 и 1978 гг. Особые споры вызвало одновременное использование и противопоставление понятий «качество» и «количество».

В широком смысле под качеством образования понимается соотношение цели и результата, мера достижения целей при том, что цели (результаты) заданы только операционально и спрогнозированы в зоне потенциального развития обучаемого.

Иначе говоря: образование, полученное школьником, признается качественным, если его результаты соответствуют операционально заданным

целям и спрогнозированы в зоне потенциального развития ученика. При этом результаты образования обязательно должны включать в себя и оценки того, какой ценой эти результаты достигнуты.

Из этого определения следует, что необходимо измерять в одинаковых единицах: цель, поставленную перед учебным заведением; результат, достигнутый этим учебным заведением. В связи с этим образовательное учреждение, школа должна:

- 1) научиться прогнозировать, проектировать, моделировать те качества подготовки выпускника, которые предполагается получить на «выходе» образовательного процесса;
- 2) обеспечивать достижение требуемого качества образования;
- 3) повышать качество образования, приводя его в соответствие с растущими требованиями внешних заказчиков;
- 4) выявлять и оценивать реальное качество образования, его соответствие стандартам.

Однако для практических целей под качеством образования решили понимать «качественные изменения» в учебном процессе и в среде, окружающей обучаемого, которое можно идентифицировать как улучшение знаний, умений и ценностей, приобретаемых обучаемым по завершению определенного этапа.

Таким образом, понятие «качество образования» можно сузить до понятия «качество подготовки». Обсуждение понятия «качество подготовки» можно вести на трех различных уровнях:

1-й уровень – планирование обучения, когда определенные представления о планируемом качестве подготовки закладывается в образовательные программы по каждому предмету;

2-й уровень – этап реализации образовательных программ в учебном процессе. В зависимости от обстоятельств, внешнего окружения и качества

работы педагога можно говорить об определенном реализуемом качестве подготовки учеников и студентов;

3-й уровень – качество подготовки обучаемых рассматривается как результат осуществления учебного процесса. Этот уровень рассмотрения качества подготовки обычно в зарубежных исследованиях отождествляется с термином «учебные достижения». Отечественные же исследования подвергают сомнению такой подход, так как представление о качестве подготовки может ассоциироваться с внутренним состоянием обучаемого, в то время как результаты обучения проявляются во внешних наблюдаемых признаках и результатах учебного процесса. Однако это противоречие снимается при использовании специальных моделей измерения, соединяющих латентные параметры испытуемых с наблюдаемыми результатами выполнения заданий.

1.2. Оценка как элемент управления качеством учебного процесса

Целью создания системы оценки качества является создание условий повышения качества образования через усиление потоков информации о результатах обучения и соответствующих факторах.

Под «качеством» учебного процесса (в соответствии со вторым и третьим уровнями обсуждения понятия «качество подготовки», см. предыдущий параграф) понимается система устойчивых характеристик в деятельности педагога и обучаемых, методике преподавания, результатах обучения, выраженных в форме качественных и количественных показателей, которые поддерживаются в ходе учебного процесса на определенном уровне, не ниже принятых норм и требований образовательной программы.

Критерий «качества» определяет специфику достигнутых результатов и указывает ту относительную границу, при которой эти результаты соответствуют (или не соответствуют) нормативным показателям,

установленных теоретическим или опытным путем. «Качество» учебного процесса может быть выражено по показателю обученности, опосредованно отображая в нем выполнение тех требований, которые предъявляются образовательной программой к деятельности обучаемых и их результатам.

Качественная оценка – это единичная или совокупная определяющая характеристика педагогического объекта, в которой абстрагированы и обобщены его наиболее существенные свойства (качества) и представлены в форме вывода или заключения об этом объекте (иногда с указанием степени качества: больше-меньше, лучше-хуже).

Количественная оценка также имеет дело с качественной характеристикой объекта и отличается тем, что дополняет ее, уточняет и углубляет, что дает возможность ответить на вопрос: «сколько», «насколько больше», «насколько меньше». Количественную оценку нельзя противопоставлять качественной. В обоих понятиях речь идет об одних и тех же свойствах, лишь в последнем случае они выступают в единстве с числовыми значениями. Следовательно, количественная оценка по отношению к качественной всегда является вторичной, производной от первой. Важно понимать, что количественная оценка – это не число, получаемое в результате измерений и суждений, а приписанное оценочному суждению значение. Количественные манипуляции с оценками недопустимы.

Большинство оценок результатов образования, понимаемого в широком современном толковании этого понятия, такие, как оценки личностного развития, различных сторон воспитанности, готовности поступать тем или иным образом в обычной или экстремальной ситуации, креативности и т.д., осуществляется квалиметрическим (качественным, описательным) путем. Известны и широко применяются следующие квалиметрические (описательные) методики определения результатов и качества образования:

– достижение прогноза, сделанного учителями (как одними из экспертов) в зоне ближайшего развития обучаемого;

– оценка поведения в естественно возникших экстремальных ситуациях, которые всегда представляет жизнь;

– оценка поведения учеников в специально продуманных и организованных педагогами диагностических ситуациях;

– достижение высшей качественной оценки по шкале критериев, разработанных самой школой;

– использование совокупности так называемых сквозных (то есть годных для оценки любых действий, поступков и т.п.) психологических показателей, например:

а) знает, как надо (интеллектуальный показатель);

б) способен реализовать (волевой показатель);

в) хочет реализовать (эмоциональный показатель).

Принцип построения элементарных критериев оценки заключается в выражении соотношения двух или нескольких эмпирических величин:

- затраченного времени к планируемому;

- числа правильных ответов по всему множеству заданных вопросов;

- числа правильно решенных задач ко всему перечню предложенных;

- числа допущенных ошибок к возможному среднему их значению;

- совокупность выполненных действий ко всей необходимой численности в ходе достижения цели.

Указанные критерии принадлежат к виду относительных показателей, значения которых выражаются отвлеченным числом или в процентах. Они просты в использовании и удобны для проведения элементарного педагогического анализа (сравнения эффективности отдельных фрагментов однородной деятельности обучаемых), но мало пригодны для оценки сложной учебной работы.

В настоящее время основой психолого-педагогических измерений служит пооперационный или поэлементный анализ, позволяющие предмет исследования расчленять на «равные» законченные части. Однако при оценке предмета в целом не учитываются их весовые коэффициенты и не принимается во внимание характер и внутренние закономерности между структурообразующими компонентами.

В зарубежной педагогической литературе используется близкое к измерению понятие "evaluation", что можно перевести как "оценивание", имея в виду не только конечный результат, но и процесс формирования оценки. Соответственно выделяются два основных вида оценивания - формирующее и итоговое. Цель первого – оказывать, посредством оценивания, формирующее влияние на текущий процесс обучения, в смысле его улучшения, за счет установления обратной связи от студента к преподавателю. Цель второго – получить итоговые результаты обучения.

1.3. Мониторинг качества образования

Мониторинг качества образования является важнейшим элементом системы качества (системы управления качеством) в учебном заведении, обеспечивая механизм обратной связи в системе управления качеством образования.

Под мониторингом понимается отслеживание состояния объекта и фиксация проблем, осуществляемые на основе систематизации и обработки, прежде всего, существующих источников информации, а также специально организованных исследований и измерений; прогнозирование тенденций и выработка рекомендаций по корректировке развития объекта мониторинга

Объектом педагогического мониторинга являются результаты учебно-воспитательного процесса и средства, которые используются для их достижения. Он позволяет определить, каково состояние образования на данном этапе, насколько рациональны педагогические средства, в какой мере

они соответствуют заданным целям, насколько эффективны педагогические технологии.

Длительность слежения за состоянием педагогических явлений позволяет выявлять тенденции их изменения, устанавливать зависимость их от определенных факторов и условий. Таким образом, педагогический мониторинг представляет собой системную и лонгитюдную диагностику качества образования.

Можно выделить разные уровни мониторинга.

Макромониторинг организуется в масштабах общегосударственной системы, является основой разработки стратегии развития образования. Формы организации макромониторинга в России: развитие федеральной системы контроля качества образования, создание центра национального тестирования, функционирование системы лицензирования, аттестации и аккредитации образовательных учреждений.

Микромониторинг организуется на региональном и муниципальном уровнях, а также в пределах конкретного образовательного учреждения и является основой разработки путей совершенствования учебно-воспитательного процесса.

Мониторинг качества образования в учебном заведении организуется для отслеживания и анализа состояния и тенденций развития образовательной системы высшего учебного заведения.

Мониторинг качества образования в учебном заведении заключается:

- 1) в сборе информации о состоянии основных видов деятельности учебного заведения, связанных с подготовкой выпускников (в случае вуза – специалистов с высшим образованием),
- 2) анализе этой информации с точки зрения качества образования,
- 3) выявлении динамики и тенденций развития образовательной системы учебного заведения,

3) подготовке и реализации необходимых управленческих решений, направленных на повышение качества образования.

Мониторинг качества образования в учебном заведении является частью комплексного системного мониторинга качества образования в России, осуществляемого на федеральном, региональном, отраслевом (уровне УМО) уровнях, уровне образовательных областей и отдельного образовательного учреждения (вуза).

Объектами мониторинга в высшем учебном заведении являются:

- качество образования под которым понимается качество подготовленности выпускников учебных заведений с точки зрения государственных и общественных потребностей, а также потребностей личности в получении образования;

- система обеспечения качества образования, включающая в себя следующие основные структурно-функциональные аспекты деятельности учебного заведения:

- а) качество содержания образования;
- б) состояние научно-педагогических кадров и условия преподавательской деятельности;
- в) состояние организации учебного процесса и использования образовательных технологий;
- г) состояние материальной базы, библиотечного, учебно-методического и информационного обеспечения учебного процесса в вузе;
- д) состояние системы внутривузовского, внутришкольного контроля качества подготовки студентов и выпускников;
- е) состояние системы управления образовательным процессом;
- ж) состояние воспитательной работы в вузе.

Для активизации и координации работ по повышению качества образования в каждом вузе в настоящее время создается специальное

подразделение, находящиеся в ведении учебного управления – Центр мониторинга качества образования. Основными задачами Центра являются:

- совершенствование методики мониторинга качества образования на основе непрерывного контроля уровня подготовки студентов;
- внедрение сквозного непрерывного мониторинга качества образования на основе специального программного пакета;
- формирование необходимого набора банков тестовых заданий путем приобретения готовых сертифицированных банков, а также за счет разработки собственных банков с последующей их сертификацией;
- подготовка и переподготовка преподавателей и специалистов в области технологий разработки банков тестовых заданий;
- формирование условий, организация работ и методическая поддержка разработки электронных банков тестовых заданий преподавателями и сотрудниками вуза;
- совершенствование методик проведения экспертизы качества программно-дидактических тестовых материалов.

Следует отметить социальную обусловленность мониторинга качества образования, поскольку сами процессы общего и профессионального образования, его содержание и формы, структура и организация на всех уровнях (государственном, региональном, местном и т. п.) подвержены постоянному и глубокому влиянию социальных, экономических, политических, духовных процессов. Поэтому при определенной устойчивости основополагающих параметров педагогического мониторинга необходимо постоянно отслеживать и учитывать изменяющиеся условия жизни общества и школы. Иными словами, мониторинг должен быть достаточно динамичен, иначе он может быстро превратиться в некую заданную схему, и его задачи могут свестись лишь к простой регистрации наблюдаемых фактов и явлений.

Благодаря такому подходу результаты мониторинга могут дать богатый материал для анализа и оценки, например, социальной мотивации, устойчивости или изменения поведения, профессиональных устремлений учащейся молодежи, а также для определения факторов, влияющих на формирование личности будущего специалиста.

1.4. Педагогический контроль. Функции и принципы педагогического контроля.

Управление любым процессом, в том числе и качеством образования, предполагает осуществление контроля, то есть определенной системы проверки его эффективности его функционирования. Крайне необходим он и для успешного протекания педагогического процесса. Благодаря контролю удовлетворяется потребность в получении информации об эффективности функционирования всей системы учебных заведений, осуществляется обратная связь в обучении. Суть проверки результатов обучения состоит в выявлении *уровня освоения* знаний обучаемыми, который должен соответствовать образовательному стандарту по данной программе, предмету.

Основные *задачи* контроля:

- выявление уровня правильности, объема, глубины и действенности усвоенных учениками знаний,
- получение информации о характере познавательной деятельности, об уровне самостоятельности и активности учащихся в учебном процессе,
- определение эффективности методов, форм и способов их учения.

Функции контроля:

- *диагностирующая* (определение объективно существующего уровня владения знаниями, умениями и навыками на конкретном этапе обучения, выявление пробелов и трудностей в усвоении учебного материала);

- *обучающая* (систематизация знаний и умений, их обобщение, их применение в разных познавательных ситуациях);

- *корректирующая* (соотнесение реального уровня сформированности знаний, умений и навыков с прогнозируемым и их совершенствование путем коррекции учебного процесса);

- *стимулирующая* (создание положительных мотивов овладения изучаемым курсом, повышение интереса к ее изучению);

- *развивающая* (развитие умений быстрой концентрации усилий для решения в определенный срок поставленной задачи, сосредоточенности, мобилизации внутренних резервов психики, самостоятельной мыслительной деятельности, логики, умения анализировать, обобщать, систематизировать, классифицировать и т.д.);

- *воспитывающая* (воспитание воли, настойчивости выдержки, самообладания, критического отношения к учебной деятельности, культуры мышления, культуры общения с педагогом и со сверстниками, культуры организации учебной деятельности).

Важнейшие *принципы* контроля:

- *объективность* (заключается в научно-обоснованном содержании диагностических заданий, тестов, вопросов, диагностических процедур; - равном, дружеском отношении педагога ко всем обучаемым; точном, адекватном установленным критериям оценивании знаний, умений);

- *систематичность* (означает осуществление контроля на всех этапах дидактического процесса – от начального восприятия знаний до их практического применения; регулярному диагностированию должны подвергаться все обучаемые с первого до последнего дня обучения в учебном заведении; комплексный подход к проведению диагностирования: различные формы, методы и средства контролирования, проверки, оценивания используются в тесной взаимосвязи и единстве, подчиняются одной цели);

- *наглядность (гласность)* (подразумевает проведение открытых испытаний всех обучаемых по одним и тем же критериям; рейтинг каждого учащегося носит наглядный, сравнительный характер; оглашение и мотивация оценок. Оценка – это ориентир, по которому учащиеся судят об эталонах требований к ним и об объективности педагога; объявление результатов диагностических средств, их обсуждение и анализ, составление планов ликвидации пробелов. Педагог, вынося оценку, должен каждый раз обосновывать ее, руководствуясь логикой и существующими критериями).

Выполняя функцию руководства учебно-познавательной деятельностью учащихся, контроль не всегда сопровождается выставлением оценок. Он может выступать как способ подготовки учащихся к восприятию нового материала, выявления готовности обучаемых к усвоению знаний, навыков и умений, их обобщению и систематизации. В процессе контроля формируются и совершенствуются как общие, так и специальные практические навыки и умения учащихся, т.е. контроль имеет важное образовательное и развивающее значение.

Контроль выполняет также большую воспитательную роль в процессе обучения. Он способствует повышению ответственности за выполняемую работу не только учащихся, но и преподавателя, приучению обучаемых к систематическому труду и аккуратности в выполнении учебных заданий, формированию у них положительных нравственных качеств и коллективистских отношений, созданию здорового общественного мнения.

1.5. Традиционные виды, методы и формы контроля

По *месту* контроля на основных этапах обучения различают предварительный, текущий, тематический и итоговый контроль. По *характеру получения информации* в процессе контроля выделяют устный, письменный и лабораторный контроль. По средствам, используемым в ходе контроля, его подразделяют на машинный и безмашинный. По способу

организации контроля все ранее названные виды его можно разделить на программированные и непрограммированные.

На различных этапах обучения используются различные виды контроля. Предварительный контроль направлен на выявление знаний, умений и навыков обучаемых по предмету или разделу, который будет изучаться, то есть он проводится с диагностической целью перед изучением новой темы или в начале учебного года, четверти. Его цель – *ознакомиться с общим уровнем подготовки учащихся по предмету*. На основе полученных результатов преподаватель планирует, если необходимо, повторение (объяснение) материала, учитывает эти результаты в дальнейшей организации учебно-познавательной деятельности обучаемых.

Текущий контроль осуществляется в повседневной учебной работе и выражается в систематических наблюдениях преподавателя за учебно-познавательной деятельностью обучаемых на каждом занятии. Главное его назначение – *оперативное получение объективных данных об уровне знаний обучаемых и качестве учебно-воспитательной работы на занятии*. Полученная информация о том, как обучаемые усваивают учебный материал, как формируются их умения и навыки, помогает преподавателю наметить рациональные методы и приемы учебной работы, правильно дозировать материал, осуществлять постоянное руководство учебной деятельностью обучаемых, активизировать внимание и пробуждать интерес к изучаемому материалу. Можно сказать, что текущий контроль решает задачи руководства учебным процессом.

Тематический (периодический) контроль обеспечивает выявление и оценку знаний и умений обучаемых, усвоенных не на одном занятии, а в целом цикле занятий. Его цель – установить, насколько успешно обучаемые владеют **системой** определенных знаний, каков общий уровень их усвоения, отвечает ли он требованиям образовательного стандарта. Периодический контроль проводится, как правило, после изучения логически завершенной

части учебного материала- темы, подтемы, нескольких тем, раздела или полного курса.

Итоговый контроль обеспечивает усвоение знаний и умений обучаемых за более длительный период обучения: за четверть, полугодие, семестр, год или по окончании курса обучения. Цель его – установить систему и структуру знаний обучаемых. Итоговый контроль учитывает результаты всех других видов контроля. Особая форма итогового контроля – зачеты и экзамены.

Для получения информации о подготовленности учащихся, об их конкретных знаниях используются различные методы. **Методы** контроля – это способы диагностической деятельности, позволяющие осуществлять обратную связь в процессе обучения с целью получения данных об успешности обучения, эффективности учебного процесса.

Они должны обеспечивать систематическое, полное, точное и оперативное получение информации об учебном процессе.

Методы *устного контроля* – это беседа, рассказ учащегося, объяснение, чтение текста, технологической карты, схемы, сообщения об опыте и пр. Основу устного контроля составляют монологический ответ учащихся (в итоговом контроле это более полное, системное изложение) и/или вопросно-ответная форма – беседа, в которой преподаватель ставит вопросы и ожидает ответа учащихся. Устный контроль, как текущий, проводится ежеурочно в индивидуальной, фронтальной или комбинированной форме. Индивидуальный опрос учащихся позволяет учителю получить более полные и точные данные об уровне усвоения, однако он оставляет пассивными на уроке других учеников, что вынуждает учителя решать проблему их занятости во время опроса. Фронтальный опрос занимает всех учащихся сразу, но дает более поверхностное представление об усвоении.

Зачет и *устный экзамен* являются наиболее активной и обстоятельной проверкой знаний за определенный период обучения. Экзамены как способ

проверки знаний вызывают неоднозначную оценку как педагогов, так и учащихся. Отличаются два главных недостатка экзамена с традиционным выбором экзаменационных билетов учащимися. В “вытягивании” удачного/неудачного билета есть элемент случайности (“экзамен - лотерея”). Второе – экзамен является в известной степени стрессовой ситуацией для экзаменуемого, часто блокирующей его интеллектуальные возможности. Однако человек всю жизнь в процессе своей деятельности подвергается проверке и оценке, поэтому и учащиеся должны быть готовы к контролю и воспринимать ситуацию экзамена как стандартную, что, кстати, и свойственно большинству экзаменуемых. Поэтому экзамен остается эффективной формой проверки знаний, умений, развития мышления, сформированности взглядов, отношений, оценок.

У каждого экзаменатора даже при наличии общих критериев оценки знаний имеется свой подход к экзамену, свое понимание педагогических принципов требовательности и справедливости, свое отношение к оценке. Один особенно ценит точность изложения и основательность знаний, другой – оригинальность; одни экзаменаторы являются воплощением строгости, другие – самой доброжелательностью. Субъективизм и волюнтаризм в оценке устного опроса и экзамена ставит школьников в неравные условия, делает оценки несопоставимыми между собой. Все это говорит о том, что существующую систему методов контроля знаний у учащихся нужно изменить и, прежде всего, стандартизировать как по содержанию, так и по структуре и формам поведения.

Письменный контроль (контрольная работа, изложение, сочинение, диктант, реферат) обеспечивают глубокую и всестороннюю проверку усвоения, поскольку требует комплекса знаний и умений учащегося. В письменной работе нужно показать и теоретические знания и умения применять их для решения конкретных задач, проблем. Письменные работы

позволяют документально установить уровень усвоения материала, но требуют больших затрат времени.

Выполнение *практических* работ можно считать эффективным, но малоприменяемым способом проверки результатов обучения. Имеется в виду, прежде всего, проведение учеником лабораторных опытов, создание изделий, монтаж аппарата и пр.

1.6. Современные методы и формы педагогического контроля (тестовый, модульно-рейтинговый)

В условиях сокращения учебного времени в учебных заведениях одной из проблем становится повышение эффективности контроля знаний, для проведения которого необходимо сочетать традиционные и нетрадиционные его методы и формы.

К традиционным обычно относят, во-первых, субъектно-объектные отношения в системе “учитель-ученик”, во-вторых, экзамены, коллоквиумы, зачеты, контрольные работы и опросы. Им, несмотря на их значимость и важность, присущи определенные недостатки: пассивность учащихся, психологическая боязнь экзаменов, эпизодичность проверки и другие. Как правило, они носят характер интуитивного оценивания, часто недостаточно объективны. Именно педагогический субъективизм является главной причиной, по которой нынешние школьники отдают предпочтение компьютерным и тестовым формам контроля с минимальным участием педагогов.

Поэтому в системе контроля очень важным является использование нетрадиционных методов контроля, которые позволяют преодолеть данные ограничения. К одному из таких методов относится тестовый контроль.

Тестовый контроль – целенаправленное, одинаковое для всех испытуемых обследование, проводимое в строго контролируемых условиях, позволяющее объективно измерять изучаемые характеристики

педагогического процесса. Тестирование отличается от других методов обследования точностью, простотой, доступностью, возможностью автоматизации. Тест позволяет получить весьма информативный результат проверки, помогает определить, какие из исследуемых факторов влияют в наибольшей степени на успеваемость, какие – в меньшей или не влияют вовсе.

Контроль с использованием тестов или заданий тестового типа в последнее время всё больше внедряется в учебный процесс. Он находит всё новые и новые области применения, в частности при построении модульно-рейтинговой системы аттестации студентов.

Модульно-рейтинговая система складывается из двух взаимосвязанных и дополняющих одна другую частей (модульной и рейтинговой), которые могут функционировать и по отдельности, но с меньшей эффективностью.

Под модулем понимают логически завершённую часть учебного материала, обязательно сопровождаемую контролем знаний и умений студентов. Модуль совпадает с темой учебной дисциплины или блоком взаимосвязанных тем, но в отличие от темы, в рамках модуля все измеряется, все оценивается: посещение студентом занятий, выполнение им аудиторных и внеаудиторных заданий, исходный, этапный и итоговый уровень учебных достижений студентов.

В структуре модуля должны содержаться следующие элементы: цель, задачи, уровень овладения, содержание учебного материала, формы, виды, методы учебной деятельности студента, направленные на овладение учебным материалом, рекомендации по организации и методике самостоятельной работы студента по данному модулю, задания для самоконтроля и самооценки результатов обучения. Число модулей зависит от сложности и значимости учебного материала, от необходимости и возможной частоты контроля. Модуль содержит теоретическую и практическую части. Первая

формирует теоретические знания, вторая – профессионально важные умения и навыки.

Целью создания модуля является достижение заранее запланированного результата обучения. С помощью модулей достигается индивидуализация темпа обучения. Модульная система имеет целью поставить студента перед необходимостью регулярной учебной работы в течение всего семестра.

Исходя из понятия «модуль», под учебно-тематическим модулем мы понимаем комплекс учебных материалов по теме теоретического и практического характера, подлежащих усвоению на заданном уровне на основе целенаправленно организованной, лично-ориентированной учебно-познавательной деятельности студентов, а также обязательному измерению и оцениванию ее результатов.

Основной целью контроля является определение степени достижения задач обучения, воспитания и развития. Контроль по модулю может быть содержательным, деятельностным или содержательно-деятельностным. Результаты контроля являются основой для оценивания.

Одной из современных организационно-методических форм контроля является рейтинговый контроль. Понятие "рейтинг учебных достижений" является категорией педагогической квалиметрии. В условиях модульно-рейтинговой технологии обучения и контроля используются различные шкалы оценивания: количественная абсолютная, количественная относительная, порядковая. Эталоны оценки учебных достижений разрабатываются на основе образовательных стандартов, таксономии учебных целей.

Максимально возможная сумма баллов предметного рейтинга может устанавливаться по-разному: либо в соответствии с объемом часов, отводимых учебным планом на аудиторную и внеаудиторную работу по данной дисциплине, либо в соответствии с суммарной оценкой всех заданий по учебной дисциплине. При этом величина оценки за выполнение

конкретного вида учебно-познавательной деятельности выбирается самим преподавателем. Полученные студентом баллы при прохождении модульных видов контроля суммируются и составляют его рейтинг по данной дисциплине. В соответствии с рекомендациями дидактики высшей школы и требованиями педагогической квалитетрии минимальный рейтинг должен быть не менее 70 % от максимального. Перевод набранной суммы баллов в 4-балльную шкалу оценки осуществляется с учетом набранной суммы баллов в процентном отношении от максимально возможной суммы баллов.

Деятельность по разработке рейтинга по учебной дисциплине может быть представлена системой последовательных и взаимосвязанных действий: от составления рабочей учебной программы модульного типа до разработки структуры и содержания учебных занятий, предусматривающих самоконтроль и самооценку различных видов учебно-познавательной деятельности студентов.

Рейтинговый контроль создает условия для формирования навыков самоконтроля и самооценки результатов учения, превращает студента в субъекта контроля и оценивания процесса и результатов своей учебно-познавательной активности. Рейтинговый контроль позволяет повысить объективность и полноту контроля, более эффективно реализовать его организующую, обучающую, воспитывающую и развивающую функции, способствует переводу внешнезаданных целей обучения во внутреннезаданные.

Таким образом, особенности модульно-рейтингового обучения заключаются в следующем: в модульном построении содержания учебного курса и организации его изучения; акценте на самостоятельную работу студентов по модулю; индивидуализации темпа, времени овладения модулем; наличие четких критериев оценки различных видов учебно-познавательной деятельности и ее результатов; широком использовании тестового результатов обучения (при этом контролю подлежали не только

знания, умения и навыки, но и такие социально и личностно важные качества, как учебная мотивация, уровень притязаний, способность к самоконтролю и самооценке учебных достижений и т.д.); постоянном самоконтроле и самооценке результатов учебной деятельности студентами; осуществлении мониторинга и управления качеством учебных достижений студентов.

1.7. Папка индивидуальных учебных достижений «портфолио»

Одним из современных средств оценивания результатов обучения является портфолио.

Термин «портфолио» пришел в педагогику из политики и бизнеса: каждому знакомы понятия министерский портфель, портфель инвестиций. Portfolio – 1) предмет для хранения и переноски письменных работ, документов и т.д. (портфель, «дело», папка); 2) должность министра; 3) серия работ, документов, бумаг и т.д., объединенных общей темой: портфель ценных бумаг – security portfolio, portfolio investments – портфельные инвестиции, папка с рисунками.

Портфолио (в широком смысле слова) – это способ фиксирования, накопления и оценки индивидуальных достижений школьника в определенный период его обучения. Портфолио относится к разряду «аутентичных» (то есть истинных, наиболее приближенных к реальному оцениванию) индивидуализированных оценок, ориентированных не только на процесс оценивания, но и самооценивания.

Портфолио – это рабочая файловая папка, содержащая многообразную информацию, которая документирует приобретенный опыт и достижения учащихся. Портфолио позволяет учитывать результаты, достигнутые учащимся к разнообразных видах деятельности – учебной, творческой, социальной, коммуникативной и др. и является важнейшим элементом практико-ориентированного подхода к образованию. Можно отметить, что

портфолио дополняет традиционные контрольно-оценочные средства. Конечная цель учебного портфолио заключается в доказательстве прогресса обучения по результатам, по приложенным усилиям, по материализованным продуктам учебно-познавательной деятельности.

Основной смысл портфолио – «показать все, на что ты способен». Педагогическая философия портфолио предполагает:

- смещение акцента с того, что учащийся не знает и не умеет, на то, что он знает и умеет по данной теме, разделу, предмету;
- интеграцию количественной и качественной оценок;
- перенос педагогического удара с оценки на самооценку

Важная *цель* портфолио – представить отчет по процессу образования обучаемого, увидеть «картину» значимых образовательных результатов в целом, обеспечить отслеживание индивидуального прогресса ученика в широком образовательном контексте, продемонстрировать его способности практически применять приобретённые знания и умения.

Оценка тех или иных достижений (результатов), входящих в портфолио, а также всего портфолио в целом (либо за определенный период его формирования) может быть как качественной, так и количественной.

Портфолио не только является современной эффективной формой оценивания, но и помогает решать важные педагогические *задачи*:

- поддерживать высокую учебную мотивацию школьников;
- поощрять их активность и самостоятельность, расширять возможности обучения и самообучения;
- развивать навыки рефлексивной и оценочной (самооценочной) деятельности учащихся;
- формировать умение учиться – ставить цели, планировать и организовывать собственную учебную деятельность.

Состав портфолио зависит от конкретных целей обучения. Именно цели являются свидетельством усилий, достижений и прогресса в обучении

учащегося по конкретному предмету, разделу, теме. Каждый элемент портфолио желательно датировать, чтобы можно было проследить динамику прогресса учащегося. Так же желательно, чтобы при оформлении окончательного варианта портфолио включал три элемента:

- описание цели сбора портфолио, его предназначения и краткого описания;
- содержание (или оглавление) портфолио с перечислением его основных элементов;
- перспективы дальнейшего использования, т.е. взгляд в будущее.

«Портфолио документов» – портфель сертифицированных (документированных) индивидуальных образовательных достижений. Этот раздел предполагает возможность как качественной, так и количественной оценки материалов портфолио. Итоговый документ представлен в виде вкладыша в аттестат и содержит итоговый балл, а также перечень и оценки сертификатов, входящих в его состав. Документы или их копии могут быть помещены в приложении к портфолио. В «Портфолио документов» ученик представляет сертификаты официально признанных на международном, федеральном, региональном, муниципальном уровне конкурсов, соревнований, олимпиад, документы об участии и грантах, об окончании музыкальных или художественных школ, сертификаты о прохождении тестирования, например TOEFL и т.п.

«Портфолио работ» представляет собой собрание различных творческих и проектных работ ученика, а также описание основных форм и направлений его учебной и творческой активности: участие в научных конференциях, конкурсах, учебных лагерях, прохождение элективных курсов, различного рода практик, спортивных и художественных достижений и др. Данный раздел портфолио предполагает качественную оценку: например, по параметрам полноты, разнообразия и убедительности материалов, качества представленных работ, ориентированности на

выбранный профиль обучения и др. Портфолио может оформляться в виде творческой книжки ученика с приложением его работ, представленных в виде текстов, электронных версий, фотографий, видеозаписей.

В «Портфолио работ», учащийся представляет, во-первых, «ведомость» или «зачетную книжку» о прохождении им курсов по выбору. Во-вторых, может представить сами работы, проекты, исследования, которые он выполнил в ходе обучения на курсах по выбору или при обучении в учреждениях дополнительного образования, других учреждениях или выполнил самостоятельно, не получив за это подтверждающего документа. В-третьих, могут быть представлены сертификаты обучения, конкурсов, соревнований и т.п., не имеющих «официального признания» на уровне федерации, региона, муниципалитета. В-четвертых, ученики могут представлять собственные изделия: например модели, картины, стихи, музыкальные произведения собственного сочинения, фотографии, компьютерные программы и т.п.

«Портфолио отзывов» включает в себя характеристики отношения школьника к различным видам деятельности, представленные учителями, родителями, возможно одноклассниками, работниками системы дополнительного образования и др., а также письменный анализ самого школьника своей конкретной деятельности и ее результата. Портфолио может быть представлен в виде текстов заключений, рецензий, отзывов, резюме, эссе, рекомендательных писем и пр.

В «Портфолио отзывов» ученикам предлагается представлять отзывы на творческие работы, исследовательские и другие проекты, социальные практики, участие в конференциях и самых различных сферах приложения сил. Важной составляющей данного раздела является самооценка ученика, его рефлексия на собственную деятельность. Бывает, что учащийся представляет неплохие результаты по каким-нибудь предметам или в конкурсах, олимпиадах, а выполняет при этом работу с большим

нежеланием, о чем часто никто не подозревает. В результате он может получить рекомендации заниматься на профиле, где он будет успешен, но не будет испытывать морального удовлетворения. Помочь школьнику «найти себя», грамотно выстроить профориентацию будет возможно используя этот раздел портфолио, где будет представлена саморефлексия ученика на разнообразную выполняемую им деятельность, начиная с учебной и урочной и заканчивая хобби.

Метод портфолио обеспечивает дифференцированный подход к принятию решений, который повлияет на отношение детей к работе и школе в целом. Это освобождает преподавателя от ограничений, связанных со стандартизацией исследований. Наконец, используя портфолио, преподаватель имеет возможность расширить работу в классе и с каждым ребенком. Кроме того, использование портфолио позволяет сосредоточить свое внимание на ребенке и развивать межличностные отношения с ним.

Это особенно важно преподавателям и учащимся для совместной работы, для того, чтобы выделить критерии, которые будут использованы в качестве основы для оценивания развития учащихся, как полностью (т.е. в процессе всего учебного периода), так и частично (то есть, как часть законченного проекта, деятельности, или оценки для выявления степени, которую необходимо достигнуть по определенным стандартам).

1.8. Оценка: понятие, функции, критерии

Понятие «оценка» в философской литературе используется в широком и узком смысле. В широком смысле оценка понимается как сравнение с эталоном изучаемого объекта, категоризация объекта. Такую оценку называют объективной, так как она опирается на достижения науки, подтверждается практикой. В узком смысле оценка отражает значимость объекта для субъекта, соотношение объекта с целями, потребностями и

интересами субъекта. Такую оценку называют субъективной, так как она дает объективно-субъективную информацию о ценностях, а не о сущностях.

Следовательно, педагогическая оценка используется в так называемом «узком смысле» и носит, в сущности, субъективный характер.

Функции оценки: стимулирующая; фиксирующая; ориентирующая; информирующая; воспитательная, контролирующая.

Критерии выставления оценки:

- *нормативный* критерий – результат контроля оценивается с точки зрения установленной нормы, образца проверяемого объекта;

- *сопоставительный* критерий – результат контроля одного обучаемого сравнивается с аналогичным результатом другого;

- *личностный* – результат контроля обучаемого сравнивается с прошлыми результатами этого же ученика и тем самым устанавливается динамика его продвижения в овладении оцениваемыми знаниями, умениями и навыками.

Формы оценок:

- *нормативная* оценка – используется, чтобы оценить насколько знания и умения соответствуют тому уровню, который принят за норматив;

- *критериальная* оценка – используется для определения содержания знаний студента или для выявления его конкретных предметных учебных навыков;

- *процессуальная* оценка – осуществляется для того, чтобы понять, как протекает учебная деятельность студента.

Негативные стороны современной оценочной системы.

1. Различные виды познавательной деятельности, выполняемые студентом и являющиеся неравнозначными, оцениваются по единой оценочной шкале.

2. При выставлении оценки не проводится различие между собственно успеваемостью и личностью студента; не учитывается тот путь, те усилия, которые стоят за каждым из оценочных фактов.

3. Зачастую носят травмирующий характер. Процесс учебы заменяется спорами, как и за что ставятся отметки, о справедливости различных моментов этой процедуры.

В связи с этим при оценивании учащихся рекомендуется придерживаться следующих **принципов**:

- учащиеся должны постоянно иметь в виду, по каким критериям будет оцениваться их деятельность;
- оценка не должна обладать карающими функциями, как это часто встречается в современном российском образовании;
- необходим баланс проверки знаний и умений;
- оценивание должно производиться как за групповую работу, так и за индивидуальную;
- оценка должна осуществляться как на основе индивидуального прогресса, так и на основе абсолютных достижений.

Проблема оценивания чрезвычайно актуальна. Деятельность оценивания имеет наибольшую по времени продолжительность в профессиональной деятельности педагогов и ценностное отношение к миру у обучаемых формируется именно как отражение преподавательского оценивания. К сожалению, господствующим пока является представление, что вузовский преподаватель должен хорошо знать в первую очередь свой предмет, а воспитание переадресовывается к родителям, молодежным организациям, соответствующим структурам вуза и т.п.

Попытки изменения существующей систем оценивания сводятся к:

- изменению содержания, качественного наполнения оценки (например, вербализация отметки; учет суммарного «веса» рассматриваемых параметров; введение рейтинговой системы);

- изменению количественного выражения оценки (вместо пятибалльной предлагается 12, 50, 100-балльная шкала)

- включению аутентичной оценки в систему оценивания результатов обучения школьников.

Понятие аутентичной оценки введено группой американских педагогов (Линн, 1995; Шепард, 1989; Валенсия, 1990), предложивших взгляд на оценку как на естественный многоаспектный процесс. Они высказали необходимость пересмотра самого процесса оценивания и смещения акцента в сторону оценки *реальной деятельности школьников* на занятии, а не ее результата, predetermined заранее взрослыми. При этом может быть оценена любая, не только мыслительная деятельность ученика. Это позволяет ребенку реализовать имеющиеся возможности. Аутентичная оценка, по сути своей, гуманна, поскольку она позволяет каждому ученику быть успешным.

Таким образом, в современных условиях предъявляются высокие требования к методологии и методике оценивания личностных и учебных достижений обучаемых. Для обеспечения объективности и надежности оценки учебных достижений должна вестись планомерная работа по широкому внедрению тестовых методик оценки учебных достижений в педагогическую практику.

Вопросы для самоконтроля

1. Дайте определение понятию "качество образования"
2. Какие функции контроля вы считаете приоритетными для повышения качества обучения?
3. Какова роль оценки и отметки в обучении?
4. Охарактеризуйте достоинства и недостатки традиционных средств оценивания результатов обучения

Практические задания и упражнения по теме

1. Ниже представлены ситуации, выявленные с помощью контроля. Какой вид контроля вы бы использовали для их определения? В каких случаях лучше использовать тесты, а в каких - традиционные средства контроля?

- Из всего класса только 10 учащихся достигли по своим результатам необходимой скорости чтения, а 15 пока не научились хорошо читать
- Иванов на уроке математики получил отметку "5", а Петров "2".
- Сидоров с трудом различает звонкие и глухие согласные
- Мой метод преподавания не эффективен для этого класса
- Петя нуждается в моей помощи для развития навыков устного счета
- Сидорова необходимо перевести в группу успевающих учеников
- У моего коллеги, преподавателя русского языка, класс оказался сильнее моего
- Миша медленно усваивает новый материал

2. Какой вид контроля вы используете для принятия каждого из 8 решений?

3. Когда лучше использовать тесты, а в каких случаях можно полагаться на традиционные средства контроля?

ТЕМА 2. ТЕОРИЯ ТЕСТИРОВАНИЯ: ОБЩИЕ ПОЛОЖЕНИЯ

Перечень вопросов, рассматриваемых в теме:

- 1. Наука о тестах: pro & contra*
- 2. Исторические предпосылки современного тестирования в отечественном образовании*

2.1. Наука о тестах: pro & contra

Нынешняя практика тестирования дает немало фактов для разных выводов - как против тестов, так и в их пользу. Голоса "против" звучат иногда громче голосов "за". Важно понять простую истину: сами по себе - тесты полезный метод, но их применение в практике может оказаться как полезным, так и вредным. Многое зависит от цели, качества, соблюдения принципа добровольности, возможности потенциального или реального ущерба личности. Государственное тестирование не только вредно, но и недопустимо ни при каких обстоятельствах. И уже совсем недопустимо применение тестовых суррогатов, вроде КИМов ЕГЭ. Это не только мнение автора этой статьи, но и общее этическое требование культурной тестовой деятельности.

Различия в отношении к тестам привели к тому, что у многих руководителей образовательных учреждений нет определенности в вопросах разработки и применения тестов. Заметно отстающее от нужд практики научное обеспечение процесса тестового контроля стало одной из причин ситуации, в которой тесты нередко становятся предметом хотя и заинтересованного, но малопродуктивного спора, а нередко, и профанации. Данные, полученные на множестве студентов, - тестовые баллы - представляют образовательные процессы не в идеальном, а в реальном свете. Вероятно, поэтому спорность стала извечной спутницей тестов. В отечественной литературе тестам слишком долго отказывали в научности,

давая тем самым много поводов для разрушительной критики; о критике же конструктивной говорить трудно из-за практического отсутствия как фактов такой критики, так и самих журналов, в которых такая критика могла бы быть напечатанной.

Наука о тестах. Для понимания сущности тестов важно разобраться в системе понятий. Понятия вообще образуют основу любой науки, и в этом смысле деятельность по разработке и эффективному применению тестов не является исключением. Начиная с 30-х годов, наука о тестах называлась буржуазной, все цели которой считались "реакционными". И хотя такие суждения пролеткультовского толка теперь уже считаются неадекватными духу нашего времени, все-таки появляются публикации, где тестам по-прежнему пытаются отказать в научности.

Первые научные труды по теории тестов появились в начале XX века, на стыке психологии, социологии, педагогики и других так называемых поведенческих наук (Behavioral Sciences). Зарубежные психологи называют эту науку психометрикой (Psychometrika), а педагоги - педагогическим измерением (Educational Measurements). Поскольку общего названия на русском языке пока нет, автор называл эту науку тестологией, которая может быть педагогической, психологической или социологической, в зависимости от того, где применяется и развивается. Незамутненная идеологией и политикой, интерпретация названия "тестология" проста и прозрачна: наука о тестах. В XXI веке название этой науки я привёл в соответствие с её названием на Западе - Педагогические Измерения.

Педагогические Измерения призваны заниматься вопросами разработки тестов для объективного контроля подготовленности учащихся. В структуре подготовленности большое (но не исчерпывающее) место занимают знания, умения, навыки и представления. Сюда же следует добавить интеллектуальное и физическое и культурное развитие, творческие способности, воспитанность и уровень развития эмоционально-чувственной

сферы. Педагогические Измерения - это прикладная методическая теория научной педагогики. Ключевыми понятиями Педагогических Измерений являются измерение, тест, содержание и форма заданий, надежность и валидность результатов измерения. Кроме того, в Педагогических Измерениях используются такие понятия статистической науки, как выборочная и генеральная совокупность, средние показатели, вариация, корреляция, регрессия и др.

Исходное понятие теории тестов. Педагогическое (учебное) задание можно определить как средство интеллектуального развития, образования и обучения, способствующее активизации учения, повышению подготовленности учащихся, а также повышению эффективности педагогического труда. В правильно организованном процессе образования большая роль отводится педагогическим заданиям. Понятие "задание" является общим, охватывающим цель и смысл не только теста, но и всех учебных заданий. Оно включает такие педагогические средства, как вопрос, задача, учебная проблема и другие, используемые, главным образом, в собственной учебной деятельности (учении).

Задания могут формулироваться в тестовой, и, скажем так, в нетестовой форме. В российском образовании большинство учебных заданий дается учащимся в нетестовой форме. В основном, это вопросы, задачи, упражнения. В зарубежном образовании доля заданий в тестовой форме существенно выше, что объясняется соображениями проводимой там образовательной политики, имеющихся там педагогических теорий, методик, обучающей техники и технологии.

Педагогические задания выполняют как обучающие, так и контролирующие функции. Обучающие задания применяют учащиеся для активизации собственного учения, усвоения учебного материала, саморазвития, а также применяют педагоги для обучения учащихся. Все это свидетельствует об обучающем потенциале заданий. Контролирующие

задания применяются, напротив, педагогом или проверяющими органами после окончания учебного года, или другого определенного цикла (четверти), с целью диагностики уровня и структуры подготовленности. Некоторая часть заданий может использоваться для обучения и для контроля.

Что такое тест? Слово "тест" вызывает самые различные представления. Одни полагают, что это вопросы или задачи с одним готовым ответом, который надо угадать. Другие считают тест формой игры или забавы. Третьи пытаются истолковать это как перевод с английского слова "test", (проба, испытание, проверка). В общем, по этому вопросу нет единства мнений. Тем более что в учебниках педагогики об этом не пишут. А если где и пишут, то нередко написанное трудно понять. Не случайно размах мнений о тестах оказывается слишком широким: от суждений обыденного сознания до попыток научного истолкования сущности тестов.

В науке проводят существенные различия между простым переводом слова и смыслом понятия.

Чаще всего мы встречаемся с упрощенным восприятием понятия "тест" как простой выбор одного ответа из нескольких предложенных к заданию. Многочисленные примеры таких, казалось бы, "тестов" легко найти в газетно-журнальной периодике, в различных конкурсах и в многочисленных книжных публикациях под названием "Тесты". Но и это часто оказываются не тесты, а нечто внешне похожее на них. Обычно это сборники вопросов и задач, рассчитанных на выбор одного правильного ответа из числа предложенных. Они только по внешней видимости похожи на настоящий тест. Различия в понимании сущности тестов порождают различия в отношении к тестам.

В наши дни существует много видов тестов, поэтому дать универсальное определение для всех этих видов вряд ли можно.

Традиционный тест представляет собой стандартизованный метод выявления уровня и структуры подготовленности. В таком тесте все

испытуемые отвечают на одни и те же задания, в одинаковое время, в одинаковых условиях и с одинаковыми правилами оценивания ответов. Главная цель применения традиционных тестов - установить отношение порядка между испытуемыми по уровню проявляемых ими знаний. И на этой основе определить место (или рейтинг) каждого на заданном множестве тестируемых испытуемых. Для достижения этой цели можно создать бесчисленное количество тестов, и все они могут соответствовать достижению поставленной задаче.

И тогда возникает один из главных вопросов теории тестов - вопрос выбора наилучшего теста из практически неограниченного множества всех возможных тестов. Каждый тест может отличаться от других по числу заданий и другим характеристикам. С прагматической точки зрения выгодней делать тест, имеющий сравнительно меньшее число заданий, но обладающий большинством достоинств, присущих тестам с большим числом заданий. От числа заданий некоторым образом зависит точность педагогического измерения. Этот вопрос будет рассмотрен отдельно при изложении теории надежности тестовых результатов.

В тест стараются отобрать минимально достаточное количество заданий, которое позволяет сравнительно точно определить уровень и структуру подготовленности. Интерпретация результатов тестирования ведется преимущественно с опорой на среднюю арифметическую и на так называемые процентные нормы, показывающие, - сколько процентов испытуемых имеют тестовый результат худший, чем у любого другого испытуемого. Такая интерпретация тестовых результатов называется нормативно-ориентированной. Тест определяется как система параллельных заданий равномерно возрастающей трудности, позволяющая эффективно измерить уровень и качественно оценить структуру подготовленности учащихся. Это определение гомогенного педагогического теста.

2.2. Исторические предпосылки современного тестирования в отечественном образовании

Всякая попытка определить точное время возникновения тестов напоминает стремление географов найти точное начало большой реки, вытекающей множеством ручейков из обширного болота. Примерно так же обстоит дело и с тестами.

В Древнем Египте искусству жрецов обучали только тех, кто выдерживал систему определенных испытаний. Вначале кандидат проходил процедуру, которую можно было бы сейчас назвать собеседованием. При этом выясняли биографические данные, уровень образованности, оценивали внешность, умение вести беседу. Затем проверяли умения – трудиться, слушать, молчать. Проводили испытания огнем, водой и угрозой смерти. Тем, кто не уверен, что выдержит все тяготы длительного образования, предлагалось подумать – с какой стороны окончательно закрыть за собой дверь в храм – с внутренней или внешней?

Сообщается, что эту суровую систему испытаний и отбора успешно преодолел в молодые годы Пифагор. Вернувшись после учебы в Грецию, он основал свою школу, допуск в которую открывал только после серии различных испытаний, похожих на те, которые выдержал он сам.

Пифагор подчеркивал важную роль интеллектуальных способностей, утверждая, что "не из каждого дерева можно выточить Меркурия". И потому придавал значение диагностике, в первую очередь, именно этих способностей, что делалось с помощью трудных математических задач.

Как считают, Пифагор обращал также внимание на походку и на смех молодых людей, утверждая, что манера смеяться служит самым хорошим показателем характера человека. Он внимательно относился к рекомендациям родителей и учителей, тщательно вел наблюдение за каждым новичком, особенно после того, как последнего приглашали свободно высказываться и, не стесняясь, смелее оспаривать мнение собеседников.

Одной из самых распространенных форм диагностики способностей человека в те времена была физиогномика – искусство распознавания характера и способностей человека по его внешнему виду. Гиппократ, впервые употребивший это название, считал физиогномику наукой. Тогда же появились первые учебники и практические руководства по физиогномике, а также и первые специалисты.

Различные конкурсы и экзамены устраивались и в средневековом Вьетнаме. Всего за два года, в период с 1370 по 1372 г. удалось провести переаттестацию всех гражданских чиновников и военнотружущих, что позволило улучшить работу государственного аппарата. В результате этого Вьетнам стал сильным и жизнеспособным государством. Особое внимание было уделено созданию боеспособного офицерского корпуса. В XV в. экзамены были там упорядочены. Они проводились по этапам и турам. Лауреаты получали подарки от короля, их имена вносились в "золотой список", который вывешивался у Восточных ворот столицы, об их победах на конкурсе сообщалось в родную общину. Имена наиболее отличившихся участников высекались на специальных каменных стелах, установленных в Храме Литературы.

Если использование письменных контрольных работ и экзаменов в странах Востока уже в те годы считалось вполне нормальным и естественным делом, то этого нельзя сказать о странах Запада. Там возникали трудности психологического характера, вызванные попытками замены привычных устных форм контроля письменными. Например, в Англии сто лет ушло на слом консервативных традиций, препятствовавших применению письменных контрольных работ в учебном процессе, и еще сто лет на то, чтобы последние начали использоваться на приемных и выпускных экзаменах.

В той же Англии ценность письменных работ быстрее других оценил Орден Иезуитов, увидевший в них средство повышения мотивации учебной

работы. Используя свое влияние, Орден распространил практику применения письменных работ во многих других странах. Ответная реакция на это в первую очередь проявилась в Америке. Сообщается, что там, в 1762 г., состоялось открытое выступление студентов Йельского университета против использования письменной формы экзаменов, из-за которых им приходится много зубрить.

В 1884г. в США вышла первая книга с тестовыми материалами, содержащая задания и ответы к ним с оценкой по пятибалльной шкале. В этой книге содержались задания по математике, истории, грамматике, навигации, давались примерные тексты сочинений вместе с методом количественной оценки сочинений. Это был первый в истории случай использования простейших статистических расчетов в педагогической работе.

Быстро распространившись в США, такая форма контроля знаний имела не только много сторонников, но и противников. Сообщается, например, о школьном инспекторе из Чикаго, запретившем в 1881г. проводить письменный контроль и обязавшем осуществлять перевод учащихся в очередной класс не на основе проверок, а опираясь исключительно на мнения учителей и директоров школ. Распространилось мнение, что никто лучше учителя не сможет оценить способности ученика, и потому экзамены и всякие другие формы контроля являются издевательством над здравым смыслом. Этим было положено начало общественному движению против какого-либо контроля в школе. Пик такого движения в США пришелся на конец XIX века (1880-е годы), после чего вновь стали появляться предпосылки для формирования благожелательного отношения к контролю знаний.

Потребность в создании общественно-государственной системы контроля знаний была осознана в США в 1885 году. В Нью-Йорке начал действовать экзаменационный совет – один из немногих советов, которые

смогли разработать методы объективного контроля знаний. Начало общегосударственной системы оценки знаний положило создание там в 1900 году комитета по проверке знаний абитуриентов колледжей в США.

В России конца XIX – начала XX века вопрос разработки тестов и их применения не ставился и не обсуждался. “Проверка знаний, – писал один из авторов, – представляется мне по опыту самым скучным, бесполезным и даже вредным занятием. Она терпима, как необходимое зло, только там, где от учителя требуют ежемесячных отметок как средства контроля не только над учениками, но и над учителем”. Однако с той поры в российской педагогике стихийно сформировались две противоречащие тенденции: одна за контроль знаний учащихся, другая – против.

Приведенный здесь краткий исторический экскурс позволяет сделать вывод о необходимости рассматривать испытания как важную часть жизни многих – если не всех – народов мира, со времен древнейших цивилизаций и до наших дней. Однако можно ли на основании приведенных данных говорить о раннем историческом периоде возникновения тестов? Если согласиться с наиболее распространёнными сейчас определениями теста, даваемыми как перевод с английского слова "test" (испытание, проверка, проба), то на поставленный вопрос надо будет ответить утвердительно. А если не согласиться? Тогда надо допустить, что настоящие тесты появились позже, в конце XIX-начале XX века.

Тесты в России в конце XIX первой половины XX в.. История тестов началась в России, как и в других странах, в конце XIX — начала XX в. В этот период тесты широко применялись сторонниками экспериментальной педагогики, которые всячески пропагандировали их наравне с другими психометрическими методиками в педологии. Видные специалисты того времени (А.П. Болтунов, А.Ф.Лазурский, А.П. Нечаев, Ф.Е.Рыбаков и др.) в своих исследованиях личности обучаемого широко использовали тесты .

На волне стремления к прогрессу в России в первые годы советской власти был востребован потенциал, накопленный российской наукой в образовании. У истоков советской педологии, в рамках которой развивались тестовые методики, стояли известные ученые того времени (М.Я.Басо, Л.С. Выготский, А.П. Болтунов, П.П. Блонский, К.Н. Корнилов, А.П. Нечаев, А.Р. Лурия и др.), посвятившие ряд своих работ оценке результатов обучения и развития детей. Однако период активной деятельности педологов и тестологов в советской России оказался коротким. В 1936 г. вышло известное постановление ВКП(б) «О педологических извращениях в системе Наркомпросов», в котором тесты были объявлены вредным методом. Многие годы, вплоть до начала 90-х гг. XX в., отношение к тестам было негативным. Педагоги-практики вернулись к тестам в период прекращения массовых политических репрессий. Уже к началу 60-х гг. XX в. многие учителя, заботящиеся не о формальных показателях, а о реальном качестве учебного процесса, стали использовать при контроле наборы заданий в тестовой форме, которые чаще всего назывались по-разному, но только не словом «тест».

Промежуток времени с 60-х гг. до начала 90-х гг. XX в. был для отечественной школы периодом постепенной либерализации, когда учителя изыскивали различные резервы в борьбе за повышение качества обучения на фоне снижения авторитарного характера учебного процесса. Тесты в это время существовали полулегально, официально запрет на них не был отменен, но уже предпринимались попытки диссертационных исследований по проблемам тестирования, появлялись работы педагогов-новаторов, в которых тесты и анкетные опросы использовались как инструмент для подтверждения эффективности нововведений. В это время тесты разрабатывались без должного знания теории педагогических измерений, на основе опыта и здравого смысла, поэтому чаще всего тестами их было можно назвать лишь условно.

Столь же необоснованными по надежности и валидности были наборы заданий с выбором ответов, создаваемые практически повсеместно на рубеже 60 - 70-х гг. XX в. в связи с интенсивным развитием программированного обучения. Только к середине 80-х гг. XX в. в нашей стране в научных и методических работах преподавателей стали утверждаться основные положения теории педагогических измерений.

В целом рассматриваемый временной промежуток в развитии тестов был периодом прогресса, когда на смену представлениям о тесте как о простом наборе заданий пришло научное понимание этого термина. В этот период появляются научные издания по тестовой и смежной с ней проблематике, учебные пособия и многочисленные статьи отечественных авторов, как поддерживающие, так и осуждающие тесты в образовании.

Развитие тестов в конце XX в. и в наши дни. Новая история тестов в России началась в 90-е гг. XX в., когда стали больше говорить не о руководстве, а о научно обоснованном управлении учебным процессом, в информационном обеспечении которого важная роль по праву принадлежит тестам. Новое понимание возможностей тестов в образовании способствовало росту научных исследований. В 90-е гг. XX в. появляются работы по проблемам измерений и тестирования в образовании, защищаются многочисленные кандидатские и докторские диссертации, издаются монографии, учебные пособия, журналы, проводятся конференции и симпозиумы.

Последнее десятилетие XX в. в нашей стране совпало с периодом бурного развития структур, занимающихся практической работой по созданию и применению тестов. К основным событиям этого периода можно отнести открытие в 1990 г. первой в стране кафедры педагогических измерений в Исследовательском центре проблем качества подготовки специалистов Московского института стали и сплавов; создание при Московском государственном университете в этом же году Центра

тестирования «Гуманитарные технологии» (проект по компьютерному аттестационному тестированию для старшеклассников «Телетестинг»); привлечение в 1991 г. Центра качества образования Института общего среднего образования РАО к участию в сравнительных международных исследованиях по оценке учебных достижений; открытие в 1995 г. Центра тестирования выпускников общеобразовательных учреждений, преобразованного впоследствии в Федеральный центр тестирования со статусом государственного учреждения Министерства образования и науки Российской Федерации; основание в 1998 г. государственной системы тестирования иностранных граждан по русскому языку как иностранному при поддержке ведущих вузов Москвы, Санкт-Петербурга и других городов России. Перечень этот можно было бы смело продолжить. Начиная с 90-х гг. XX в. практически во всех регионах России создаются центры тестирования и аттестации учащихся, методические лаборатории по диагностике и центры качества образования.

Значимым событием в области подготовки кадров по педагогическим измерениям стало открытие в 2001г. на факультете повышения квалификации Российского университета дружбы народов кафедры тестологии. На этой кафедре впервые в нашей стране началась реализация профессиональной образовательной программы, рассчитанной на 1480 часов, для получения дополнительной квалификации «Тестолог (специалист в области педагогических измерений)».

Для повышения качества тестовых материалов, разрабатываемых в России, в 2000 г. при Исследовательском центре проблем качества подготовки специалистов Московского института стали и сплавов (технологического университета) был открыт Центр сертификации педагогических тестовых материалов (ПТМ) и создан Координационный совет Минобрнауки России (в настоящий момент – Минобрнауки) по вопросам сертификации качества педагогических тестовых материалов.

Коренные изменения в отношении учителей к тестам произошли в 2001г. в связи с началом эксперимента по введению единого государственного экзамена, благодаря которому тесты получили официальное признание в России. За годы эксперимента значительно повысилась степень доверия к результатам тестирования со стороны органов управления образованием, образовательных учреждений, самих учителей, которые убедились в высокой объективности и обоснованности баллов ЕГЭ, в их высокой прогностичности при отборе абитуриентов вузов. В 2002г. был открыт Федеральный институт педагогических измерений (директор А.Г.Ершов), приоритетным направлением деятельности которого является научное, методическое и организационное сопровождение процесса создания контрольных измерительных материалов (КИМ) для ЕГЭ.

В целом, подводя итоги развития тестирования в России, можно сказать, что на сегодняшний день среди ученых-педагогов в нашей стране, наконец, появилось понимание того, что теория педагогических измерений - наука, обладающая своей методологией, методами и аппаратом, необходимым для разработки качественных педагогических тестов. Сегодня большими тиражами издаются инновационные работы по педагогическим тестам отечественных ученых и зарубежных авторов, а также сборники материалов КИМ ЕГЭ, специальные журналы по тестовой проблематике. Таким образом, можно считать, что на данный момент в России сформировалось сообщество профессионалов — специалистов по разработке и применению тестов.

Вопросы для самоконтроля

1. Как вы считаете, будет ли возрастать роль тестов в российских школах?
2. Этапы введения тестов в практику образования.
4. Каково ваше отношение к тестам в образовании? Под влиянием каких факторов оно сложилось?

5. Какие периоды Можно выделить в отечественной и зарубежной истории развития тестов? Каковы их отличительные черты?

Практические задания и упражнения по теме

Подготовить доклады с презентацией на следующие темы:

- Развитие педагогического тестирования в России и за рубежом
- Тестирование в психологии, образовании, армии
- Взаимосвязь психологических и педагогических измерений
- Создание классической теории тестов
- Создание современной теории тестов

ТЕМА 3. ПЕДАГОГИЧЕСКИЙ ТЕСТ КАК СИСТЕМА ЗАДАНИЙ

Перечень вопросов, рассматриваемых в теме

1. Основные положения классической теории тестов
2. Обзор основных понятий общей теории педагогических измерений
3. Этапы формирования теста
4. Классификация теорий педагогических измерений по В.С.Аванесову.

Сущность и методологические принципы интерактивного обучения

3.1. Основные положения классической теории тестов

Начало этой теории положил Ch. Spearman в 1904 году. Существенный вклад в её развитие внес Н. Gulliksen. Эти и другие авторы предполагали наличие истинного компонента измерения, в каждом результате. Они же заложили основы аксиоматики психолого-педагогических измерений

В ней существует семь основных *аксиом*:

1. Устойчивое существование *латентного качества*.
2. Возможность представления *латентного качества* как непрерывно измеряемой величины.
3. Существование *истинного* значения *латентного качества* у каждого испытуемого. Измеряемая величина предполагается варьирующейся на множестве испытуемых.
4. Неизбежная погрешность измерения, в соответствии с которым измеряемое значение X не равно истинному T . Для каждого i -го испытуемого $X_i = T_i + E_i$, где E_i - некоторая случайная ошибка измерения, состоящая из суммы возможных ошибок различного происхождения, значения и знака.
5. Случайные погрешности измерений распределены по нормальному закону.
6. Истинные и ошибочные компоненты измерений не коррелируют между собой.

Из четвертой и шестой аксиом вытекает седьмая аксиома.

7. Дисперсия тестовых баллов состоит из двух компонентов.

Latent Trait Theory (Теория измерения латентных качеств). В начале 40-х годов D.N.Lawly задался целью найти некое свойство, присущее заданию, которое было бы инвариантным (неизменяемым) относительно уровня подготовленности группы испытуемых. Таким свойством, как выяснилось, обладают так называемые *характеристические кривые тестовых заданий* и *теста* в целом. Автор стал рассматривать параметр крутизны кривой в качестве *меры способности дифференцировать студентов по уровню их подготовленности*. Поскольку все без исключения, теории педагогических измерений направлены на оценку *истинных компонентов измерений* латентных качеств, а также исходят из аксиомы устойчивого существования истинных компонентов измерения, то стало понятным, что название *Latent Trait Theory* нельзя признать правильным.

Поэтому в последние годы стали применять второе название *Latent Trait Theory – Item Response Theory (IRT)*. В современных публикациях эту теорию часто называют «современной теорией», хотя сейчас существуют и другие «современные теории» (Random Sampling Theory; Theory of Generalizability), но практические преимущества некоторых положений этой теории в области педагогических измерений последних лет трудно переоценить

Главным в *IRT* являются утверждение о зависимости вероятности правильного ответа испытуемого от уровня подготовленности испытуемого и от параметров задания. Эту зависимость удобно представлять в виде *логистической функции*. Число рассматриваемых параметров ставится в соответствие одна из возможных моделей измерения.

Создатель *IRT* G.Rasch попытался формализовать идею зависимости результатов сопряжения множества испытуемых с множеством заданий, используя матрицу тестовых данных. Подход G.Rasch был априорным, теоретическим, направленным на создание математической модели

измерения и получения такой единицы педагогического измерения, с помощью которой можно было бы, *на одной шкале, сравнить уровень знаний студента с уровнем трудности задания.*

В *IRT* ставится важный исследовательский вопрос - как соотносятся между собой эмпирически наблюдаемые результаты испытуемого (сырые баллы) с соответствующими значениями θ_i того же испытуемого на латентной переменной θ ?

F.M.Lord сделал вывод, что *линия регрессии* тестового балла на *латентную переменную* имеет, в общем случае, вид логистической функции, откуда можно понять, что тестовый и истинный баллы связаны нелинейно. В.А. Аванесов предложил новый перевод *IRT* на русский язык как *математическая теория тестовых заданий*

Большой вклад в развитие классической теории тестов внес Льюис Гуттман (Louis Guttman, 1916-1987).

Всесторонне и полно классическая теория тестов впервые изложена в фундаментальном труде Гарольда Гулликсена (Gulliksen H., 1950 г.). С тех пор теория несколько видоизменялась, в частности совершенствовался математический аппарат. Классическая теория тестов в современном изложении приведена в книге Crocker L., Aligned J. (1986 г.). Из отечественных исследователей впервые описание этой теории дал В.Аванесов (1989г.). В работе Чельшковой М.Б. (2002 г.) приведены сведения о статистическом обосновании качества теста.

Классическая теория тестов основывается на следующих пяти основных положениях.

1. Эмпирически полученный результат измерения (X) представляет собой сумму истинного результата измерения (T) и ошибки измерения (E):

$$X = T + E$$

Величины T и E обычно неизвестны.

2. Истинный результат измерения можно выразить как математическое ожидание $E(X)$:

$$T = E(X)$$

3. Корреляция истинных и ошибочных компонентов по множеству испытуемых равна нулю, то есть

$$\rho_{TE} = 0.$$

4. Ошибочные компоненты двух любых тестов не коррелируют:

$$\rho_{E1,E2} = 0$$

5. Ошибочные компоненты одного теста не коррелируют с истинными компонентами любого другого теста:

$$\rho_{E1,T2} = 0$$

Кроме этого, основу классической теории тестов составляют два определения – параллельных и эквивалентных тестов.

ПАРАЛЛЕЛЬНЫЕ тесты должны соответствовать требованиям (1-5), истинные компоненты одного теста (T_1) должны быть равны истинным компонентам другого теста (T_2) в каждой выборке испытуемых, отвечающих на оба теста.

Предполагается, что $T_1 = T_2$ и, кроме того, равны дисперсии $S_1^2 = S_2^2$.

ЭКВИВАЛЕНТНЫЕ тесты должны соответствовать всем требованиям параллельных тестов за исключением одного: истинные компоненты одного теста не обязательно должны равняться истинным компонентам другого параллельного теста, но отличаться они должны на одну и ту же константу C .

Условие эквивалентности двух тестов записывается в следующем виде:

$$T_1 = T_2 + C_{12}$$

где C_{12} - константа различий результатов первого и второго тестов.

На основе приведенных положений построена теория надежности тестов.

Далее, примем в качестве исходного положения следующее утверждение

$$S_X^2 = S_T^2 + S_E^2$$

то есть, дисперсия полученных тестовых баллов равна сумме дисперсий истинных и ошибочных компонентов.

Перепишем это выражение в следующем виде:

$$\frac{S_T^2}{S_X^2} = 1 - \frac{S_E^2}{S_X^2}$$

Правая часть этого равенства представляет собой надежность теста (r). Таким образом надежность теста можно записать в виде:

$$r = 1 - \frac{S_E^2}{S_X^2}$$

На основе этой формулы в последующем были предложены различные выражения для нахождения коэффициента надежности теста. Надежность теста представляет собой его важнейшую характеристику. Если неизвестна надежность, то результаты тестирования невозможно интерпретировать. Надежность теста характеризует его точность как измерительного инструмента. Высокая надежность означает высокую повторяемость результатов тестирования в одинаковых условиях.

В классической теории тестов важнейшей проблемой является определение истинного тестового балла испытуемого (Т). Эмпирический тестовый балл (Х) зависит от многих условий – уровня трудности заданий, уровня подготовленности испытуемых, количества заданий, условий проведения тестирования и т.д. В группе сильных, хорошо подготовленных испытуемых, результаты тестирования будут как правило, лучше, чем в

группе слабо подготовленных испытуемых. В этой связи остается открытым вопрос о величине меры трудности заданий на генеральной совокупности испытуемых. Проблема заключается в том, что реальные эмпирические данные получают на вовсе не случайных выборках испытуемых. Как правило, это учебные группы, представляющие собой множество учащихся достаточно сильно взаимодействующих между собой в процессе учения и обучающиеся в условиях, часто не повторяющихся для других групп.

Найдем S_E из уравнения

$$s_E = s_X \sqrt{1 - r}$$

Здесь в явной форме показана зависимость точности измерения от величины стандартного отклонения S_X и от надежности теста r .

3.2. Обзор основных понятий общей теории педагогических измерений

Традиционной особенностью оценки знаний в нашей стране было оценивание, фактически, только объема знаний, выявляемого на том или ином уровне усвоения. Вместе с тем, при обучении **планируется некоторая идеальная структура знаний**. Выявление **меры соответствия** индивидуальной структуры знаний структуре знаний, планируемой при обучении, становится возможным **благодаря тестовым методам**.

Многолетний опыт использования педагогических тестов в зарубежной педагогике показывает, что посредством тестов возможно организовать непрерывную регистрацию «успехов и неудач» обучающихся, и, следовательно, **сделать вывод об их динамике**.

Понятия *тест*, *тестовое задание*, а также *латентные качества*, *истинный балл*, являются основными понятиями продолжающей формироваться «Общей теорией педагогических измерений». Наряду с «Общей теорией педагогических измерений» существуют разнообразные

«Частные теории педагогических измерений», занимающиеся тем или иным аспектом теории измерений.

В непосредственной связи с этими гуманитарными теориями, занимающимися формированием понятийного аппарата, находятся *общие и специальные математические теории педагогических измерений*, занимающиеся математическими оценками истинных баллов испытуемых, качества тестов, и.т.п. Для нахождения этих оценок используются различные математические методы (например, методы математической статистики), а также были созданы и активно используются специальные математические модели оценки истинных баллов испытуемых и трудностей заданий, как, например, в *Item Response Theory (IRT)*

Начнем с введения основных понятий общей теории измерений, рассмотрим основные моменты процесса построения тестовых заданий и теста в целом, вкратце опишем историю и классификацию теперь уже многочисленных теорий педагогических измерений, а затем займемся конкретной математической теорией в приложении к этой проблеме.

Понятие латентных качеств человека.

Латентные качества. Латентными называются различные качества личности, не поддающиеся непосредственному измерению. Например, требуется оценить такие качества как «уровень знаний по той или иной дисциплине», «подготовленность студентов», «интеллектуальное развитие». Попытки измерения подобных качеств на уровне обыденного сознания оканчиваются словесными или численными оценками, содержащими в себе немалые погрешности.

Понятия педагогического измерения, истинного балла.

В основе перехода к идее измерений лежит довольно условное и крайне необходимое предположение о непрерывном характере распределения баллов, полученных в результате тестирования репрезентативной выборки испытуемых. Соответственно возникла необходимость введения определения

педагогического измерения. Существует несколько достаточно близких друг другу определений педагогического измерения.

Например,

Педагогическое измерение — это процесс установления соответствия между оцениваемыми характеристиками обучающихся и точками эмпирической шкалы, в которой отношения между единичными оценками характеристик выражены свойствами числового ряда.

Роль единицы измерения играют тестовые задания, объектом оценки являются сами ученики, а результатом измерения — шкала баллов тестируемых учеников.

Педагогическое измерение – процесс отображения числами уровней проявления интересующих латентных качеств личности. В частности это касается измерения уровня знаний ученика (его истинного балла) в той или иной области. Средством получения этого истинного балла ученика является выполнение учеником определенным образом сконструированной системы тестовых заданий (теста).

Также существует несколько определений *истинного балла (true score)* испытуемых.

1) *Истинный балл* - это значение некоторого **параметра** ученика (*латентного качества*) в определенный момент времени; каждому ученику можно поставить в соответствие **единственное** на момент измерения **значение параметра**.

2) *Истинный балл* - это **предел среднего значения** наблюдаемых баллов, достигаемый при **бесконечном увеличении числа выполнений** учеником одного и того же теста.

3) В целом же можно считать, что *истинный балл* — это **показатель испытуемого в гипотетической генеральной совокупности заданий бесконечного теста**.

Данные определения носят теоретический характер. При обращении к практике, оценка истинного балла вычисляется по-разному, в зависимости от применяемой математической теории.

Шкала как средство отображения истинных баллов. Типы шкал

Исходным материалом для выявления истинного балла испытуемых являются сырые баллы испытуемых, полученные в результате тестирования. Процедуру интерпретации результатов тестирования должна предварять процедура выравнивания или перевода сырых баллов в одну из *стандартных шкал*, используемых в педагогических измерениях.

Шкала - средство фиксации результатов измерений путем упорядочения их в определенную систему чисел, в которой отношения между отдельными результатами выражены некоторыми числовыми значениями.

Традиционно существуют 4 типа шкал (две дискретные шкалы: номинальная и порядковая; 2 непрерывные шкалы: интервальная и шкала отношений)

В математической теории педагогических измерений **IRT** после обработки исходных сырых баллов тестируемых некоторой итерационной процедурой получают оценки уровней знаний испытуемых и трудностей заданий, выраженные значениями на одной и той же специальной *шкале логитов*, относящейся к типу шкалы отношений.

3.3. Этапы формирования теста

Первым этапом оценки *истинного балла* ученика является выполнение им *теста*. Процесс построения правильного *теста* является довольно трудоемким.

На первом этапе формируются *задания в тестовой форме* или *предтестовые задания* по данной дисциплине. Большое значение здесь играет правильный выбор *содержательной части* этих заданий. Обычно выбор содержания оценивается впоследствии независимыми экспертами.

Затем определенным образом оцениваются *задания в тестовой форме* и из этого множества формируется набор *тестовых заданий*. Затем формируется *тест*.

1. По *форме* все известные в теории и практике *тестовые задания* можно разделить на *четыре основные группы*.

Первую форму образуют задания с *выбором одного или нескольких правильных ответов*. Если к заданиям даются готовые ответы на выбор (обычно один правильный, остальные неправильные), то такие задания лучше называть заданиями с выбором *одного* правильного ответа. Задания с выбором *нескольких* правильных ответов труднее по содержанию. Этим заданиям предшествует инструкция — «Обвести кружком номера всех правильных ответов»

Вторую форму образуют задания, в которых правильный ответ надо дописать, обычно это одно слово или один знак. Заданиям этой формы предшествует стандартная инструкция — «*Дополнить*»

Третью форму образуют задания, состоящие из элементов двух столбцов. Таким заданиям предшествует инструкция — «Установить соответствие»

Четвертую форму образуют задания процессуального или алгоритмического толка. Каждому заданию предшествует инструкция — «Установить правильную последовательность». Испытуемый ставит цифры рангов в прямоугольниках, стоящих слева перед каждым элементом задания.

Сформировав набор *тестовых заданий* в одной или нескольких из этих форм, их проверяют на соответствие следующим требованиям:

- краткость;
- технологичность;
- правильность формы;
- логическая форма высказывания;

- одинаковость правил оценки ответов;
- наличие определённого места для ответов;
- правильность расположения элементов задания;
- одинаковость инструкции для всех испытуемых;
- адекватность инструкции форме и удержанию задания.

2. Затем из сформированного набора *предтестовых заданий* выбираются *тестовые задания*. Далеко не каждое *предтестовое задание* может быть *тестовым*.

Предтестовое задание превращается в *тестовое* после доведения его характеристик до уровня, соответствующего определенным научно обоснованным критериям качества.

Тестовое задание — это составная *единица теста*, отвечающая перечисленным выше требованиям формы, содержания и, кроме того, статистическим требованиям:

- 1) известной трудности;
- 2) достаточной вариации тестовых баллов;
- 3) положительной корреляции баллов задания с баллами по всему тесту.

Таким образом, для формирования *тестовых заданий* необходимы многократные эмпирические проверки *предтестовых заданий*, сопровождаемые обработкой и интерпретацией эмпирических результатов, и их переработка до тех пор, пока они не обретут *статус тестовых*.

3. Наличие достаточного числа *тестовых заданий*, полученных в результате обязательной эмпирической апробации *заданий в тестовой форме*, позволяет перейти к разработке *теста как системы*, обладающей *целостностью, составом и структурой*.

Исследование *системообразующих характеристик тестового задания* проводится методами корреляционного, факторного и латентно-структурного анализа

Педагогический тест — это система тестовых заданий возрастающей трудности, специфической формы, определённого содержания, создаваемая с целью аргументированной оценки уровня и структуры подготовленности обучаемых

Уровень и структура знаний выявляются при анализе ответов каждого ученика на все задания теста. Чем больше правильных ответов, тем выше индивидуальный тестовый балл испытуемых. Обычно этот тестовый балл ассоциируется с понятием «уровень знаний» и проходит процедуру уточнения на основе той или иной модели педагогического измерения.

Один и тот же уровень знаний может быть получен за счет ответов на различные задания. Например, в тесте из тридцати заданий испытуемый получил десять баллов. Можно предположить, что эти баллы, были получены за счет правильных ответов на первые десять сравнительно легких заданий. Полученную в этом случае последовательность единиц, а затем нулей называют *правильным профилем знаний*

Если же обнаруживается противоположная картина, когда испытуемый правильно отвечает на трудные задания и неправильно — на легкие, то это противоречит логике теста, и потому такой профиль знаний можно назвать *инвертированным*. Он встречается редко и чаще всего по причине ошибочности теста.

3.4. Классификация теорий педагогических измерений по В.С. Аванесову

В.С. Аванесов предложил следующую классификацию сейчас уже многочисленных теорий педагогических измерений.

1. Общая теория педагогических измерений.

Главным предметом такой теории должна быть *разработка понятийного аппарата теории педагогических измерений; изучение влияния*

заданий в тестовой форме, тестовых заданий и тестов на активизацию учебного процесса.

Другие предметы общей теории - *разработка целей и задач педагогических измерений*; создание программ проведения массового тестирования, изучение их положительных и отрицательных воздействий на личность, общество и государство.

К сугубо научным вопросам общей теории можно отнести *анализ понятий, положенных в основу измерений; анализ содержания заданий и теста в целом; разработка педагогических критериев эффективности и качества проводимых измерений*.

Попытки разработки общей теории педагогических измерений представлены в ряде публикаций. Некоторые построения *классической теории тестов* могут считаться частью общей теории педагогических измерений.

2. Частные педагогические теории.

С общей теорией непосредственно связаны частные теории, например, *логико-педагогическая теория композиции заданий в тестовой форме*. Это важно, поскольку каждое *нарушение формы, содержания*, того и другого – источники *погрешностей измерения*.

3. Общие математические теории педагогического измерения. (Random Sampling Theory; Theory of Generalizability)

Общие теории педагогических измерений непосредственно связаны с общими математическими теориями педагогических измерений, поскольку конечным результатом всех построений должно быть наиболее точное измерение *истинных баллов*. Для этого в математических теориях используются различные модели оценки этих баллов.

В частности, варианты *классической теории тестов* полезно рассматривать в связке с положениями *Random Sampling Theory* и более

общей Theory of Generalizability (Dependability), учитывая их общую опору на статистические теории индуктивного вывода.

4. Специальные математические теории педагогического измерения (Item Analysis; Item Response Theory)

К ним можно отнести частную математико-педагогическую *теорию тестовых заданий*, в которую целесообразно включить *часть положений классической теории тестов*, известной на Западе под названием «**Item Analysis**», а также постулаты, модели и вычислительные методы **IRT**.

3.5. Классификация тестов и тестовых заданий

Тест обладает составом, целостностью и структурой. Он состоит из заданий, правил их применения, оценок за выполнение каждого задания и рекомендаций по интерпретации тестовых результатов. Целостность теста означает взаимосвязь заданий, их принадлежность общему измеряемому фактору. Каждое задание теста выполняет отведенную ему роль и потому ни одно из них не может быть изъято из теста без потери качества измерения. Структуру теста образует способ связи заданий между собой. В основном, это так называемая факторная структура, в которой каждое задание связано с другими через общее содержание и общую вариацию тестовых результатов.

Существуют два основных вида тестов: традиционные и нетрадиционные.

Традиционный тест представляет собой единство, по меньшей мере, трех систем:

- содержательной системы знаний, описываемой языком проверяемой учебной дисциплины;
- формальной системы заданий возрастающей трудности;
- статистических характеристик заданий и результатов испытуемых.

Традиционный педагогический тест нужно рассматривать в двух существенных смыслах: - как метод педагогического измерения и как

результат применения теста. Удивительно, что тексты на русском языке тяготеют к смыслу метода, в то время как в большинстве работ западных авторов понятие тест чаще рассматривается в смысле результатов. Между тем, оба эти смысла характеризуют тест с разных сторон, потому что тест надо понимать одновременно и как метод, и как результат педагогического измерения. Одно дополняет другое. Тест, как метод, не мыслится без результатов, подтверждающих качество его самого и качество оценок измерения испытуемых различного уровня подготовленности.

В приведенном выше определении традиционного теста получили развитие несколько идей.

Первая идея - тест рассматривается не как обычная совокупность или набор вопросов, задач и т.п., а в виде понятия "система заданий". Такую систему образует не всякая совокупность, а только та, которая обуславливает возникновение нового интегративного качества, отличающего тест от элементарного набора заданий и от других средств педагогического контроля. Из множества возможных систем наилучшую образует та целостная совокупность, в которой качество теста проявляется в сравнительно большей степени. Отсюда вытекает мысль о выделении первого из двух главных системообразующих факторов - наилучшего состава тестовых заданий, образующих целостность. Исходя из этого, можно дать одно из самых коротких определений: тест - это система заданий, образующих наилучшую методическую целостность. Целостность теста - это устойчивое взаимодействие заданий, образующих тест как развивающуюся систему.

Вторая идея состоит в том, что в данном определении теста совершен отход от укоренившейся традиции рассмотрения теста как простого средства проверки, пробы, испытания. Всякий тест включает в себя элемент испытания, он не сводится весь к нему. Ибо тест - это еще и концепция, содержание, форма, результаты и интерпретация - все, требующее

обоснования. Этим подразумевается, что тест является качественным средством педагогического измерения. В соответствии с положениями теории, тестовые оценки не являются точными оценками испытуемых. Правильно говорить, что они лишь репрезентируют эти значения с некоторой точностью.

Третья идея, развиваемая в нашем определении традиционного теста - это включение нового понятия - эффективность теста, который ранее в литературе по тестам не рассматривался в качестве критерия анализа и создания тестов. Ведущая идея традиционного теста - минимумом числом заданий, за короткое время, быстро, качественно и с наименьшими затратами сравнить знания как можно большего числа учащихся.

По существу, этим отражается идея эффективности педагогической деятельности в области контроля знаний. Хотелось бы думать, что против самой этой идеи возражать уже некому и незачем. Если наш учитель может разъяснить учебный материал не хуже своего зарубежного коллеги, то хорошо проверить требуемые знания, у всех учеников, по всему изученному материалу, он не в состоянии из-за господствующей у нас классно-урочной системы занятий, нехватки компьютерной техники, тестов и программ для организации автоматизированного самоконтроля - самой гуманной формы контроля знаний. Не в состоянии он это делать и физически. В силу, мягко скажем, ошибочной социальной политики зарплата наших учителей уже давно не компенсирует затраты даже физической энергии, необходимой для хорошего преподавания, не говоря уже о повышенных затратах энергии интеллектуальной, что способно совершать только раскованное, а не озабоченное поисками хлеба мышление. Как отмечается в литературе, квалифицированный работник получает у нас в три-четыре раза меньше того уровня зарплаты, за границами которого нарушается нормальная жизнедеятельность и начинается разрушение трудового потенциала.

Хотя в литературе имеются сотни примеров определений теста, с которыми либо трудно, либо вообще нельзя согласиться, это совсем не означает, что данное определение традиционного теста - истина в последней инстанции. Как и все остальные понятия, оно нуждается в постоянном улучшении. Просто оно представляется более аргументированным, чем некоторые другие известные понятия педагогического теста. Впрочем, стремление к улучшению понятий - явление совершенно нормальное и необходимое для нормально развивающейся практики и науки. Конструктивные попытки дать другие определения теста или оспорить уже имеющиеся всегда полезны, но именно этого нам не достает.

К *традиционным* тестам относятся тесты гомогенные и гетерогенные. Гомогенный тест представляет собой систему заданий возрастающей трудности, специфической формы и определенного содержания - система, создаваемая с целью объективного, качественного, и эффективного метода оценки структуры и измерения уровня подготовленности учащихся по одной учебной дисциплине. Легко видеть, что в своей основе определение гомогенного теста совпадает с определением традиционного теста.

Гомогенные тесты распространены больше других. В педагогике они создаются для контроля знаний по одной учебной дисциплине или по одному разделу такой, например, объемной учебной дисциплины, как физика. В гомогенном педагогическом тесте не допускается использование заданий, выявляющих другие свойства. Наличие последних нарушает требование дисциплинарной чистоты педагогического теста. Ведь каждый тест измеряет что-то заранее определенное.

Гетерогенный тест представляет собой систему заданий возрастающей трудности, специфической формы и определенного содержания - система, создаваемая с целью объективного, качественного, и эффективного метода оценки структуры и измерения уровня подготовленности учащихся по

нескольким учебным дисциплинам. Нередко в такие тесты включаются и психологические задания для оценки уровня интеллектуального развития.

Обычно гетерогенные тесты используются для комплексной оценки выпускника школ, оценки личности при приеме на работу и для отбора наиболее подготовленных абитуриентов при приеме в вузы. Поскольку каждый гетерогенный тест состоит из гомогенных тестов, интерпретация результатов тестирования ведется по ответам на задания каждого теста (здесь они называются шкалами) и кроме того, посредством различных методов агрегирования баллов делаются попытки дать общую оценку подготовленности испытуемого.

Напомним, что традиционный тест представляет собой метод диагностики испытуемых, в котором они отвечают на одни задания, в одинаковое время, в одинаковых условиях и с одинаковой оценкой. При такой ориентации задачи определения точного объема и структуры освоенного учебного материала отступают, по необходимости, на задний план. В тест отбирается такое минимально достаточное количество заданий, которое позволяет сравнительно точно определить, образно говоря, не "кто что знает", а "кто знает больше". Интерпретация результатов тестирования ведется преимущественно на языке тестологии, с опорой на *среднюю арифметическую, моду или медиану и на так называемые процентильные нормы*, показывающие - сколько процентов испытуемых имеют тестовый результат хуже, чем у любого взятого для анализа испытуемого с его тестовым баллом. Такая интерпретация называется *нормативно-ориентированной*. Здесь вывод достраивается рейтингом: задания ответы выводы о знаниях испытуемого рейтинг, понимаемый как вывод о месте или ранге испытуемого.

Тесты можно классифицировать по следующим основаниям

1. *Предметная область применения тестов*: монопредметные, полипредметные, интегративные.

Интегративным можно назвать тест, состоящий из таких заданий, правильные ответы на которые требуют интегрированных (взаимосвязанных, обобщенных) знаний двух или большего числа учебных дисциплин. Использование таких тестов в школе, как контролирующих, так и обучающих, - отличное средство реализации межпредметных связей в обучении.

Интегративные тесты. Интегративным можно назвать тест, состоящий из системы заданий, отвечающих требованиям интегративного содержания, тестовой формы, возрастающей трудности заданий, нацеленных на обобщенную итоговую диагностику подготовленности выпускника образовательного учреждения. Диагностика проводится посредством предъявления таких заданий, правильные ответы на которые требуют интегрированных (обобщенных, явно взаимосвязанных) знаний двух и большего числа учебных дисциплин. Создание таких тестов дается только тем преподавателям, которые владеют знаниями ряда учебных дисциплин, понимают важную роль межпредметных связей в обучении, способны создавать задания, правильные ответы на которые требуют от учащихся знаний различных дисциплин и умений применять такие знания.

Интегративному тестированию предшествует организация интегративного обучения. К сожалению, существующая сейчас классно-урочная форма проведения занятия, в сочетании с чрезмерным дроблением учебных дисциплин, вместе с традицией преподавания отдельных дисциплин (а не обобщенных курсов), ещё долго будут тормозить внедрение интегративного подхода в процессы обучения и контроля подготовленности. Преимущество интегративных тестов перед гетерогенными заключается в большей содержательной информативности каждого задания и в меньшем числе самих заданий. Потребность создания интегративных тестов возрастает по мере повышения уровня образования и числа изучаемых учебных дисциплин. Поэтому попытки создания таких тестов отмечаются, в

основном, в высшей школе. Особенно полезны интегративные тесты для повышения объективности и эффективности проведения итоговой государственной аттестации учащихся и студентов.

Методика создания интегративных тестов сходна с методикой создания традиционных тестов, за исключением работы по определению содержания заданий. Для отбора содержания интегративных тестов использование экспертных методов является обязательным. Это связано с тем, что только эксперты могут определить адекватность содержания заданий целям теста. Но, прежде всего, самим экспертам важно будет определиться с целями образования и изучения тех или иных образовательных программ, а затем и договориться между собой по принципиальным вопросам, оставив для экспертизы лишь вариации в понимании степени значимости отдельных элементов в общей структуре подготовленности. Согласованный, по принципиальным вопросам, отобранный состав экспертов в зарубежной литературе нередко панелью. Или учитывая различия в смысле последнего слова, в русском языке, такой состав можно назвать представительной экспертной группой. Группа подбирается так, чтобы адекватно представлять подход, используемый при создании соответствующего теста.

2. *Общая ориентация замысла построения теста:* нормативно-ориентированные или критериально-ориентированные (предметно-ориентированные). Их главное отличие друг от друга заключается в способе интерпретации результатов измерения – тестовых баллов. (Любая интерпретация представляет собой сопоставление индивидуального результата измерения с некоторым «эталонном».)

При критериально-ориентированном подходе создаются тесты для сопоставления учебных достижений каждого ученика с планируемым к усвоению объемом знаний, умений или навыков. В этом случае в качестве интерпретационной системы отсчета используется конкретная область содержания, а не та или иная выборка учеников. При этом упор делается на

то, что может выполнить ученик и что он знает, а не на то, как он выглядит на фоне других.

Есть свои трудности и при критериально-ориентированном подходе. Как правило, они связаны с отбором содержания теста. В рамках критериально-ориентированного подхода в тесте стараются отразить все содержание контролируемого курса или, по крайней мере, то, что можно принять за этот полный объем. Процент правильного выполнения заданий рассматривают как уровень подготовки или как степень овладения общим объемом содержания курса. Конечно, в рамках критериально-ориентированного подхода для последней интерпретации есть все основания, так как тест включает все то, что можно условно принять за 100%.

Критериально-ориентированные тесты закрывают довольно широкий спектр задач. В частности, они помогают собрать полную и объективную информацию об учебных достижениях каждого учащегося в отдельности и группы учеников; сравнить знания, умения и навыки ученика с требованиями, заложенными в государственных образовательных стандартах; отобрать учеников, достигших планируемого уровня подготовленности; оценить эффективность профессиональной деятельности отдельных преподавателей и групп преподавателей; оценить эффективность различных программ обучения.

Понимание критериальной и нормативной ориентированности как разных подходов к интерпретации результатов мы находим у В.С. Аванесова.

С точки зрения интерпретации результатов:

Для критериально-ориентированной интерпретации вывод выстраивается вдоль логической цепочки: задания – ответы – выводы о соответствии испытуемого заданному критерию” (без сравнения индивидуальных результатов с результатами других испытуемых).

Для нормативно-ориентированной интерпретации вывод достраивается рейтингом: задания – ответы – выводы о знаниях испытуемого – рейтинг,

понимаемый как вывод о месте или ранге испытуемого”. Заметим, что норма устанавливается заранее и эмпирически.

Описание этих двух подходов в этой схематичной форме вполне отражает смысл разницы в интерпретации.

Для нормативно-ориентированного подхода характерно сопоставление индивидуального результата с результатами других учащихся. Нормативно-ориентированные тесты лучше всего подходят для решения задач, связанных со сравнением уровня учебных достижений, ранжированием и отбором:

- сравнение индивидуальных результатов учащихся со среднегрупповыми;
- сравнение учебных достижений отдельных испытуемых, проведение на этой основе ранжирования учащихся;
- сравнение между собой достижений учебных групп и учебных заведений;
- отбор фиксированного количества испытуемых, например, при приеме в учебные заведения или переходе на следующую ступень (уровень) образования.

В основе критериально-ориентированного подхода лежит сопоставление индивидуальных результатов тестирования с общим объемом знаний, который должен быть усвоен учащимся на данном этапе обучения.

Критериально-ориентированные тесты используются для решения задач, связанных с оценкой доли учебного материала, усвоенного испытуемым.

При нормативно-ориентированном подходе разрабатываются тесты для сравнения испытуемых по уровню учебных достижений.

Главным отличительным признаком предметно-ориентированного тестирования является интерпретация выполнения теста с точки зрения его смыслового содержания. Упор делается на строго определенную содержательную область (что тестируемые могут и что знают), а не на то, как они выглядят на фоне других.

3. *Дидактико-психологическая ориентация теста*: тест достижений для контроля знаний теории; тест достижений для контроля умений и навыков различной степени сложности по данному предмету, тест обучаемости (диагностики реальных учебных возможностей по данному кругу предметных или цикловых знаний – математической, лингвистической и т.п.).

4. *Ориентация на определенный этап контроля*: тесты предварительного контроля, тесты текущего контроля, тесты итогового контроля.

5. *Доминирующая деятельность испытуемого при выполнении тестов* – устные, письменные, компьютерные.

6. *Количество объектов контроля*: тесты, имеющие один объект контроля (например, количество выполняемых на должном уровне операций) или несколько (качество, количество, скорость, строгую последовательность, осознанность тех же операций).

7. *Степень гомогенности тестовых заданий*: тесты с однородными или разнородными формами построения заданий.

8. *Скоростной фактор*: скоростные (с обязательным фиксированием времени выполнения) и нескоростные.

9. *Форма организации тестирования*: массовые, индивидуальные, групповые.

Отдельно выделяют так называемые адаптивные тесты, основанные на принципе индивидуализации обучения. Каждый учитель понимает, что хорошему ученику нет смысла давать легкие и очень легкие задания, так же как нет смысла давать трудные задания слабому ученику. В теории педагогических измерений была найдена мера трудности заданий и мера уровня знаний, сопоставимые в одной шкале. После появления компьютеров эта мера легла в основу методики адаптивного контроля знаний, где трудность и число предъявляемых заданий регулируются в зависимости от ответов учеников.

Целесообразность адаптивного контроля вытекает из необходимости рационализации традиционного тестирования. Каждый учитель понимает, что хорошо подготовленному ученику нет необходимости давать легкие и очень легкие задания. Потому что слишком высока вероятность правильного решения. К тому же, легкие материалы не обладают заметным развивающим потенциалом. Симметрично, из-за высокой вероятности неправильного решения нет смысла давать трудные задания слабому ученику. Известно, что трудные и очень трудные задания снижают учебную мотивацию многих учащихся. Нужно было найти сопоставимую, в одной шкале, меру трудности заданий и меру уровня знаний. Эта мера была найдена в теории педагогических измерений. Датский математик Г. Раск назвал эту меру словом "логит". После появления компьютеров эта мера легла в основу методики адаптивного контроля знаний, где используются способы регулирования трудности и числа предъявляемых заданий, в зависимости от ответа учеников. При успешном ответе следующее задание ЭВМ подбирает более трудным, при неуспешном - легким. Естественно, этот алгоритм требует предварительного опробования всех заданий, определения их меры трудности, а также создания банка заданий и специальной программы.

Использование заданий, соответствующих уровню подготовленности, существенно повышает точность измерений и минимизирует время индивидуального тестирования до, примерно, 5 - 10 минут. Адаптивное тестирование позволяет обеспечить компьютерную выдачу заданий на оптимальном, примерно 50%-ом уровне вероятности правильного ответа, для каждого ученика.

В западной литературе выделяется три варианта адаптивного тестирования. Первый называется пирамидальным тестированием. При отсутствии предварительных оценок всем испытуемым дается задание средней трудности и уже затем, в зависимости от ответа, каждому испытуемому дается задание легче или труднее; на каждом шаге полезно

использовать правило деления шкалы трудности пополам. При втором варианте контроль начинается с любого желаемого, испытуемым, уровня трудности, с постепенным приближением к реальному уровню знаний. Третий вариант - когда тестирование проводится посредством банка заданий, разделенных по уровням трудности.

Таким образом, адаптивный тест представляет собой вариант автоматизированной системы тестирования, в которой заранее известны параметрами трудности и дифференцирующей способности каждого задания. Эта система создана в виде компьютерного банка заданий, упорядоченных в соответствии с интересующими характеристиками заданий. Самая главная характеристика заданий адаптивного теста - это уровень их трудности, полученный опытным путем, что означает: прежде чем попасть в банк, каждое задание проходит эмпирическую апробацию на достаточно большом числе типичных учащихся интересующего контингента. Слова "интересующего контингента" призвано представлять здесь смысл известного в науке понятия более строгого понятия "генеральная совокупность".

Распространенная у нас образовательная модель адаптивной школы Е.А. Ямбурга, исходит, по существу, из общих идей адаптивного обучения и адаптивного контроля знаний. Истоки такого подхода можно проследить с момента возникновения педагогических трудов Коменского, Песталоцци и Дистервега, которых объединяют идеи природосообразности и гуманности обучения. В центре их педагогических систем был Ученик. Например, в малоизвестной у нас работе А. Дистервега "Дидактические правила" можно прочесть такие слова: "Преподавай сообразно природе... Учи без пробелов... Начинай преподавание с того, на чем остановился ученик... Прежде чем приступить к преподаванию, нужно исследовать точку исхода... Без знания того, на чем остановился ученик, невозможно порядочно обучить его". Недостаточная информированность о реальном уровне знаний учеников и

естественные различия в их способностях усвоить предлагаемые знания стали главной причиной появления адаптивных систем, основанных на принципе индивидуализации обучения. Этот принцип трудно реализуем в традиционной, классно-урочной форме.

До появления первых компьютеров наиболее известной системой, близкой к адаптивному обучению, была так называемая "Система полного усвоения знаний".

Беспалько В.П. предлагает классификацию тестов, основанную на различных уровнях усвоения знаний, однако исходя из принятых выше определений понятий теста и тестового задания, данная классификация представляет собой скорее классификацию тестовых заданий.

Тесты 1 уровня – выполнение деятельности по узнаванию.

Тесты 2 уровня – работа на уровне репродукции.

Тесты 3 уровня – продуктивная деятельность.

Тесты 4 уровня – работа на уровне творческой деятельности.

Рассмотрим наиболее популярную классификацию тестовых заданий. В рамках данной классификации тестовые задания можно разделить на две группы:

- тестовые задания закрытого типа (с предписанными ответами, когда испытуемому необходимо выбрать из предложенных вариантов ответа тот или иной вариант);

- тестовые задания открытого типа (со свободными ответами, когда испытуемому необходимо самостоятельно дописать слово, словосочетание, предложение, знак, формулу и т.д.).

В рамках каждого типа можно выделить несколько видов тестовых заданий в зависимости от формы вариантов ответов.

Выбор типа и вида тестового задания определяется, прежде всего, целями, в соответствии с которыми проводится тестирование, характером материала, усвоение которого необходимо выявить, возрастными

особенностями испытуемых. Немаловажное значение играет финансовое, кадровое и ресурсное обеспечение, а также запас времени, которым располагает разработчик.

Виды тестовых заданий представлены в таблице 1:

Таблица 1.

Форма тестового задания		Инструкция
Закрытого типа	альтернативный выбор	испытуемый должен ответить «да» или «нет»
	установление соответствия	испытуемому предлагается установить соответствие элементов двух списков
	множественный выбор	испытуемому необходимо выбрать один или несколько правильных ответов из приведенного списка
	установление последовательности	испытуемый должен расположить элементы списка в определенной последовательности
Открытого типа	дополнение	испытуемый должен сформулировать ответы с учетом предусмотренных в задании ограничений (например, дополнить предложение)
	свободное изложение	испытуемый должен самостоятельно сформулировать ответ; никакие ограничения на них в задании не накладываются

Трудность в применении заданий открытого типа заключается в формализации ответов: их неоднозначность затрудняет стандартизацию, что является серьезной проблемой в случае проведения компьютерного тестирования, поскольку при создании эталона к такому заданию необходимо предусмотреть все возможные варианты ответа, учитывая

наличие и отсутствие знаков препинания, пробелов, характеристики шрифта и т.д.

Тестирование является значительным шагом на пути развития методики контроля за усвоением учащимися учебного материала. Введение тестирования позволяет осуществить плавный переход от субъективных и во многом интуитивных оценок к объективным обоснованным методам оценки результатов обучения. Однако, как и любое другое педагогическое нововведение, этот шаг должен осуществляться на строго научной базе, опираясь на результаты педагогических экспериментов и научных исследований. Тестирование не должно заменить традиционные методы педагогического контроля, а должно лишь в некоторой степени дополнить их.

Вопросы для самоконтроля

1. Какие компоненты педагогических измерений вы знаете?
2. Чем результаты оценивания качества подготовленности студентов, полученные на основе педагогических измерений, отличается от результатов традиционных экзаменов?
3. Каковы функции входного теста?
4. Каковы цели разработки корректирующих тестов?

Практические задания и упражнения по теме

1. Составить тест по одной из учебных дисциплин начальной школы с заданиями закрытой формы (задания с выбором одного или нескольких правильных ответов). Это простейший вид задания, в котором правильный ответ уже содержится, и задача испытуемого состоит в его узнавании. Рассмотрим основные элементы заданий с выбором правильного ответа. К ним относятся инструкции для испытуемых, содержание заданий,

форма, содержание и число ответов, а также оценки за правильность выполнения

2. Составить по 3 тестовых задания к каждому принципу по учебным дисциплинам начальной школы.

3. Используя задания открытой формы, составить тест (заданий) для промежуточного контроля по учебной дисциплине начальной школы.

4. Используя ЗАДАНИЯ НА УСТАНОВЛЕНИЕ ПРАВИЛЬНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ, составить тест для промежуточного контроля по учебной дисциплине начальной школы.

5. Используя ЗАДАНИЯ НА УСТАНОВЛЕНИЕ СООТВЕТСТВИЯ, составить тест (10 заданий) для промежуточного контроля по учебной дисциплине начальной школы.

Методические справки к заданиям по теме 3 представлены приложении 4.

ТЕМА 4. ХАРАКТЕРИСТИКИ ПЕДАГОГИЧЕСКОГО ТЕСТА И МЕТОДИКА ИХ РАСЧЕТА

Перечень вопросов, рассматриваемых в теме

- 1. Принципы отбора и критерии оценки содержания теста*
- 2. Процедуры статистической обработки результатов тестирования знаний и методы оценки качества теста*
- 3. Алгоритм расчета характеристик педагогического теста*

4.1. Принципы отбора и критерии оценки содержания теста

Акцент на содержательном подходе может оказать благотворное влияние на педагогическое тестирование в целом. От такого подхода выигрывает, например, интерпретация тестовых баллов при текущем контроле. Ученик получает информацию не о том, как он выглядит на фоне других, а о том, что он может делать и что знает по сравнению с заданными требованиями к уровню подготовки по предмету. Разумеется, такая интерпретация не исключает сочетания с отнесением результатов к нормам, что, как правило, происходит при текущем контроле знаний учеников в повседневном учебном процессе. В этом случае тестирование интегрировано с обучением и помогает учащемуся выявить возможные затруднения, а также своевременно исправить ошибки в усвоении содержания учебного материала.

В современной тестологии (В.С.Аванесов, М.Б.Челышкова, А.Н.Майоров и др.) различают 4 типа заданий в тестовой форме: задания на выбор одного или нескольких правильных ответов, задания в открытой форме или на дополнение, задания на установление правильной последовательности и задания на установление соответствий. Наиболее распространенной является первая форма.

Рассмотрим подробно каждую форму заданий по классификации В.С. Аванесова.

Задания на выбор одного или нескольких правильных ответов для компьютерного контроля знаний подходят больше всего. Такие задания удобно разделить на следующие виды: задания с двумя, тремя, четырьмя пятью и большим числом ответов. Инструкцией для этой формы заданий служит предложение: «Обведите (отметьте, укажите) номер правильного ответа».

Пример 1. Отметьте номер правильного ответа.

Место, которое занимает цифра в записи числа, называют

- 1) позицией;
- 2) разрядом;
- 3) положением;
- 4) знакоместом.

Задание должно быть сформулировано кратко и четко, так, чтобы его смысл был понятен при первом прочтении.

Содержание задания формулируется как можно яснее и как можно короче. Краткость обеспечивается тщательным подбором слов, символов, графиков, позволяющих минимумом средств добиваться максимума ясности смысла задания. Необходимо полностью исключить повторы слов, использование малопонятных, редко употребляемых, а также неизвестных учащимся символов, иностранных слов, затрудняющих восприятие смысла. Хорошо, когда задание содержит не более одного придаточного предложения.

Для достижения краткости в каждом задании лучше спросить о чем-нибудь одном. Утяжеление заданий требованиями что-то найти, решить и затем еще и объяснить отрицательно сказываются на качестве задания, хотя с педагогической точки зрения легко понять причину такой формулировки.

Еще лучше, когда короткими являются и задание и ответ. Неправильный, но правдоподобный ответ в американской тестовой литературе называется словом *дистрактор* (от английского глагола *to distract* -

отвлекать). В общем случае, чем лучше подобраны дистракторы, тем лучше бывает и задание. Талант разработчика проявляется в первую очередь в разработке эффективных дистракторов. Обычно считают, что чем выше доля выбора неправильного ответа, тем он лучше сформулирован. Следует отметить, что это верно только до известного предела; в погоне за привлекательностью дистракторов нередко теряется чувство меры. Привлекательность каждого ответа проверяется эмпирически.

Задания с выбором одного или нескольких ответов являются самой критикуемой формой. Сторонники привычных подходов утверждают, что по-настоящему проверить знания можно только в процессе непосредственного общения с учеником, задавая ему уточняющие вопросы, что помогает лучше прояснить подлинную глубину, прочность и обоснованность знаний. С подобными утверждениями надо согласиться. Однако есть еще вопросы экономии живого труда учителей и учащихся, экономии временных затрат и проблемы повышения эффективности образовательного процесса.

Нередко считается, что найти правильный ответ гораздо легче, чем формулировать его самому. Однако в хорошо сделанных заданиях незнающему ученику неправильные ответы часто кажутся более правдоподобными, чем правильные. Талант разработчика теста раскрывается в процессе создания именно неправильных, но очень правдоподобных ответов. Другое возражение - что тестовое задание с выбором одного или нескольких правильных ответов годится только для оценки знаний так называемого низшего уровня.

Выделяется вариант заданий *с выбором одного, наиболее правильного ответа из числа предложенных*. Соответственно пишется и инструкция к таким заданиям: «Обвести номер наиболее правильного ответа». Естественно предполагается, что все остальные ответы к заданиям правильные, но в различной мере.

Существует три основания для введения таких заданий в практику.

Первое - это старая идея исключения из заданий неправильных ответов, которые слабые учащиеся могут, якобы, запомнить. Если следовать этому очень спорному тезису, то неправильные ответы при тестировании вообще давать нельзя.

Второе основание для введения таких заданий в практику более реалистично. Оно касается необходимости формировать у учащихся не только умения отличать правильные ответы от неправильных, но и умения дифференцировать меру правильности ответов. Это действительно важно, как в общем среднем, так и в высшем профессиональном образовании.

Третье основание для применения заданий с выбором наиболее правильного ответа - это стремление проверить с их помощью полноту знаний.

Сколько бы ни были убедительны основания для введения таких заданий в практику, последние вряд ли могут найти широкое применение.

В заданиях *открытой формы* готовые ответы не даются: их должен придумать или получить сам тестирующийся. Иногда вместо термина «задания открытой формы» используют термины: «задания на дополнение» или «задания с конструируемым ответом». Для открытой формы принято использовать инструкцию, состоящую из одного слова: «Дополните».

Пример 2. Дополните.

В двоичной системе счисления $10-1=$ _____.

Задания на дополнение бывают двух заметно отличающихся видов:

1) с ограничениями, налагаемыми на ответы, возможности получения которых соответствующим образом определены по содержанию и форме представления;

2) задания со свободно конструируемым ответом, в котором необходимо составить развернутый ответ в виде полного решения задачи или дать ответ в виде микросочинения.

В заданиях с ограничениями заранее определяется, что однозначно считать правильным ответом, и задается степень полноты представления ответа. Обычно он бывает достаточно кратким - одно слово, число, символ и т.д. Иногда - более длинным, но не превышающим двух-трех слов. Естественно, что регламентированная краткость ответов выдвигает определенные требования к сфере применения, поэтому задания первого вида в основном используются для оценки достаточно узкого круга умений.

Отличительная особенность заданий с ограничениями на дополняемые ответы заключается в том, что они должны порождать только один, запланированный разработчиком правильный ответ.

Задания второго типа со свободно конструируемым ответом не имеют никаких ограничений на содержание и форму представления ответов. За определенное время учащийся может писать что угодно и как угодно. Однако тщательная формулировка подобных заданий предполагает наличие эталона, в качестве которого обычно выступает наиболее правильный ответ с описывающими его характеристиками и признаками качества.

В заданиях на установление соответствия преподаватель проверяет знание связей между элементами двух множеств. Элементы для сопоставления записываются в два столбца: слева обычно приводятся элементы задающего множества, содержащие постановку проблемы, а справа - элементы, подлежащие выбору.

К заданиям дается стандартная инструкция: «Установите соответствие».

Пример 3. Установите соответствие

Свойство	Формула
а) коммутативности	1) $(a+b)+c = a+(b+c)$
б) ассоциативности	2) $a+b = b+a$
в) дистрибутивности относительно	3) $(a+b)c = ac+bc$

сложения

$$4) a(b+c) = ab+ac$$

$$5) ab = ba$$

$$6) (ab)c = a(bc)$$

а) - _____, б) - _____, в) - _____.

Следует отметить, что желательно, чтобы в правом столбце элементов было больше, чем в левом. В этой ситуации возникают определенные трудности, связанные с подбором правдоподобных избыточных элементов. Иногда на один элемент левого множества необходимо выбрать несколько правильных ответов из правого столбца. Кроме того, соответствия могут быть расширены на три и большее число множеств. Эффективность задания существенно снижается, если неправдоподобные варианты будут легко различаться даже незнающими учащимися.

Эффективность задания также снижается в тех случаях, когда число элементов в левом и правом столбцах одинаково и при установлении соответствия для последнего элемента слева просто не из чего выбирать. Последнее правильное или неправильное соответствие устанавливается автоматически благодаря последовательному исключению элементов для предыдущих соответствий.

Тестовые задания на установление правильной последовательности предназначены для оценки уровня владения последовательностью действий, процессов и т.п. В заданиях приводятся в произвольном, случайном порядке действия, процессы, элементы, связанные с определенной задачей. Стандартная инструкция к этим заданиям имеет вид: «Установите правильную последовательность действий».

Пример 4. Установите правильную последовательность

Команда полного ветвления на учебного алгоритмического языка имеет формат:

? иначе <серия 2>

? кв

? то <серия 1>

? если <условие>

Задания на установление правильной последовательности получают доброжелательную поддержку у многих преподавателей, что объясняется важной ролью упорядоченного мышления и алгоритмов деятельности.

Цель введения таких заданий в учебный процесс - формирование алгоритмического мышления, алгоритмических знаний, умений и навыков.

Алгоритмическое мышление можно определить как интеллектуальную способность, проявляющуюся в определении наилучшей последовательности действий при решении учебных и практических задач. Характерные примеры проявления такого мышления - успешное выполнение различных заданий за короткое время, разработка самой эффективной программы для ЭВМ и т.п.

Выбор форм заданий определяется многими весьма противоречивыми факторами, в числе которых особенности содержания, цели тестирования, а также - специфика контингента испытуемых. Проверка проще при использовании заданий закрытой формы, однако, такие задания менее информативны. Задания открытой формы более информативны, но сложнее организовать их проверку. Еще более сложной задачей является создание компьютерных программ для проверки правильности ответов на такие задания. Это связано с богатством словарного запаса испытуемых (при ответе могут быть использованы синонимы), внимательностью (опечатки, несоответствие регистров) и т.п.

Для успешной ориентировки в формах заданий можно использовать специальную таблицу (см. таблицу 2) сопоставительного анализа заданий, предложенную М.Б. Чельшковой.

По мнению разработчика, настоящая таблица носит сугубо ориентировочный характер, однако, ее использование может облегчить процесс подбора тестовых заданий различной формы для решения тех или иных диагностических задач.

Соответствие заданий в тестовой форме требованиям педагогической корректности содержания и формы являются необходимыми, но недостаточными условиями для того, чтобы называть их тестовыми.

Превращение заданий в тестовой форме в тестовые задания начинается с момента статистической проверки каждого задания на наличие у них тестобразующих свойств.

Наличие достаточного числа тестовых заданий позволяет перейти к разработке теста как системы, обладающей целостностью, составом и структурой. На третьем этапе отбираются задания и создают тесты, повышаются качество и эффективность теста.

Целостность теста образует взаимосвязь ответов испытуемых на задания теста, наличие общего измеряемого фактора, влияющего на качество знаний.

Состав теста образует правильный подбор заданий, позволяющий минимально необходимым числом отобразить существенные элементы языковой компетентности испытуемых.

Таблица 2

Сопоставительный анализ характеристик тестовых заданий

Характеристики	Задания закрытой формы	Задания на дополнение	Задания на установление соответствия	Задания на установление последовательности
Проверка знания фактов	Годны	Годны	Годны	Годны
Применение знаний по образцу	Годны	Годны	Годны	Годны
Применение знаний в нестандартных ситуациях	Негодны	Годны	Негодны	Годны
Простота конструирования	Есть	Есть	Нет	Нет
Исключение угадывания	Не исключено	Исключено	Не исключено	Не исключено
Объективность	Да	Нет	Да	Да

Характеристики	Задания закрытой формы	Задания на дополнение	Задания на установление соответствия	Задания на установление последовательности
оценки				
Исключение описок	Нет	Да	Нет	Нет
Возможность оригинального ответа	Нет	Да	Да/Нет	Нет

Уровень и структура знаний выявляются при анализе ответов каждого испытуемого на все задания теста. Чем больше правильных ответов, тем выше индивидуальный тестовый балл испытуемых. Обычно этот тестовый балл ассоциируется с понятием "уровень знаний" и проходит процедуру уточнения на основе той или иной модели педагогического измерения. Один и тот же уровень знаний может быть получен за счет ответов на различные задания. Например, в тесте из тридцати заданий испытуемый получил десять баллов. Эти баллы скорее всего, получены за счет правильных ответов на первые десять, сравнительно легких заданий. Присущую для такого случая последовательность единиц, а затем нулей можно назвать правильной структурой подготовленности испытуемого. Если же обнаруживается противоположная картина, когда испытуемый правильно отвечает на трудные задания и неправильно - на легкие, то это противоречит логике теста и потому такой профиль знаний можно назвать инвертированным. Он встречается редко, и чаще всего, по причине ошибочности теста, в котором задания расположены с нарушениями требования возрастающей трудности. При условии, что тест сделан правильно, каждый профиль свидетельствует о структуре знаний. Эту структуру можно назвать элементарной (поскольку есть еще факторные структуры, которые выявляются с помощью методов факторного анализа).

По мнению ученых-тестологов наиболее важными характеристиками тестовых заданий являются **надежность и валидность**.

Надежным считается тест, который дает постоянные результаты, оценки при повторных предъявлениях. Существует несколько способов расчета коэффициента надежности теста. Например, коэффициент надежности K определяется как отношение дисперсии истинной компоненты к дисперсии измеренных тестовых баллов.

Валидность – комплексная характеристика теста, отражающая обоснованность, значимость результатов, адекватность теста целям измерения.

Уровень знаний в значительной степени зависит от личных усилий и способностей, в то время как структура знаний заметно зависит от правильной организации учебного процесса, от индивидуализации обучения, от мастерства педагога, от объективности контроля - в общем, от всего того, чего обычно не хватает. Путь к достижению этого идеала лежит через трудности создания качественных тестов.

Разработка тестов начинается с анализа содержания преподаваемых знаний и овладения принципами формулирования тестовых заданий. К сожалению, на тесты все еще смотрят как на средство, которое легко придумать, в то время как сильная сторона тестов - их эффективность, проистекающая из теоретической и эмпирической обоснованности.

На третьем этапе от разработчиков тестов потребуется некоторая математико-статистическая подготовка, знания теории тестов. Теорию тестов можно определить как совокупность непротиворечивых понятий, форм, методов, аксиом, формул и утверждений, способствующих повышению эффективности и качества тестового процесса. Кроме того, может потребоваться и некоторый опыт применения методов многомерного статистического анализа, и опыт правильной интерпретации тестовых результатов.

Часто возникает вопрос: «Как поведут себя удаляемые задания в других группах испытуемых?» Ответ зависит от качества подбора групп, а точнее от

статистического плана формирования выборочных совокупностей. Верный ответ на этот вопрос следует искать в смысле понятия «target group»; это множество испытуемых в генеральной совокупности, для которых предназначен разрабатываемый тест.

Соответственно, если задания проектируемого теста ведут себя неодинаково в разных группах, то это является, скорее всего, указанием на ошибки в формировании выборок испытуемых. Последние должны быть такими же однородными, как и испытуемые в целевой группе. На языке статистики это означает, что испытуемые в целевой и в экспериментальных группах должны принадлежать одной генеральной совокупности.

Логарифмические оценки, называемые логитами, таких, казалось бы, реально несопоставимых феноменов как уровень знаний испытуемого с уровнем трудности каждого задания, были использованы для непосредственного сопоставления уровня трудности с уровнем подготовленности испытуемого.

По мнению Беспалько В.П. и Татур Ю.Г., тестирование должно быть измерением качества усвоения знаний, умений и навыков. Сравнение правил выполнения задания (задачи), предложенного в тексте, с эталоном ответа позволяет определить коэффициент усвоения знаний (K_{us}).

Определение K_{us} является операцией измерения качества усвоения знаний. K_{us} поддается нормировке ($0 < K_{us} < 1$), процедура же контроля усвоения легко автоматизируется. По коэффициенту судят о завершенности процесса обучения: если $K_{us} > 0,7$, то процесс обучения можно считать завершенным.

При создании теста внимание разработчика, прежде всего, привлекают вопросы отбора содержания, которое можно определить как оптимальное отображение содержания учебной дисциплины в системе тестовых заданий. Требование оптимальности предполагает использование определенной

методики отбора, включающей вопросы целеполагания, планирования и оценки качества содержания теста.

Этап целеполагания является наиболее трудным и вместе с тем наиболее важным: от результатов его выполнения в первую очередь зависит качество содержания теста. В процессе целеполагания преподавателю необходимо решить вопрос о том, какие результаты учеников он хочет оценить с помощью теста.

Основания для ошибок в выводах педагога далеко не всегда связаны с технологическими недостатками традиционных средств контроля. Иногда они обусловлены недоработками педагога на этапе целеполагания, когда центр тяжести проверки смещается на второстепенные цели обучения, а иногда этап целеполагания отсутствует вовсе, поскольку часть педагогов уверена в непогрешимости своего опыта и интуиции, особенно при условии многолетней работы в школе. Однако никакие даже очень совершенные методы контроля и никакой опыт не дадут оснований для надежных выводов о достижении целей обучения до тех пор, пока нет уверенности в правильной постановке целей контроля и в их правильном, несмещенном отображении в содержании теста.

При создании теста ставится задача отобразить в его содержании то главное, что должны знать ученики в результате обучения, поэтому ограничиться простым перечислением целей обучения нельзя. В тест хотелось бы включить все, но, к сожалению, это невозможно, поэтому часть целей приходится просто отбросить и не проверять степень их достижения учащимися. Для того чтобы не утратить самое главное, необходимо структурировать цели и ввести определенную иерархию в их взаимное расположение. Без сомнения, здесь нет и не может быть готовых общих рецептов, поскольку в каждой дисциплине свои приоритеты. К тому же отдельные цели заметно связаны между собой, и потому простого

представления о системе целей как об упорядоченной совокупности без рассмотрения связей между элементами явно недостаточно.

После определения целей тестирования и их конкретизации необходимо разработать план и спецификацию теста.

При разработке плана делается примерная раскладка процентного соотношения содержания разделов и определяется необходимое число заданий, по каждому разделу дисциплины исходя из важности раздела и числа часов, отведенных на его изучение в программе.

Раскладку начинают с подсчета планируемого исходного числа заданий в тесте, которое затем в процессе работы над тестом будет неоднократно меняться в сторону увеличения или уменьшения. Обычно предельное число не должно превышать 60-80 заданий, поскольку время тестирования выбирают в пределах 1,5 - 2 ч., а на выполнение одного задания должно отводиться в среднем не более 2 мин.

После выполнения первого шага по планированию содержания разрабатывается спецификация теста, в которой фиксируется структура, содержание проверки и процентное соотношение заданий в тесте. Иногда спецификацию делают в развернутой форме, содержащей указания на тип заданий, который будет использоваться для оценки достижений учащихся в соответствии с намеченными целями создания теста, время выполнения теста, число заданий, особенности проведения тестирования, которые могут повлиять на характеристики теста и т.д.

Спецификация в развернутой форме включает:

- 1) цель создания теста, обоснование выбора подхода к его созданию, описание возможных сфер применения теста;
- 2) перечень нормативных документов, используемых при планировании содержания теста;
- 3) описание общей структуры теста, включающее перечень субтестов (если они есть) с указанием подходов к их разработке;

4) количество заданий различной формы с указанием числа ответов к закрытым заданиям, общее число заданий в тесте;

5) число параллельных вариантов теста либо ссылку на кластер, содержащую число и номера заданий кластера;

6) вес каждого задания, рекомендуемый автором теста;

7) рекомендуемое время выполнения теста, в том числе на каждый субтест, среднее время выполнения одного задания с учетом специфики формы;

8) соотношение заданий по различным разделам и видам учебной деятельности школьников;

9) рекомендации по контингенту учащихся для апробации теста;

10) охват требований стандартов (для аттестационных тестов);

11) перечень требований, не вошедших в тест (для аттестационных тестов);

12) рекомендуемую автором стратегию расположения заданий в тесте.

Знания и умения делятся таким образом:

А - знания понятий, определений, терминов;

В - знание законов и формул;

С - умение применять законы и формулы для решения задач;

Д - умение интерпретировать результаты на графиках и схемах;

Е - умение проводить оценочные суждения.

Нередко устанавливаются следующие пропорции:

А - 10%, В - 20%, С - 30%, Д - 30%, Е - 10%.

Помимо критериев, есть общие принципы, способствующие в определенной степени правильному отбору содержания тестов.

Принцип репрезентативности регламентирует не только полноту отображения, но и значимость содержательных элементов теста. Содержание заданий должно быть таким, чтобы по ответам на них можно было сделать

вывод о знании или незнании всей программы проверяемого раздела или курса.

Принцип системности предполагает подбор содержательных элементов, отвечающих требованиям системности и связанных между собой общей структурой знаний. При соблюдении принципа системности тест можно использовать для выявления не только объема знаний, но и для оценки качества структуры знаний учеников.

После отбора содержания теста начинается наиболее ответственный этап создания предтестовых заданий. Эта работа поручается обычно самым опытным преподавателям с большим стажем работы в школе. Однако для создания заданий одного опыта недостаточно. Необходимы также специальные знания по теории и методике разработки педагогических тестов, обеспечивающие профессиональный подход к созданию предтестовых заданий.

В.С. Аванесов выделил 3 критерия отбора содержания тестовых заданий:

- 1) определенность содержания теста;
- 2) непротиворечивость содержания заданий;
- 3) обоснованность содержания тестовых заданий.

Определенность содержания теста образует предмет педагогического измерения. В случае гомогенного теста возникает вопрос об уверенности в том, что все задания теста проверяют знания именно по определенной учебной дисциплине, а не по какой-то другой. Довольно часто случается так, что правильные ответы на некоторые задания требуют знаний не только интересующей дисциплины, но и ряда других, обычно смежных и предшествовавших учебных дисциплин. Близость и связанность которых затрудняет точное определение предметной принадлежности измеряемых знаний.

Во всяком тестовом задании заранее определяется, что однозначно считается ответом на задание, с какой степенью полноты должен быть правильный ответ. Не допускается определение понятия через перечисление элементов, не входящих в него.

Непротиворечивость содержания заданий требует, чтобы относительно одной и той же мысли не возникали суждения, одновременно утверждающие и отрицающие ее. Недопустимо существование двух исключających ответов на одно и то же задание теста. Если испытуемым дается инструкция: "Обведите кружком номер правильного ответа", а затем в одном из ответов утверждается, что правильного ответа нет, то это порождается пример непоследовательности мышления разработчика теста. В некоторых тестах встречаются ответы, вообще не связанные с содержанием задания. Таки ответы довольно легко распознаются испытуемыми как ошибочные, и потому тест оказывается неэффективным. Для повышения эффективности тест предварительно проходит апробацию на типичной выборке испытуемых. И если обнаружатся такие ответы к заданиям, которые испытуемые вообще не выбирают, то такие ответы из теста удаляются. Потому что они не выполняют функцию так называемых дистракторов, призванных отвлечь внимание незнающих испытуемых от правильного ответа. Кроме того, такие дистракторы вредны для теста, ибо снижают точность.

Обоснованность содержания тестовых заданий означает наличие у них оснований истинности. Обоснованность связана с аргументами, которые могут быть приведены в пользу той или другой формулировки заданий теста. При отсутствии доказательных аргументов в пользу правильности сформулированного задания оно в тест не включается, ни под каким предлогом. То же происходит, если в процессе экспертного обсуждения возникает хотя бы один контраргумент, или допускается условие, при котором данное утверждение может оказаться двусмысленным или ложным. Идея обоснованности содержания теста тесно переплетается с принципом

содержательной правильности тестовых заданий, о чем уже говорилось в предыдущей статье. Напомним, что в тест включается только то содержание учебной дисциплины, которое является объективно истинным и что поддается некоторой рациональной аргументации. Соответственно, спорные точки зрения, вполне приемлемые в науке, не рекомендуется включать в содержание тестовых заданий.

Неистинность содержания тестовых заданий отличается от некорректности их формулировки. Неистинность, как отмечалось выше, определяется соответствующим ответом, в то время как некорректно сформулированное задание может продуцировать ответы как правильные, так и неправильные, а то и вызывать недоумение. Сюда же можно отнести неточно или двусмысленно сформулированные задания, порождающие несколько правильных или условно правильных ответов. Отсюда возникает необходимость вводить дополнительные условия истинности, что удлиняет само задание и усложняет его семантику. Некорректность формулировки обычно выясняется в процессе обсуждения содержания заданий с опытными педагогами-экспертами. Успех такого обсуждения возможен при создании соответствующей культурной среды, где допустимы только конструктивные и тактичные суждения. Увы, опыт убеждает, что такое встречается не часто. Между тем, только совместное и доброжелательное обсуждение материалов разработчиками и экспертами способно породить атмосферу поиска наилучших вариантов содержания теста. Этот поиск практически бесконечен, и здесь нет истины в последней инстанции.

Знания, необходимые для целей тестирования, можно разделить на три вида: предлагаемые, приобретаемые и проверяемые.

Предлагаемые знания даются учащимся в форме учебных пособий, материалов, текстов, лекций, рассказов и т.п., отражающих основную часть образовательной программы. Эти знания формулируются, кроме того, в

системе заданий, по которым сами учащиеся могут проверить степень своей подготовленности.

Приобретаемые знания являются обычно только частью предлагаемых знаний, большей или меньшей, в зависимости от учебной активности учащихся. С развитием компьютерного обучения появились условия для превышения объема приобретаемых знаний над объемом предлагаемых знаний. Это новая ситуация, связанная с возможностями массового погружения учащихся в мировое образовательное пространство, в котором ведущая роль заданий в процессе приобретения знаний уже осознана достаточно хорошо. Решение учебных заданий является главным стимулом для активизации учения, собственной деятельности учащихся. Эта деятельность может протекать в форме работы с учителем, в группе или самостоятельно. Распространенные в литературе рассуждения об уровнях усвоения относятся исключительно к приобретаемым знаниям.

Проверяемые знания образуют основное содержание того документа, который может называться называется программой экзамена или тестирования, в зависимости от избираемой формы контроля знаний. Главной признаком проверяемых знаний является их актуальность, что означает готовность испытуемых к практическому применению знаний для решения заданий, используемых в момент проверки. В высшей школе этот же признак иногда называют оперативностью знаний.

В процессе тестирования обычно проверяются только такие знания, которые находятся в оперативной памяти, те, что не требует обращения к справочникам, словарям, картам, таблицам и т.п. В числе проверяемых знаний можно выделить еще нормативные знания, которые подлежат обязательному усвоению учащимися и последующему контролю со стороны органов управления образованием посредством экспертно подобранной и утвержденной руководящим органом системы заданий, задач и других контрольных материалов.

Кроме того, выделяются свойства знаний. В.И. Гинецинский выделяет следующие свойства знаний:

- рефлексивность (я не только знаю нечто, но и знаю, что я это знаю);
- транзитивность (если я знаю, что некто знает нечто, то из этого следует, что я знаю это нечто);
- антисимметричность (если я знаю кого-то, то это не значит, что он меня знает).

Для определения степени обученности по каждой учебной дисциплине выделяют объем знаний, которые необходимы для усвоения согласно учебной программе, что составляет базовый объем знаний. Базовые знания представляют минимум государственного образовательного стандарта. Однако и среди базовых знаний выделяют те, которые должны оставаться в памяти по любой дисциплине, в совокупности образуют мировоззренческие знания. Б.У. Родионов и А.О. Татур выделяют несколько звеньев мировоззренческих знаний: базовые знания, программные знания, сверхпрограммные знания. Педагогические тесты - единственный инструмент, позволяющий не только измерить обученность, но и умение использовать знания.

4.2. Принципы отбора и критерии оценки содержания теста

Рассмотрим самые простые и необходимые процедуры статистической обработки результатов тестирования знаний и методы оценки качества теста в соответствии с классической теорией тестирования.

Обозначим через x_{ij} числовую оценку успешности выполнения j -го задания, выполненного i -м испытуемым. Результаты тестирования обычно представляются в виде матрицы $\{x_{ij}\}$ с n строками и m столбцами ($i=1, \dots, n$; $j=1, \dots, m$). В практике тестирования принято, как правило, пользоваться дихотомической шкалой оценок результатов, когда множество возможных оценок состоит всего из двух элементов $\{0;1\}$: 0 – задание не выполнено, 1 –

выполнено правильно. Это, конечно, не единственно возможная шкала. Расчет, однако, ведется по формулам, приведенным ниже, независимо от выбранной для оценок шкалы.

Статистическая обработка матрицы результатов тестирования

Процесс статистической обработки матрицы результатов тестирования будем рассматривать последовательно, по шагам.

1 шаг. Вычисляются индивидуальные баллы испытуемых y_i ($i=1, \dots, n$), показывающие результат выполнения теста каждым студентом:

$$y_i = \sum_{j=1}^m x_{ij}$$

Поскольку для проверки статистических гипотез, которые применяются в классической теории тестов, используют предположение о нормальном распределении суммарных баллов испытуемых, то рекомендуется исследовать распределение частот. Для сравнения распределения баллов с нормальным можно использовать любой из критериев, применяемых обычно для этой цели.

2 шаг. Вычисляются средние результаты суммарных баллов испытуемых:

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

3 шаг. Вычисляются средние результаты испытуемых по каждому заданию:

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

Для дихотомических данных величины, вычисляемые по аналогичной формуле, обозначаются через p_j и традиционно называются в тестологии мерой трудности задания j ($j=1,2,\dots,m$):

$$p_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

Заметим, однако, что чем больше величина коэффициента p_j , тем большая часть испытуемых успешно справляется с заданием j . Так что на самом деле коэффициенты p_j ($j=1,2,\dots,m$) должны интерпретироваться как показатели легкости заданий.

4 шаг. Вычисляется дисперсия и стандартное отклонение суммарных баллов испытуемых:

дисперсия

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

стандартное отклонение

$$s_y = \sqrt{s_y^2}$$

5 шаг. Вычисляется дисперсия результатов испытуемых по j-ому заданию ($j=1, \dots, m$). Если успешность выполнения задания оценивается баллами 0 или 1, мера вариации определяется по формуле:

$$s_j^2 = p_j \times (1 - p_j)$$

Когда множество оценок состоит из более чем двух значений, применима формула:

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}$$

Вычислив дисперсию, можно найти и стандартное отклонение:

$$s_j = \sqrt{s_j^2}$$

6 шаг. Определяется связь каждого j-го задания ($j=1, \dots, m$) с суммой баллов по всему тесту. Для этого можно использовать коэффициент корреляции Пирсона:

$$R_j = \frac{\frac{\sum_{i=1}^n (x_{ij} \times y_i)}{n} - \bar{x}_j \times \bar{y}}{s_j \times s_y} \times \frac{n}{n-1}$$

7 шаг. Определяется попарная корреляционная связь заданий между собой. Здесь тоже можно использовать коэффициент корреляции Пирсона r_{jk} , ($j, k=1, 2, \dots, m$):

$$r_{jk} = \frac{\frac{\sum_{i=1}^n (x_{ij} \times x_{ik})}{n} - \bar{x}_j \times \bar{x}_k}{s_j \times s_k} \times \frac{n}{n-1}$$

Для дихотомических оценок успешности выполнения заданий тот же результат можно получить, оценив эту связь посредством коэффициента корреляции ($j, k=1, 2, \dots, m$) для такого рода данных:

$$\varphi_{jk} = \frac{A \times D - B \times C}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

где А– количество испытуемых, верно выполнивших задания j и k ;

В, - количество испытуемых, верно выполнивших задание j и неверно - задание k ;

С - количество испытуемых, неверно выполнивших задание j и верно задание k ;

Д - количество испытуемых, неверно выполнивших задания j и k .

Очевидно, величины А, В, С и D вычисляются по формулам:

$$A = \sum_{i=1}^n x_{ij} \times x_{ik}$$

$$B = \sum_{i=1}^n x_{ij} \times (1 - x_{ik})$$

$$C = \sum_{i=1}^n (1 - x_{ij}) \times x_{ik}$$

$$D = \sum_{i=1}^n (1 - x_{ij}) \times (1 - x_{ik})$$

8 шаг. Вычисляется индекс $I_j (j=1, 2, \dots, m)$ дискриминативности задания, то есть его различающая способность, указывающая на возможность разделять отдельных испытуемых по уровню выполнения теста в целом. Для этого из общей совокупности испытуемых выделяют две подгруппы – тех, кто получил самые высокие суммарные баллы, и тех, кто получил самые низкие. Тогда индекс дискриминативности может быть определен как разность между относительными численностями испытуемых, правильно выполнивших задание j в этих двух подгруппах. Например, упорядоченную совокупность суммарных баллов делят на три части и сравнивают результаты выполнения каждого задания j первой и последней третями испытуемых. В этом случае для дихотомических данных индекс приобретает вид:

$$I_j = \frac{\sum_{i=1}^{n/3} x_{ij} - \sum_{i=2n/3+1}^n x_{ij}}{n/3}$$

Чем больше коэффициент I_j , тем больше дискриминативность задания.

При наличии больших выборочных совокупностей дихотомических данных и нормального распределения индивидуальных сумм баллов рекомендуют рассчитывать для всех заданий бисериальные коэффициенты корреляции B_j ($j=1, 2, \dots, m$):

$$B_j = \frac{M_{j1} - M_{j0}}{s_y} \sqrt{\frac{n_{j0} \times n_{j1}}{n(n-1)}}$$

где M_{j1} – среднее арифметическое сумм баллов по всему тесту для испытуемых, получивших по данному заданию 1 балл;

M_{j0} – среднее арифметическое сумм баллов по всему тесту для испытуемых, получивших по данному заданию 0 баллов;

n_{j1} – число испытуемых, получивших по данному заданию 1 балл;

n_{j0} – число испытуемых, получивших по данному заданию 0 баллов.

Очевидно, входящие в формулу величины могут быть рассчитаны следующим образом:

$$n_{j1} = \sum_{i=1}^n x_{ij}, \quad n_{j0} = n - n_{j1},$$

$$M_{j1} = \frac{\sum_{i=1}^n x_{ij} \times y_i}{n_{j1}}, \quad M_{j0} = \frac{\sum_{i=1}^n (1 - x_{ij}) \times y_i}{n_{j0}}$$

9 шаг. Очередной шаг делается на основе вектора корреляций $\{R_j\}$ (или $\{B_j\}$), корреляционной матрицы j_{jk} (или r_{jk}) и вектора коэффициентов трудности $\{p_j\}$. Из собрания тестовых заданий удаляются задания, не обладающие дискриминативностью, то есть задания слишком легкие ($p_j > 0,9$) и слишком трудные ($p_j < 0,2$). Затем исключаются задания, плохо

коррелирующие с суммой баллов ($R_j < 0,15$), и имеющие отрицательные коэффициенты корреляции j_{jk} (или r_{jk}).

10 шаг. Для укороченного списка заданий вновь подсчитываются суммарные баллы испытуемых y_i . Затем составляется новая, упорядоченная, матрица данных тестирования, в которой столбцы располагаются в порядке возрастания трудности заданий, а строки – в порядке уменьшения, сверху вниз, суммарных баллов испытуемых. Для редуцированной матрицы пересчитываются средний суммарный балл, дисперсия суммарных баллов и коэффициенты корреляции заданий с суммой баллов.

Проверка качества теста

Чтобы полученное собрание тестовых заданий можно было считать тестом, оно должно удовлетворять определенным критериям надежности и валидности.

Надежность теста r тем выше, чем более согласованы результаты одного и того же человека при повторной проверке знаний посредством того же теста или эквивалентной его формы (параллельного теста). Согласованность результатов можно измерять коэффициентом корреляции Пирсона.

Если значения коэффициента r попадают в интервал 0,80-0,89, то говорят, что тест обладает хорошей надежностью, а если этот коэффициент не меньше 0,90, то надежность можно назвать очень высокой.

Другие, более практичные, методы оценки надежности теста, основаны на однократном применении единственной формы теста.

При применении метода расщепления откорректированную выше описанным образом тестовую матрицу разбивают на две половины, состоящие из заданий с четными и нечетными номерами. Коэффициент корреляции $r_{1/2}$ Пирсона между двумя совокупностями суммарных баллов результатов сам по себе уже может служить оценкой надежности всего теста.

Оценку надежности полного теста можно делать также с использованием коэффициента корреляции $r_{1/2}$, по формуле Спирмана-Брауна:

$$\rho = \frac{2r_{1/2}}{1 + r_{1/2}}$$

Другой способ оценки надежности расщепленного теста основан на формуле Рюлона:

$$\rho = 1 - \frac{S_d^2}{S_y^2}$$

где

S_y^2 - дисперсия суммарных баллов результата,

S_d^2 - дисперсия разностей между результатами каждого испытуемого по обеим половинам теста.

Она вычисляется по формуле:

$$S_d^2 = \frac{\sum_{i=1}^n [(y_i^{\text{чет}} - y_i^{\text{нечет}}) - \frac{\sum_{i=1}^n (y_i^{\text{чет}} - y_i^{\text{нечет}})}{n}]^2}{n - 1}$$

Здесь $(y_i^{\text{чет}} - y_i^{\text{нечет}})$, $(i=1,2,...,n)$ - разность сумм баллов в строках с номером i субматриц с четными и нечетными заданиями.

Еще один метод определения надежности, основанный на однократном предъявлении единственной формы теста, носит имя Кьюдера-Ричардсона. Он использует данные о выполнении испытуемыми каждого задания. Коэффициент надежности Кьюдера-Ричардсона вычисляется по следующей формуле:

$$\rho = \frac{s_y^2 - \sum_{j=1}^m s_j^2}{2s_y^2} + \sqrt{\left(\frac{s_y^2 - \sum_{j=1}^m s_j^2}{2s_y^2} \right)^2 + \frac{\sum_{j=1}^m R_j^2 s_j^2}{2s_y^2}}$$

Показано, что такой коэффициент равен среднему арифметическому значений коэффициентов надежности, найденных по методу расщепления при всех возможных разбиениях теста.

Чем выше показатель надежности, тем меньше стандартная ошибка измерения индивидуального результата. Показатель надежности можно использовать для построения доверительного интервала, в пределах которого с выбранной вероятностью P находится истинное значение оценки знаний испытуемого:

$$y_i \pm t_p \times s_y \sqrt{1 - \rho}$$

где t_p – значение статистики Стьюдента, найденное для выбранной вероятности P , когда число испытуемых равно n .

Валидность теста показывает, насколько хорошо тест делает то, для чего он был создан. 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:

$$V = \frac{\frac{\sum_{i=1}^n (Y_i \times y_i)}{n} - \bar{Y} \times \bar{y}}{s_Y \times s_y} \times \frac{n}{n-1}$$

где

$\bar{Y}$ - средняя арифметическая экспертных оценок,

s_Y - стандартное отклонение этих оценок:

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

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

Из двух тестов, предназначенных для одной и той же цели, более эффективен тот, который быстрее, дешевле и качественнее измеряет знания данной группы испытуемых.

4.3. Алгоритм расчета характеристик педагогического теста

Покажем алгоритм приведения теста в параметрам педагогического.

Считаем, что в результате тестирования получена матрица результатов. Она представляет собой исходную матрицу с непроранжированными заданиями.

Матрица сопряжения результатов тестирования, которая представлена в таблице 3, предварительно упорядочена по количеству правильных ответов на задание.

Таблица 3

Матрица сопряжения результатов тестирования

№ задания по тесту	0	7	2	9	4	5	1	6	8	3	Xi
ФИО	1	2	3	4	5	6	7	8	9	10	
<i>ФИО 1</i>	1	1	1	1	1	1	1	0	0	0	7
<i>ФИО 2</i>	1	1	0	0	0	0	0	0	0	0	2
<i>ФИО 3</i>	0	0	0	0	0	0	0	1	0	0	1
<i>ФИО 4</i>	1	1	0	1	1	1	1	1	1	1	9
<i>ФИО 5</i>	1	0	1	0	1	1	0	0	0	0	4
<i>ФИО 6</i>	1	1	1	0	0	0	0	0	0	0	3
<i>ФИО 7</i>	1	1	1	1	0	0	0	0	0	0	4
<i>ФИО 8</i>	1	1	1	1	0	0	0	0	0	0	4
<i>ФИО 9</i>	1	1	1	1	1	1	1	1	1	0	9
<i>ФИО 10</i>	1	1	1	1	1	0	1	0	0	0	6
гj	9	8	7	6	5	4	4	3	2	1	49

Xi -индивидуальный балл

гj - количество правильных ответов на j-задание

m - число заданий

n - число испытуемых

После предварительного упорядочения по гj матрицу следует упорядочить по индивидуальному баллу Xi. Получили упорядоченную матрицу сопряжения результатов тестирования, которая представлена в таблице 4.

Таблица 4.

Упорядоченная матрица сопряжения результатов тестирования

№ задания по тесту	0	7	2	9	4	5	1	6	8	3	Xi
n / m	1	2	3	4	5	6	7	8	9	10	
<i>ФИО 4</i>	1	1	0	1	1	1	1	1	1	1	9
<i>ФИО 9</i>	1	1	1	1	1	1	1	1	1	0	9
<i>ФИО 1</i>	1	1	1	1	1	1	1	0	0	0	7
<i>ФИО 10</i>	1	1	1	1	1	0	1	0	0	0	6
<i>ФИО 5</i>	1		1	0	1	1	0	0	0	0	4

№ задания по тесту	0	7	2	9	4	5	1	6	8	3	Xi
ФИО 7	1	1	1	1	0	0	0	0	0	0	4
ФИО 8	1	1	1	1	0	0	0	0	0	0	4
ФИО 6	1	1	1	0	0	0	0	0	0	0	3
ФИО 2	1	1	0	0	0	0	0	0	0	0	2
ФИО 3	0	0	0	0	0	0	0	1	0	0	1
rj	9	8	7	6	5	4	4	3	2	1	49
Qi	1	2	3	4	5	6	6	7	8	9	
Pj	0,9	0,8	0,7	0,6	0,5	0,4	0,4	0,3	0,2	0,1	
qj	0,1	0,2	0,3	0,4	0,5	0,6	0,6	0,7	0,8	0,9	

Q_j - количество неправильных ответов на j-задание

$$Q_j = n - r_j$$

P_j - относительная частота правильных ответов

$$P_j = r_j / n$$

q_j - относительная частота неправильных ответов

$$q_j = Q_j / n$$

4.3. Алгоритм расчета характеристик педагогического теста

4.3.1. Расчет надежности теста

$r_{н.т.}$ - коэффициент надежности теста, должен быть более 0,8

$$r_{н.т.} = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m P_j \cdot q_j}{S_x^2} \right)$$

S_x^2 - дисперсия

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - x_{cp})^2}{n-1}$$

x_{cp} - средний балл испытуемых

$$X_{cp} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Надежность рассчитываемого теста ($\Gamma_{н.т.} = 0,66$) мала.

4.3.2. Расчет валидности теста

На основании данных, приведенных в таблице, рассчитаем валидность теста:

$$r_v = \frac{M_B - M_H}{S_x} \cdot \sqrt{p_j \cdot q_j}$$

M_B - средний балл по всему тесту у испытуемых, успешно ответивших на j -задание.

M_H - средний балл по всему тесту у испытуемых, не ответивших на j -задание.

Рассчитывается валидность каждого задания, после этого вычисляется валидность всего теста (таблица 5).

Таблица 5.

Упорядоченная матрица сопряжения результатов тестирования с расчетом валидности каждого задания

№ исп./дозы	0	7	2	9	4	5	1	6	8	3	X_i
n / m	1	2	3	4	5	6	7	8	9	10	
4	1	1	0	1	1	1	1	1	1	1	9
9	1	1	1	1	1	1	1	1	1	0	9
1	1	1	1	1	1	1	1	0	0	0	7
10	1	1	1	1	1	0	1	0	0	0	6
5	1		1	0	1	1	0	0	0	0	4
7	1	1	1	1	0	0	0	0	0	0	4
8	1	1	1	1	0	0	0	0	0	0	4

№ исп./дозы	0	7	2	9	4	5	1	6	8	3	Xi
6	1	1	1	0	0	0	0	0	0	0	3
2	1	1	0	0	0	0	0	0	0	0	2
3	0	0	0	0	0	0	0	1	0	0	1
rj	9	8	7	6	5	4	4	3	2	1	49
Qi	1	2	3	4	5	6	6	7	8	9	
Pi	0,9	0,8	0,7	0,6	0,5	0,4	0,4	0,3	0,2	0,1	
qj	0,1	0,2	0,3	0,4	0,5	0,6	0,6	0,7	0,8	0,9	
gv	0,21	0,34	0,43	0,51	0,23	0,08	0,67	0,34	0,23	0,54	

Валидность теста должна быть более 0,4.

Валидность этого теста 0,35.

Последовательно исключаем задания, не удовлетворяющие намного (валидность менее 0,3) критерию валидности. (Процедура удаления заданий, не удовлетворяющих критериям валидности сама по себе значительно сложнее).

После исключения заданий в тесте осталось 6 заданий, которые удовлетворяют критериям валидности. Результат в таблице 6.

Таблица 6.

Упорядоченная матрица сопряжения результатов тестирования по валидным заданиям

№ задания	7	2	9	1	6	3	Xi	6	8	3	Xi
n / m	2	3	4	7	8	10		8	9	10	
4	1	0	1	1	1	1	5	1	1	1	9
9	1	1	1	1	1	0	5	1	1	0	9
1	1	1	1	1	0	0	4	0	0	0	7
10	1	1	1	1	0	0	4	0	0	0	6
7	1	1	1	0	0	0	3	0	0	0	4
8	1	1	1	0	0	0	3	0	0	0	4
6	1	1	0	0	0	0	2	0	0	0	4
5	0	1	0	0	0	0	1	0	0	0	3
2	1	0	0	0	0	0	1	0	0	0	2
3	0	0	0	0	1	0	1	1	0	0	1
rj	8	7	6	4	3	1		3	2	1	49
Qj	2	3	4	6	7	9		7	8	9	
Pi	0,8	0,7	0,6	0,4	0,3	0,1		0,3	0,2	0,1	

№ задания	7	2	9	1	6	3	Xi	6	8	3	Xi
qj	0,2	0,3	0,4	0,6	0,7	0,9		0,7	0,8	0,9	
rv	0,34	0,43	0,51	0,67	0,34	0,54		0,34	0,23	0,54	

После удаления невалидных заданий надежность теста повысилась до 0,89. Валидность этого теста стала 0,472.

Количество заданий в тесте должно остаться больше половины. Если для получения хорошей надежности теста остается меньше половины заданий, то необходимо конструировать другой тест (добавлять новые задания и переделывать имеющиеся).

Выводы

Полученный тест (6 заданий) обладает свойствами педагогического теста, т.е. надежность теста достаточная (0,89), валидность теста (0,472) отвечает требуемому критерию. Но диагностическая способность теста низкая, т.к.:

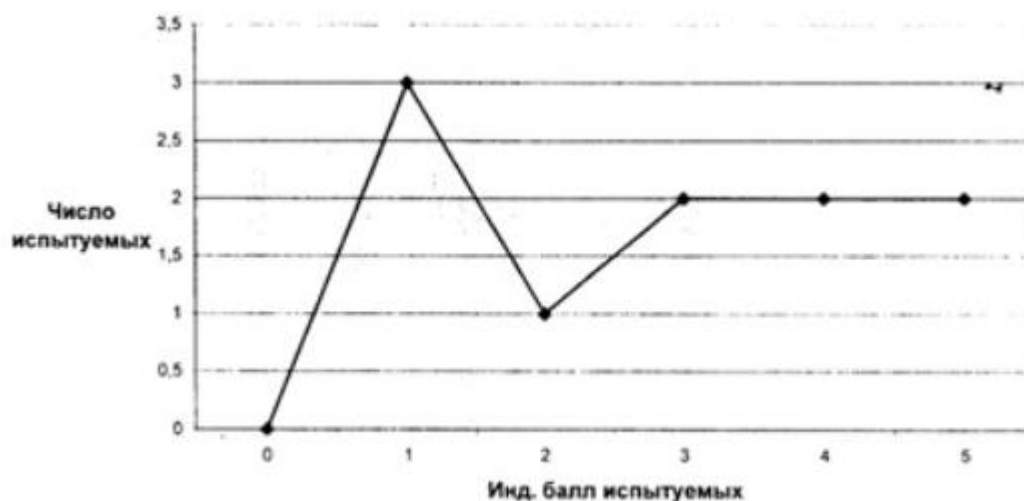
- распределение индивидуальных баллов испытуемых относительно средних баллов отличается от нормального, что видно из таблицы 7;
- исключены многие задания, т.е. тест не охватывает всю тему;
- тестирование проводится в области достаточно легких заданий.

Таблица 7.

Распределение числа испытуемых по индивидуальным баллам

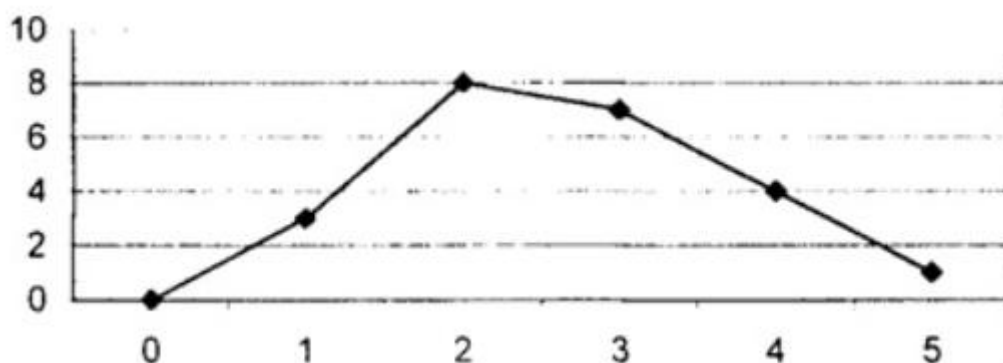
Индивидуальный балл, Xi	0	1	2	3	4	5
Кол-во испытуемых	0	3	1	2	2	2

Графически это выглядит следующим образом:



Нормальное распределение могло быть, например, следующим:

Индивидуальный балл, X_i	0	1	2	3	4	5
Кол-во испытуемых	0	3	8	7	4	1



Чтобы завершить разработку педагогического теста, нужно дополнить тест с отобранными заданиями новыми заданиями, провести новое тестирование и обработку результатов для большего количества испытуемых. Эту работу нужно будет проводить до тех пор, пока тест не будет удовлетворять всем требованиям валидности и надежности.

Вопросы для самоконтроля

1. Как связаны между собой показатели трудности и валидности?
2. Укажите два наиболее существенных недостатка классической теории тестов.

Практические задания и упражнения по теме

1. Составить итоговый тест по учебной дисциплине. Привести расчет валидности теста
2. Составить «портфолио» по дисциплине «Современные средства оценки результатов обучения».

Вам необходимо составить учебное «портфолио» по дисциплине «Современные средства оценки результатов обучения». Ответ оформить в виде папки

В учебные портфолио могут быть включены различные категории и наименования продуктов учебно-познавательной деятельности:

- работы студента (самостоятельные и аудиторные работы);
- оригинальные решения или выполнение сложных занимательных учебных заданий по конкретной теме (по выбору);
- выполнение заданий самостоятельно сверх учебной программы;
- рефераты различной направленности и сложности; наглядные пособия, изготовленные по конкретной тематике;
- краткие эссе или резюме о прочитанных книгах по данной теме;
- задания для одноклассников, составленные самим студентами по конкретной теме;
- описание проведенных мини экспериментов и лабораторных работ;
- аудио и видеоматериалы с выступлениями;
- работы, имеющие интегративные связи со смежными дисциплинами;
- дипломы, поощрения, награды, благодарственные письма по конкретным предметам.

Вопросы для итоговой аттестации
(ориентировочные формулировки вопросов к экзамену)

1. Что в педагогике понимается под средствами обучения?
2. Раскройте дидактические функции средств обучения.
3. Какие средства обучения относят к материальным, к идеальным?
4. Назовите средства обучения математике.
5. Что понимается под мониторингом качества образования?
6. Какова основная цель педагогического мониторинга?
7. Перечислите недостатки «традиционного» мониторинга.
8. Какие меры по повышению качества образования предпринимаются в отечественном образовании?
9. Перечислите основные направления деятельности федеральной системы тестирования.
10. Раскройте основные понятия в области педагогического тестирования.
11. Чем отличаются критериально- и нормативно- ориентированные тесты?
12. Раскройте критерии качества теста
13. Какие виды тестовых заданий существуют. Приведите примеры.
14. Что понимается под трудностью и сложностью задания?
15. Сколько уровней трудности выделяют? Раскройте каждый уровень.
16. Перечислите основные требования к тестовым заданиям.

Информационные источники

Основная литература:

1. Дистанционные образовательные технологии: проектирование и реализация учебных курсов / Лебедева М. Б., Агапонов С. В., Горюнова М. А., Костиков А. Н., Костикова Н. А., Никитина Л. Н., Соколова И. И., Степаненко Е. Б., Фрадкин В. Е., Шилова О. Н. / Под общ. ред. М. Б. Лебедевой. - СПб.: БХВ-Петербург, 2010. - 336 с. <http://znanium.com/bookread.php?book=350822>
2. Киселев, Г. М. Информационные технологии в педагогическом образовании [Электронный ресурс] : Учебник для бакалавров / Г. М. Киселев. - М.: Дашков и К, 2013. - 308 с. // <http://znanium.com/bookread.php?book=415216>
3. Крокер, Л. Введение в классическую и современную теорию тестов [Электронный ресурс] : учебник / Л. Крокер, Дж. Алгина; под общей ред. В. И. Звонникова и М. Б. Челышковой. - М.: Логос, 2010. - 668 с. // <http://znanium.com/bookread.php?book=468889>
4. Методологические основы психологии: Учебное пособие к практическим и семинарским занятиям для студентов психологических факультетов / Т.И.Чиркова - М.: Вузовский учебник: НИЦ ИНФРА-М, 2013. - 416 с. // <http://znanium.com/bookread.php?book=366333>
5. Создание системы оценивания ключевых компетенций учащихся массовой школы: Монография / А.В. Пашкевич. - М.: ИЦ РИОР: НИЦ Инфра-М, 2013. - 166 с. // <http://znanium.com/bookread.php?book=361932>
6. Современные средства оценивания результатов обучения : учебное пособие для студентов высших учебных заведений, обучающихся по педагогическим специальностям / В. И. Звонников, М. Б. Челышкова .- 4-е изд., стер. - Москва : Академия, 2011 .- 222

Дополнительная литература:

1. Методология качественных исследований в психологии: Учебное пособие / Н.П. Бусыгина. - М.: НИЦ Инфра-М, 2013. - 304 с. // <http://znanium.com/bookread.php?book=357385>

2. Модульные технологии: проектир. и разработка образоват. программ: Уч. пос. / О.Н. Олейникова, А.А. Муравьева, Ю.В. Коновалова, Е.В. Сартакова. - М.: Альфа-М: ИНФРА-М, 2010. - 256 с. <http://znanium.com/bookread.php?book=185177>

3. Общая и профессиональная педагогика: Учебник / Г.Н. Жуков, П.Г. Матросов. - М.: Альфа-М: НИЦ ИНФРА-М, 2013. - 448 с. // <http://znanium.com/bookread.php?book=403199>

4. Осипова, С. И. Математические методы в педагогических исследованиях [Электронный ресурс] : Учебное пособие / С. И. Осипова, С. М. Бутакова, Т. Г. Дулинец, Т. Б. Шаипова. - Красноярск: Сиб. федер. ун-т, 2012. - 264 с. // <http://znanium.com/bookread.php?book=442057>

5. Сердюков, В. А. ЕГЭ для родителей абитуриентов (математика, физика, информатика) [Электронный ресурс] / В. А. Сердюков. - М.: Дашков и К, 2013. - 152 с. // <http://znanium.com/bookread.php?book=430235>

Открытые отечественные и зарубежные образовательные ресурсы

- Архив учебных программ и презентаций - http://www.rusedu.ru/subcat_28.html

- Научный журнал "Наука и школа" (ВАК). - <http://nauka-i-shkola.ru/archive>

- Педагогическая библиотека - <http://www.pedlib.ru/katalogy/katalog.php?id=1&page=1>

- Социальная сеть работников начального образования - <http://nsportal.ru/nachalnaya-shkola>

- Учительский портал - <http://www.uchportal.ru/>

- Электронная версия газеты "Начальная школа" -
<http://nsc.1september.ru/index.php>

- Электронная версия журнала "Начальная школа плюс до и после" -
<http://www.school2100.ru/izdaniya/magazine>

Приложение 1

Образец оформления титульного листа к реферату

**Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Казанский (Приволжский) федеральный университет»**

Институт психологии и образование
Кафедра начального образования

Направление подготовки: 44.03.05 Педагогическое образование с двумя
профилями подготовки
(Начальное образование и иностранный (английский) язык)
(бакалавриат)

**РЕФЕРАТ НА ТЕМУ
«НАЗВАНИЕ РАБОТЫ»**

Выполнил:

студент __ курса группы __

И.И. Иванов

Проверил:

д.п.н., профессор В.Г. Закирова

Казань-20__

Приложение 2

Образец оформления доклада

Доклад на тему:

«Тема доклада»

Студентки группы №_____ Ивановой А.А.

Актуальность, новизна. Почему именно эту форму нужно в настоящее время изучать? Что сделано из того, что другими не было сделано, какие результаты получены впервые?

Автор формы (краткая справка). Дать краткую характеристику педагога, чей опыт представлен в Вашем выступлении.

Например, Карасева Ольга Николаевна, учитель высшей категории Гимназии №24 г. Казани. Является автором пособий по организации безопасной среды в начальной школе.

Описание формы. Необходимо ярко, интересно описать, прорекламировать выбранную форму работы.

Ваше отношение. Какие преимущества и недостатки? Где можно применять?

Источник. Ссылка на сайт, видео, библиографическое описание статьи. Библиографическое описание по ГОСТ Р 7.0.5-2008

Темы рефератов

1. История развития тестирования в России.
2. Основные подходы к оценке качества подготовки и способы их реализации.
3. Система оценивания и контроль качества образования в образовательных учреждениях разного типа и уровня.
4. Формы организации процесса обучения: традиции и новации.
5. Современные тенденции в оценивании учебных достижений.
6. Технологизация контрольно-оценочной деятельности на занятиях.
7. Виды контроля в учебном процессе.
8. Функции оценки в современном учебном процессе.
9. Педагогическое тестирование: проблемы и актуальность.
10. Тестирование как метод педагогического контроля.
11. Проблема выбора способа оценивания для различных профилей обучения.
12. Самостоятельная работа как условие эффективного усвоения учебного материала.
13. Критерии отбора содержания для составления тестовых заданий.
14. Понятие «портфолио» в современном образовательном процессе.
15. Эксперимент по введению Единого государственного экзамена: концепция, реализация, проблемы

Методические справки к заданиям по теме 3.

Задания закрытой формы (задания с выбором одного или нескольких правильных ответов)

Это простейший вид задания, в котором правильный ответ уже содержится, и задача испытуемого состоит в его узнавании. Рассмотрим основные элементы заданий с выбором правильного ответа. К ним относятся инструкции для испытуемых, содержание заданий, форма, содержание и число ответов, а также оценки за правильность выполнения.

Инструкции. В заданиях с выбором ответа применяют один из двух вариантов инструкций, соответствующих двум вариантам заданий. Инструкция помещается перед заданием, она печатается шрифтом, отличающимся от шрифта самого задания

Содержательная основа задания. Задание представляет собой часть утвердительного предложения, но не вопрос. Предлагаемые варианты ответов дополняют его до полного утвердительного предложения. Содержательная основа задания должна быть такой, чтобы для выбора правильного ответа достаточно было вспомнить и применить лишь то, что звучало на лекциях и было написано в рекомендованных для изучения пособиях. В конце содержательной части задания нет знаков препинания.

Ответы. Ответы должны быть содержательными и, по возможности, короткими. В конце ответов нет знаков препинания. Места для правильных ответов в разных заданиях выбираются случайным образом. Располагать ответы можно в одну, две и три колонки. Минимум средств (слов, символов, рисунков и графиков) должен обеспечивать максимальную ясность смысла задания. При подготовке ответов избегают повторов слов, применения малопонятных, редко употребляемых слов, а также неизвестных студентам или учащимся символов, иностранных слов, затрудняющих восприятие смысла.

Для объективной оценки знаний обучаемым предлагаются так называемые ответы-дистракторы (от англ. to distract - отвлекать). Дистракторы – это правдоподобные, но неправильные ответы. В хорошо составленном задании правильные и неправильные ответы испытуемыми, плохо знакомыми с предметом, выбираются с равной вероятностью (не

должно быть вариантов ответов, вводящих в заблуждение знающих учеников).

Ответы к заданиям могут быть выражены словами, числами, графиками. Они не должны представляться в форме "да" или "нет", "верно" или "неверно". Исключается применение таких вариантов ответа, как "правильного ответа нет", "все ответы правильные" или "все ответы неправильные".

Перечислим ряд требований, предъявляемых к заданиям закрытой формы:

1) в тексте задания должна быть устранена всякая двусмысленность или неясность формулировок;

2) основная часть задания формулируется предельно кратко, как правило, не более одного предложения из семи-восьми слов;

3) задание имеет предельно простую синтаксическую конструкцию, в основной текст задания вводится не более одного придаточного предложения;

4) в основную часть задания следует включать как можно больше слов, оставляя для ответа не более двух-трех наиболее важных, ключевых слов для данной проблемы;

5) все ответы к одному заданию должны быть приблизительно одной длины либо правильный ответ может быть короче других, но не во всех заданиях теста;

6) из текста задания необходимо исключить все вербальные ассоциации, способствующие выбору правильного ответа с помощью догадки;

7) частота выбора одного и того же номера места для правильного ответа в различных заданиях теста должна быть примерно одинакова либо номер места для правильного ответа выбирается в случайном порядке;

8) из ответов обязательно исключаются все повторяющиеся слова путем ввода их в основной текст заданий;

10) в ответах не рекомендуется использовать слова «все», «ни одного», «никогда», «всегда» и т. п., так как в отдельных случаях они способствуют угадыванию правильного ответа;

11) из числа неправильных исключаются ответы, вытекающие один из другого;

12) при формулировке дистракторов не рекомендуется использовать выражения «ни один из перечисленных», «все перечисленные» и т.п., так как они способствуют угадыванию правильного ответа;

13) из числа тестовых исключаются задания, содержащие оценочные суждения и мнения ученика по какому-либо вопросу;

14) все дистракторы к каждому заданию должны быть равновероятно привлекательными для испытуемых, не знающих правильного ответа;

15) ни один из дистракторов не должен являться частично правильным ответом, превращающимся при определенных дополнительных условиях в правильный ответ;

16) основная часть задания формулируется в форме утверждения, которое обращается в истинное или ложное высказывание после подстановки одного из ответов;

17) ответ на одно задание не должен служить ключом к правильным ответам на другие задания теста, т.е. не следует использовать дистракторы из одного задания в качестве ответов к другим заданиям теста;

18) все ответы должны быть параллельными по конструкции и грамматически согласованными с основной частью задания теста.

Принципы конструирования заданий с выбором одного или нескольких правильных ответов.

При конструировании заданий с выбором одного правильного ответа применяются две группы *принципов композиции*.

1. *Первая группа* используется при разработке ответов к заданиям. К ней относятся принцип противоречия, принцип противоположности, принцип однородности, принцип кумуляции, принцип градуирования, принцип удвоенного противопоставления.

2. *Вторая группа*, состоящая из принципа фасетности и принципа импликации, используется при разработке содержания заданий.

Рассмотрим принципы композиции разработки ответов.

1. Принцип противоречия. Если задание содержит **два ответа**, то, согласно *принципу противоречия*, второй ответ образуется из первого простым прибавлением отрицательной частицы “не”, отрицающих предлогов и слов так, что этим ответом отрицается смысл не самого задания, а содержания первого ответа:

Примеры:

Обвести кружком номер правильного ответа:

ПРЯМЫЕ $y=1+3x$ И $y=1-5x$

1) параллельны

2) не параллельны

ОРГАНИЧЕСКОЕ СТЕКЛО – ЭТО ПЛАСТМАССА, СОСТОЯЩАЯ ИЗ ТЕРМОПЛАСТИЧНОЙ СМОЛЫ

1) с наполнителями

2) без наполнителей

В этих заданиях отрицается смысл не самого задания, а содержания первого ответа.

2. Принцип противоположности.

Примеры ответов, построенных по *принципу противоположности*.

ПРОИЗВОДНАЯ ЛИНЕЙНОЙ ФУНКЦИИ

1) постоянная

2) переменная

ЯЧЕЙКИ ПРИ ОБРАЗОВАНИИ ПЕНОПЛАСТОВ

1) открыты

2) закрыты

ОТЛОЖЕНИЕ ЛИГНИНА В ОБОЛОЧКЕ РАСТИТЕЛЬНОЙ КЛЕТКИ
ПРОИСХОДИТ

1) снаружи

2) внутри

В заданиях с **тремя ответами** второй ответ может быть противоположен первому, а третий ответ – первому и второму, так что множество ответов полно.

Примеры:

ЕСЛИ $b > 1$, то функция $f(x) = \log_b(x)$

1) убывает

2) возрастает

3) немонотонна

Противоположность может быть введена также внутрь самих ответов.

ПРИ ОСТРОМ ИНФАРКТЕ МИОКАРДА АРТЕРИАЛЬНОЕ
ДАВЛЕНИЕ

1) повышается

2) понижается

3) вначале повышается, затем понижается

4) вначале понижается, затем повышается

3. Принцип однородности состоит в том, что ответы в заданиях должны быть однородными по форме. Не следует использовать в разных ответах одновременно глагол, прилагательное и существительное.

Примеры:

РАЗНОСТЬ ПОЛИНОМОВ – ФУНКЦИЯ

1) рациональная

2) полиномиальная

3) трансцендентная

КИСЛАЯ СРЕДА ОБРАЗУЕТСЯ ПРИ ГИДРОЛИЗЕ

1) карбоната натрия

3) нитрита свинца

2) сульфида калия

4) ацетата натрия

Усиливает эффективность заданий, сконструированных по принципу **однородности**, использование сходных по написанию и звучанию слов, похожих формул.

Примеры:

$$z=x^2+y^2 - \text{УРАВНЕНИЕ}$$

- 1) параболы
- 2) параболида
- 3) параболоида

ТИП ПИТАНИЯ ТИОНОВЫХ БАКТЕРИЙ

- 1) фотоавтотрофный
- 2) фотогетеротрофный
- 3) хемоавтотрофный
- 4) хемогетеротрофный

4. Принцип кумуляции. Суть принципа в том, что каждый следующий ответ содержит на один элемент больше, чем предыдущий:

Примеры:

ПРИ ЙОДОМЕТРИИ ОПРЕДЕЛЯЮТ СОДЕРЖАНИЕ

- 1) окислителей
- 2) окислителей, восстановителей
- 3) окислителей, восстановителей, кислот

СТЕПЕНЬ ПЛОДРОДИЯ ПОЧВ ОБУСЛОВЛИВАЕТ

- 1) совокупность элементов
- 2) структура и совокупность элементов
- 3) микроорганизмы, структура и совокупность элементов

5. Принцип сочетания. Применяя *принцип сочетания*, используют соединение двух, трех или четырех слов в каждом ответе. Можно, например, сочетать более или менее однородные и правдоподобные пары ответов. Ответы сочетаются также по правилу цепочки, причем последнее слово в первом ответе становится первым во втором ответе, последнее во втором – первым в третьем и т.д. Понятия сочетаются по два и по три.

Примеры:

$$\text{ФУНКЦИЯ } y=-2x^2+5x-1$$

- 1) непрерывна, дифференцируема
- 2) дифференцируема, монотонна
- 3) монотонна, непрерывна

ВОДА, ОЧИЩЕННАЯ В АПТЕКЕ, ЕЖЕДНЕВНО АНАЛИЗИРУЕТСЯ
НА СОДЕРЖАНИЕ

- 1) хлоридов и сульфатов
- 2) сульфатов и нитратов
- 3) нитратов и хлоридов

6. Принцип градуирования. Ответы в задании упорядочиваются по возрастанию какого-то количественного признака.

Пример:

НА РЕАЛИЗАЦИЮ ПРИРОДООХРАННЫХ МЕРОПРИЯТИЙ
КРАЕВОГО ЗНАЧЕНИЯ ОТ ВСЕЙ СУММЫ, ПОСТУПИВШЕЙ В
ЭКОЛОГИЧЕСКИЙ ФОНД, ВЫДЕЛЯЕТСЯ

- 1) 9%
- 2) 10%
- 3) 27%
- 4) 30%
- 5) 54%
- 6) 60%

ПРИ УВЕЛИЧЕНИИ ВЛАЖНОСТИ СВЫШЕ 30% ПРОЧНОСТЬ
ДРЕВЕСИНЫ

- 1) возрастает
- 2) не изменяется
- 3) снижается

7. Принцип удвоенного противопоставления применяется в заданиях с четырьмя ответами, части которых, построенные по принципу противоположности, сочетаются попарно.

Примеры:

СКОРОСТЬ РЕАКЦИИ ЗАВИСИТ ОТ ТЕМПЕРАТУРЫ

- 1) прямо пропорционально, линейно
- 2) прямо пропорционально, экспоненциально
- 3) обратно пропорционально, линейно
- 4) обратно пропорционально, экспоненциально

ЕСЛИ ДЛЯ ЛЮБОЙ ТОЧКИ (x,y) ИЗ НЕКОТОРОЙ ОКРЕСТНОСТИ
ТОЧКИ (a,b) ВЫПОЛНЯЕТСЯ УСЛОВИЕ $f(x,y) > f(a,b)$, ТО В ТОЧКЕ (a,b)

ФУНКЦИЯ

$f(x,y)$

ИМЕЕТ

1) относительный максимум

2) относительный минимум

3) абсолютный максимум

4) абсолютный минимум

Вторая группа принципов используется при разработке содержания заданий и состоит из принципа фасетности и принципа импликации.

8. Принцип фасетности содержания задания позволяет использовать фасеты в основном тексте задания. **Фасет** – это **форма записи нескольких вариантов одного и того же задания. Множество слов и словосочетаний, образующих фасет, помещается в столбик и обрамляется фигурными скобками.**

С помощью фасета создаются параллельные задания.

Фасет содержания задания может иметь большое количество вариантов. Например, список математических выражений может быть сделан сколь угодно длинным.

Пример:

$$\left\{ \begin{array}{l} 2x + 3y - 9z - 1 = 0 \\ 2x + 3y - 9z = 0 \\ \dots \end{array} \right\}$$

ПЛОСКОСТЬ

ЧЕРЕЗ НАЧАЛО КООРДИНАТ

1) проходит 2) не проходит

$$\left\{ \begin{array}{l} хлопкового \\ вискозного \\ шерстяного \end{array} \right\}$$

ПРИ НАМОКАНИИ

ВОЛОКНА ЕГО ПРОЧНОСТЬ

1) увеличивается

2) остается без изменений

3) уменьшается

9. Принцип импликации. Когда в содержательной части задания используется оборот “Если..., то” или эквивалентные ему, то имеет место применение *принципа импликации*.

Пример:

41. ЕСЛИ ПРИМЕНИТЬ ПОВЫШЕННЫЕ ДОЗЫ ФОСФОРНЫХ УДОБРЕНИЙ, ТО СНИЗИТСЯ ДОСТУПНОСТЬ

- 1) бора
- 2) меди
- 3) цинка
- 4) молибдена

При конструировании заданий в тестовой форме, как правило, применяются **сочетания принципов**, о которых шла речь выше.

Пример сочетания принципа фасетности и импликации.



ЕСЛИ В ПОЛДЕНЬ СТАТЬ СПИНОЙ К СОЛНЦУ, ТО БУДЕТ

- 1) юг
- 2) север
- 3) запад
- 4) восток

В заданиях с **выбором нескольких правильных ответов** применимы те же принципы конструирования, что и в заданиях с выбором одного правильного ответа. При этом верным может быть один ответ, несколько ответов или даже все предложенные ответы.

Примеры:

К ПОЛИНОМАМ ОТНОСЯТСЯ ФУНКЦИИ

- 1) $y = \frac{ax+b}{cx^3+dx}$
- 2) $y = ax+b$
- 3) $y = ax^2+bx+c$
- 4) $y = \frac{1}{ax+b}$

Если требуется указать не один верный ответ, а несколько из большого числа вариантов ответов, то без знания учебного материала угадать ответ маловероятно, так что информационная ценность заданий этого вида, вообще говоря, выше, чем у заданий с выбором одного правильного ответа.

В заданиях с выбором нескольких правильных ответов используется практически только два принципа композиции – **однородность ответов и фасеты** в основной части задания.

Разработчик теста с выбором одного правильного ответа вправе использовать любую обоснованную систему оценки выполнения заданий. Хотя в большинстве случаев за правильно выполненное задание принято давать один балл, а за неправильно выполненное – ноль (дихотомная оценка).

Задание с выбором нескольких правильных ответов считается выполненным правильно, если точно выбраны все без исключения правильные ответы. Как правило, один балл дается за правильно выполненное задание, ноль – за неправильно выполненное. Можно использовать и более широкую шкалу оценок. Например, за полностью выполненное задание давать 3 балла, за одну ошибку (один лишний ответ или один пропущенный) - 2 балла, и так далее, до нуля. Лишь бы правила оценки были определены заранее и известны испытуемым перед началом тестирования.

Задания открытой формы (задания на дополнения)

Задание открытой формы конструируется в виде утверждения, рядом с которым готовые ответы с выбором не приводятся. Испытуемый сам дописывает в отведенном для этого месте свой ответ так, чтобы в результате получилось истинное высказывание. Эта форма задания сводит возможность догадки к минимуму. С помощью заданий открытой формы проверяют знание названий, формул, имен, фактов, свойств, признаков, дат, причинно-следственных отношений.

1. Каждое задание должно быть нацелено только на одно дополнение, место для которого обозначается прочерком или точками.
2. Прочерк ставится на месте ключевого элемента, знание которого является наиболее существенным для контролируемого материала.
3. Все прочерки в открытых заданиях для одного теста рекомендуется делать равной длины.
4. Дополнения лучше ставить в конце задания или как можно ближе к концу.
5. После прочерка, если это необходимо, указываются единицы измерения.
6. Текст задания должен обладать предельно простой синтаксической конструкцией и содержать минимальное количество информации, необходимое для правильного выполнения задания.
7. В тексте задания нет повторы и двойное отрицание.

Инструкция к заданиям открытой формы может иметь вид:

ДОПОЛНИТЕ:

Если несколько заданий открытой формы следуют одно за другим, то инструкция пишется один раз перед всей группой таких заданий.

Чем ближе к концу фразы находится место для ответа, тем лучше понимается суть задания. В конце задания открытой формы ставится точка.

Примеры:

НИ ОДНОГО ЭЛЕМЕНТА НЕ СОДЕРЖИТ _____ МНОЖЕСТВО.

При создании заданий открытой формы используются такие *принципы композиции*, как логическая определенность содержания задания, фасетность, параллельность, обратимость, логическая соразмерность объема определяющего понятия объему определяемого, краткость,

неотрицательность и импликация. Рассмотрим на примерах применение этих принципов.

Выполняя задание, построенное в соответствии с *принципом логической определенности содержания*, испытуемый, знающий содержание предмета, легко находит ответ. Содержание и форма правильно сконструированного задания ведут его мысль прямо к правильному результату.

Пример:

СОВОКУПНОСТЬ ОБЪЕКТОВ, ОБЪЕДИНЕННЫХ ОБЩИМ ПРИЗНАКОМ ИЛИ СВОЙСТВОМ, НАЗОВЕМ _____.

Использование *принципа фасетности* обогащает возможности создания параллельных заданий.

Примеры (из арифметики):

$6+5=$ ____ . $6+$ ____ $=11$. ____ $+5=11$.

Посредством такого приема создаются параллельные задания для разных вариантов теста.

Еще один принцип конструирования заданий в тестовой форме – *принцип краткости*: чем меньше слов, тем меньше недоразумений:

Посредством заданий открытой формы можно проверять знание терминов.

Пример:

КРИВАЯ, К КОТОРОЙ ПРИБЛИЖАЕТСЯ ГРАФИК ДАННОЙ ФУНКЦИИ, НИКОГДА НЕ СЛИВАЯСЬ И НЕ ПЕРЕСЕКАЯ ЕЕ -

Несколько последовательных заданий открытого типа, предназначенных для проверки знания терминов, можно объединить в таблицу под условным названием “КАК НАЗЫВАЕТСЯ”. Ответы в этой таблице вписываются испытуемым на специально предназначенные для этого места во втором столбце.

Пример:

КАК НАЗЫВАЕТСЯ ПРОЦЕСС, ПРОИСХОДЯЩИЙ	ОТВЕТЫ:
1) ПРИ ПОСТОЯННОМ ДАВЛЕНИИ	_____
2) ПРИ ПОСТОЯННОМ ОБЪЕМЕ	_____
3) ПРИ ПОСТОЯННОЙ ТЕМПЕРАТУРЕ	_____

При оценке правильности выполнения заданий открытой формы, обычно используется дихотомическая шкала $\{0,1\}$. Однако в случаях, когда в

одном задании проверяется более одного элемента знаний, допустимо использовать и более широкие шкалы.

Задания на дополнение кажутся ученикам более трудными, так как в них исключается догадка. Это свойство делает задания на дополнение исключительно привлекательными для педагогов, особенно для тех, кто привык в своей работе опираться на традиционные средства контроля и не доверяет тестам.

Есть и недостатки. Они связаны с трудностями, возникающими при оценке ответов учеников. Дописывая ответ на месте прочерка, ученик может выбрать синонимы пропущенного запланированного разработчиком слова или изменить порядок следования элементов в пропущенной формуле, что значительно затрудняет проверку и оценку результатов учеников.

Основной сферой применения заданий на дополнение остается текущий контроль. Реже они используются в итоговом контроле из-за трудностей с проверкой, которая становится крайне трудоемкой, довольно субъективной и требует привлечения большого количества педагогов в случаях массового тестирования учеников

В тех случаях, когда требуется установить правильную последовательность действий или слов в определениях, используются задания на установление правильной последовательности. Это более сложный тип задания в тестовой форме, в процессе выполнения которого испытуемый конструирует ответ из предложенной неупорядоченной последовательности слов.

Задания на установление правильной последовательности используются для проверки знаний хода процесса, цепочки событий, действий и операций, а также определений и понятий. Они помогают формировать у учащихся алгоритмические мышление, знание и умение. Задания этой формы полезны как в качестве средства контроля знаний, так и в качестве средства обучения.

Инструкции к заданиям этой формы имеют вид:

УСТАНОВИТЕ ПРАВИЛЬНУЮ ПОСЛЕДОВАТЕЛЬНОСТЬ:

Пример:

“ПРАВИЛО ТРЕХ ШАГОВ” ОТЫСКАНИЯ ПРОИЗВОДНОЙ

-- избавиться от x в знаменателе

-- найти предел выражения при $x \rightarrow 0$

-- найти отношение приращения функции к приращению аргумента

ФАЗЫ ДЕЛЕНИЯ КЛЕТКИ ПРОТЕКАЮТ В СЛЕДУЮЩЕЙ ПОСЛЕДОВАТЕЛЬНОСТИ

-- анафаза

-- метафаза

-- профаза

-- телофаза

Испытуемый должен расставить в отведенных для ответов местах в начале каждой строчки (___) порядковые номера элементов последовательности действий.

Посредством заданий этой формы можно проверять знание последовательности доказательства теорем. Задания на установление правильной последовательности полезны при проверке знаний формулировок утверждений и определений.

Инструкция к заданию может сопровождаться необходимыми пояснениями. Название задания пишется заглавными буквами; оно декларирует, знание чего должен продемонстрировать испытуемый, причем ключевое слово в нем должно иметь именительный падеж.

Ранжируемые элементы в задании ставятся в случайном порядке. Чтобы окончания слов не служили подсказкой, их пишут в именительном падеже. Предлоги и союзы из множества ранжируемых элементов могут исключаться.

Чаще всего используется дихотомическая оценка правильности выполнения задания: 0 или 1. Однако можно применить и другое правило оценивания. Например, правильно выполненное задание оценивать 3 баллами, ошибка в конце задания снижает оценку на 1 балл, ошибка в середине – на 2 балла, а ошибка в начале – к снижению оценки выполнения всего задания до нуля.

Каждая из рассмотренных выше форм заданий может быть применена для проверки знаний. Выбор зависит от цели применения теста, специфики учебной дисциплины, от технологии проверки, сбора и обработки эмпирических данных, от технического и материального обеспечения процесса применения теста. Задания с выбором ответов наиболее удобны при итоговом контроле в силу ряда технологических преимуществ, повышающих эффективность процессов применения итогового теста и оценки результатов учеников. Для таких заданий легко организовать компьютеризированный сбор и анализ результатов. Задания на дополнение предпочтительны при обучающем контроле. Результаты их выполнения требуют ручной обработки. Как правило, для оценки результатов их выполнения приходится привлекать экспертов, а это требует дополнительных материальных затрат и времени на проверку. Задания на установление соответствия эффективны при самоконтроле и текущем контроле знаний.

В каждом тесте можно использовать несколько форм, но с точки зрения тестологов желательно выбрать одну форму и выдержать ее от первого до последнего задания. Совмещение различных форм в одном тесте значительно затрудняет подведение итоговых результатов тестирования.

Задания, где элементам одного множества требуется поставить в соответствие элементы другого множества, называют заданиями на установление соответствия. Задания на установление соответствия эффективны при самоконтроле и текущем контроле знаний. С их помощью проверяются так называемые ассоциативные знания, то есть знания о связи формы и содержания, сущности и явления, о соотношении между различными предметами, свойствами, законами. Испытуемый должен собрать ответ из элементов списка левой и соответствующих им элементов правой колонки.

Инструкция к заданиям этой формы имеет вид:

УСТАНОВИТЕ СООТВЕТСТВИЕ:

Номер задания ставится рядом с заголовком первого столбца. Заголовки пишутся прописными буквами. Заголовок каждого столбца относится ко всем элементам соответствующего множества. Элементы должны соответствовать названиям столбцов и быть короткими. Элементы левой колонки нумеруются цифрами, элементы правой – прописными буквами. Номер и буква отделяются от элементов столбца круглой скобкой и пробелом.

При разработке заданий на соответствие следует руководствоваться следующими правилами:

- 1) задание формулируется так, чтобы все содержание можно было выразить в виде двух множеств с соответствующими названиями;
- 2) элементы задающего столбца располагаются слева, а элементы для выбора справа;
- 3) желательно, чтобы каждый столбец имел определенное название, обобщающее все элементы столбца;
- 4) необходимо, чтобы правый столбец содержал хотя бы несколько дистракторов. Еще лучше, когда число элементов правого множества примерно в два раза больше числа элементов левого столбца;
- 5) необходимо, чтобы все дистракторы в одном задании были равновероятно правдоподобны;
- 6) элементы столбцов должны быть выбраны по одному основанию для включения только гомогенного материала в каждое задание теста;
- 7) в дополнительной инструкции к заданию необходимо сообщить ученику о наличии дистракторов в правом столбце, а также о том,

используется ли каждый элемент правого столбца только один раз или более одного раза;

8) все задания необходимо располагать на одной странице, не допуская переноса отдельных элементов.

На каждый элемент слева должен найтись, по крайней мере, один элемент справа, а каждому элементу справа должен соответствовать только один элемент слева. Испытуемый пишет ответы буквами из правого столбца в соответствующих пробелах рядом с номерами.

Пример:

УСТАНОВИТЕ СООТВЕТСТВИЕ

Процесс	Газовый закон
1) изотермический	А) $p/T = \text{const}$
2) изобарный	Б) $V/T = \text{const}$
3) изохорный	В) $pV = \text{const}$

ОТВЕТЫ: 1 ____; 2 ____; 3 ____.

В определенном смысле задания на установление соответствия довольно близки к заданиям в закрытой форме. И в том, и в другом случае при их выполнении ученик совершает одно и то же действие — выбирает из числа предложенных преподавателем правильный ответ. Не случайно задания на соответствие легко переделываются в ряд заданий закрытой формы, а закрытые задания в том случае, когда они имеют к нескольким содержательным основам одно и то же множество ответов, можно трансформировать в два множества для установления соответствия.

Оценивая правильность выполнения задания на установление соответствия, можно использовать дихотомическую шкалу: 1 за безусловно выполненное задание, 0 даже за одну допущенную ошибку. Однако можно выставить оценки на основе более широкой порядковой шкалы: 0 — “совершенно неверно”, 1- “есть верные элементы”, 2 — “много верных элементов” или “совершенно верно”, и т.д.

