

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

**Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ
(РГТЭУ)**

Научно-информационный материал

н5.74.7. Раздел: Подготовка населения по вопросам информационных технологий путем осуществления комплекса мероприятий по оказанию образовательных услуг малоимущим слоям населения САО г. Москвы на базе РГТЭУ для овладения основами информационных технологий и расширения возможностей трудоустройства.

н5.74.7.2.1.1. Подраздел: Разработка научно - информационных материалов.

"Методика разработки проведения тестовых занятий с обучаемыми".

Состав научно-образовательного коллектива:

Мосягин Александр Борисович - руководитель научно-образовательного Коллектива, к.т.н., доцент., заместитель декана ФИТ.

Нечаев Дмитрий Юрьевич - к.т.н., профессор, декан ФИТ.

Членов Анатолий Николаевич - д.т.н., профессор кафедры ИТиТ.

Подгрушный Александр Васильевич-к.т.н., доцент кафедры ИТиТ.

Коробков Вадим Владимирович-к.т.н., доцент кафедры ИТиТ.

Зыков Владимир Иванович, д.т.н., профессор кафедры ИиИБ.

Петренко Андрей Николаевич-к.т.н., доцент кафедры ИиИБ.

Олейников Владимир Тарасович-к.т.н., доцент кафедры ИиИБ.

Ковергина Оксана Сергеевна-лаборант кафедры ИТиТ.

Назарова Ламара Резовна-лаборант кафедры ИТиТ.

Москва, 2011г.

Оглавление

Введение.....	3
1. Обоснование введения системы тестирования и порядок её эксплуатации.....	5
1.1. Особенности оценки качества образования.....	5
1.2. Понятие теста в современной педагогической науке.....	8
1.3. Определение педагогического теста.....	10
1.4. Теоретические основы тестирования.....	14
1.5. Виды тестов.....	16
1.5.1. Традиционные тесты.....	16
1.5.2. Нетрадиционные тесты.....	21
1.6. Определение принципов разработки содержания теста для различных слоев потребителей.....	27
1.7. Результаты исследования подходов к определению трудности теста. Определение форм тестовых заданий для различных слоев населения.....	34
1.8. Разработка логических требований к содержанию теста.....	41
1.8.1. Определенность содержания теста.....	42
1.8.2. Непротиворечивость содержания заданий.....	44
1.8.3. Обоснованность.....	45
1.9. Композиция заданий в тестовой форме.....	47
2. Логико-информационная структура АПИМ.....	55
2.1. Требования, связанные с динамическим характером АПИМ.....	58
2.2. Требования по работе с плохо алгоритмируемыми задачами.....	62
3. Разработка модели оценивания при аттестационно-педагогических измерениях. Оценка эффективности предложений.....	63
3.1. Разработка требований к модели аттестационного педагогического измерителя.....	63
3.2. Предлагаемая модель методики обработки результатов аттестационно-педагогического измерения...	64
3.3. Математическая модель оценки результатов освоения дисциплины.....	66
3.4. Сравнение модельных и эмпирических результатов аттестационных педагогических измерений.....	71

Введение

За последнее десятилетие в России произошли значительные изменения в содержании, структуре и управлении системой образования. Важным шагом реформирования стала замена государственного контроля деятельности образовательных учреждений механизмом государственной регламентации их деятельности и государственной гарантии качества образования. Процесс трансформации содержания и структуры образовательной системы потребовал нового механизма управления качеством образования.

В этой системе аттестация выступает как один из этапов оценки выполнения образовательной организацией определенных требований государства. В результате аттестации устанавливается степень соответствия *содержания, уровня и качества подготовки* слушателей требованиям государственных образовательных стандартов (ГОС).

Возможны два подхода к проведению аттестации. Наиболее распространенным является подход, при котором *содержание* обучения оценивается через *анализ учебно-методической документации на соответствие требованиям ГОС*. При этом кадровый, информационный и материально-технический потенциал вуза характеризует степень обеспечения необходимого *уровня подготовки*, а экзаменационные оценки промежуточных аттестаций слушателей являются показателем *качества подготовки*. При рассматриваемом подходе оценка качества подготовки имеет низкую точность измерения, связанную, прежде всего с тем, что знание дисциплины в полном объеме оценивается по результатам выполнения *непредставительной выборки контрольных заданий экзаменационного билета*. Поэтому, при всей значимости такой оценки качества подготовки во время аттестации, в данном случае она не является основным фактором для принятия решения об аттестации. В связи с этим

повышенное внимание аттестующих органов к двум другим факторам, которые привлекаются для усиления (или ослабления) значения результатов прямого измерения, вызвано необходимостью компенсации большой погрешности последнего.

При другом подходе к аттестации оценка *уровня подготовки сотрудников основывается на объективных методах проведения педагогических измерений*. Для этого по каждой дисциплине учебного плана разрабатываются типовые **аттестационные педагогические измерительные материалы**, которые содержат генеральную совокупность контрольных заданий, обеспечивая необходимую точность измерений. Если при этом выполняются условия валидности по содержанию и уровню трудности требованиям ГОС, то тем самым с помощью аттестационных измерительных материалов оценивается качество освоения установленного содержания учебного материала на уровне требований ГОС.

Выполнение задачи проверки уровня знаний учащихся требует использования системы мероприятий, связанных единой целью, т.е. должно осуществляться в режиме технологии.

Разрабатываемая технология включает в себя модель формирования аттестационных педагогических измерительных материалов, методику обработки результатов и организацию педагогических измерений на основе выборочных методов как по отношению к дисциплинам, так и к контингенту обучающихся. Предложенная модель формирования аттестационных педагогических измерительных материалов позволяет конструировать измерители, валидные по отношению к требованиям ГОС для контролируемых дисциплин. Разработанная методика обработки результатов выявляет степень соответствия уровня подготовки слушателей требованиям ГОС и позволяет осуществить интеграцию результатов измерений по дисциплине, циклу дисциплин, профессиональной образовательной программе и учебному заведению в целом.

Все вышеизложенное доказывает высокую степень актуальности и практической значимости, выбранной для дипломного проектирования темы.

1. Обоснование введения системы тестирования и порядок её эксплуатации

1.1. Особенности оценки качества образования

Понятие "качество образования" имеет самые различные толкования, связанные в первую очередь с целями системы образования. Согласно исследованиям Международного института планирования образования, качество образования очень широкое понятие, дать однозначное определение которому просто невозможно. Для практических целей оно трактуется как "качественные изменения" в учебном процессе и в среде, окружающей обучающегося, которые можно идентифицировать как улучшение знаний, умений и ценностей, приобретаемых обучаемым по завершении определенного этапа.

Для характеристики качества образования вводятся определенные показатели, которые можно разделить на 3 группы: показатели вложения в образование, показатели качества учебного процесса и показатели результатов обучения. К последним относятся результаты оценки качества подготовки. В дальнейшем понятие "качество подготовки" рассматривается как результат сужения понятия "качество образования".

Анализ этого понятия можно вести на различных уровнях, один из которых проводится с позиции оценки качества результатов осуществленного учебного процесса. Подход к процедуре аттестации через педагогические измерения результатов учебного процесса - это основной принцип рассматриваемой здесь аттестационной технологии.

При оценке качества подготовки слушателей следует обратить внимание на две составляющие этой подготовки: образовательную и

профессиональную. Главной частью образования является образовательная составляющая, профессиональная направленность которой отражается в содержании образовательной программы в виде определенного набора дисциплин.

В результате профессиональной подготовки слушатель должен быть отнесен к определенной квалификационной категории, отражающей степень освоения им профессии по выбранной специальности. Однако попытки создания типовых систем оценки уровня освоения профессиональной составляющей подготовки слушателей (комплексные контрольные задания) не привели к положительному результату.

Следует отметить, что идет поиск путей усиления профессиональной составляющей подготовки слушателей.

Взаимодействие основных компонентов и подсистем в современной системе образования России приведено на рис. 1.1.

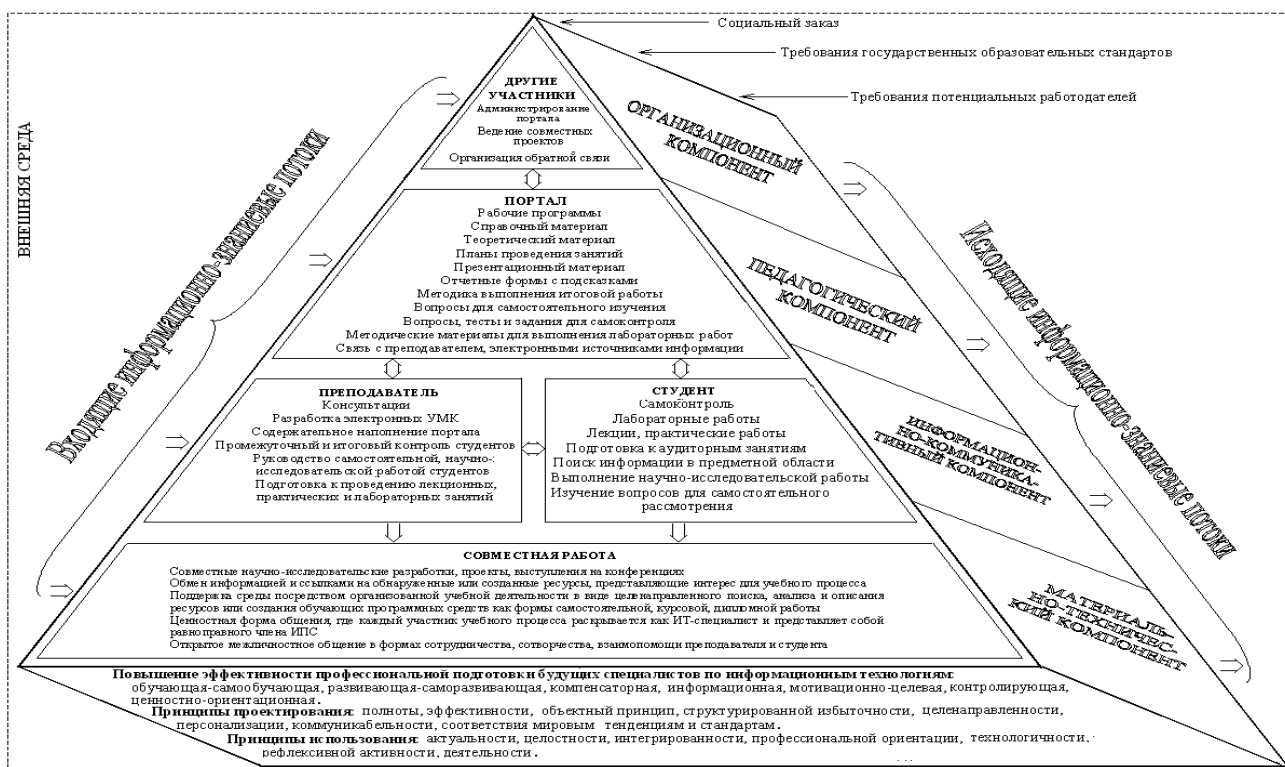


Рис.1.1. Взаимодействие основных компонентов и подсистем в современной системе образования России

Вузами накоплен большой опыт оценивания образовательной составляющей подготовки учащихся, что позволило создать типовые требования к содержанию и уровню обученности. Это нашло отражение, в частности, и в факте появления государственного образовательного стандарта высшего профессионального образования (ГОС ВПО), по отношению к требованиям которого проводится оценка содержания, уровня и качества подготовки слушателей.

Концепция единого интернет экзамена по оценке уровня освоения ГОС ВПО, предложенная Рособрнадзором и министерством образования приведена на рис. 1.2.

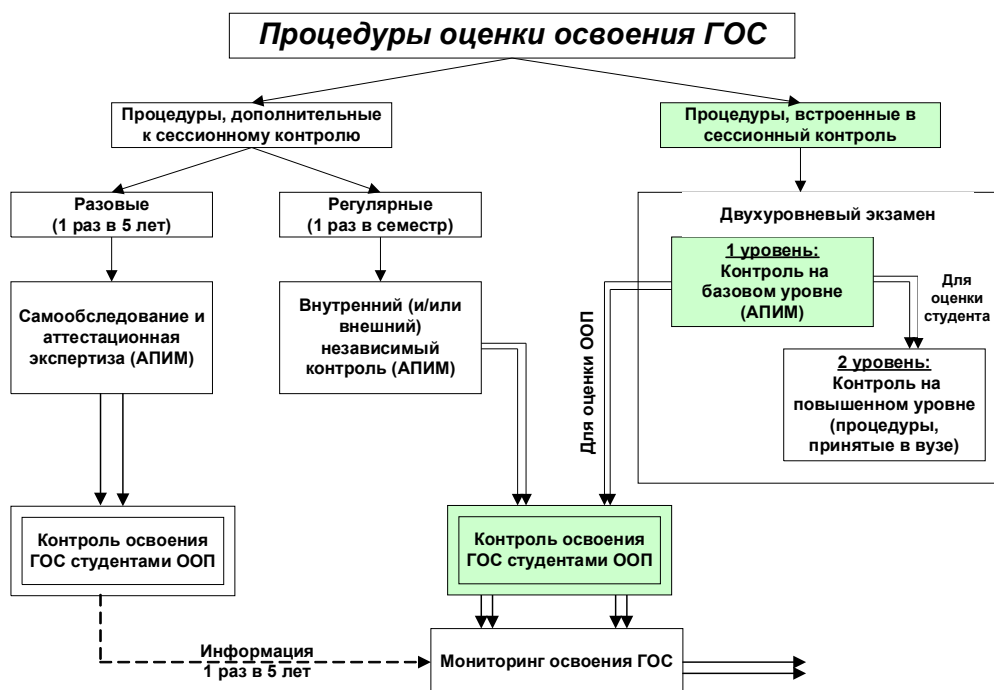


Рис. 1.2. Концепция единого интернет экзамена по оценке уровня освоения ГОС ВПО

Одной из важнейших составляющих подобной оценки и является система тестирования учащихся. Разработке принципов построения и использования тестирующих материалов посвящена данная выпускная квалификационная работа.

1.2. Понятие теста в современной педагогической науке

Что такое тест? Слово “тест” вызывает у преподавателей 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. 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 вопросу. Многочисленные примеры таких, казалось бы, "тестов" легко найти в газетно-журнальной периодике, в различных конкурсах и в многочисленных книжных публикациях под названием “Тесты”. Но 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 диагностики 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, устанавливаемых между испытуемыми по уровню проявляемых при тестировании 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 данной цели.

Тест определяется как система заданий возрастающей трудности, позволяющая качественно и эффективно измерить уровень и оценить структуру подготовленности учащихся. Такой тест назовем педагогическим. Смысл всех терминов, включенных в понятие педагогического теста, рассматривается в следующем разделе.

1.3. Определение педагогического теста

В предыдущем пункте педагогический тест определялся как система заданий возрастающей трудности, специфической формы, позволяющая качественно и эффективно измерить уровень и оценить структуру подготовленности учащихся. Для лучшего понимания этого определения полезно дать краткое истолкование его основных терминов.

Система означает, что в тесте собраны такие задания, которые обладают системообразующими свойствами. Здесь, в первую очередь, надо выделить общую принадлежность заданий к одной и той же системе знаний, т.е. к одной учебной дисциплине, их связь и упорядоченность. В педагогическом тесте задания располагаются по мере возрастания трудности - от самого легкого до самого трудного. Иначе говоря, главным формальным системообразующим признаком теста является различие заданий по степени их трудности. Слово “формально” появилось здесь для того, чтобы оградить читателя от соблазна трактовать тест единственным образом - только как систему заданий возрастающей трудности, без учета их содержания.

Специфическая форма тестовых заданий отличается тем, что задания теста представляют собой не вопросы и не задачи, а задания, сформулированные в форме высказываний, в зависимости от ответов, истинных или ложных. Традиционные вопросы, напротив, истинными или

ложными не бывают, а ответы на них нередко настолько неопределенны и многословны, что для выявления их правильности требуются заметные, в суммарном исчислении, затраты интеллектуальной энергии преподавателей. В этом смысле традиционные вопросы и ответы нетехнологичны, и потому их лучше не включать в тест.

Определенное содержание означает использование в тесте только такого контрольного материала, который соответствует содержанию учебной дисциплины; остальное в педагогический тест не включается, ни под каким предлогом. Например, проверка уровня интеллектуального развития является предметом психологического тестирования. Содержание теста существует, сохраняется и передается в одной из четырех основных форм заданий. Вне тестовых форм ни тест, ни его содержание не существуют. Теоретически оправданным критерием упорядочения содержания так называемого гомогенного теста является критерий трудности заданий. Внеучебное содержание (например, проверка уровня интеллектуального развития) в педагогический тест не включается. Это предмет психологического измерения.

Возрастающую трудность заданий можно образно сравнить с барьерами на беговой дорожке стадиона, где каждый последующий выше предыдущего. Пробежать дистанцию и успешно преодолеть все барьеры сможет только тот, кто лучше подготовлен. Поскольку в педагогическом тесте задания упорядочиваются по принципу возрастающей трудности, одни испытуемые "заваливаются" уже на самом легком, первом задании, другие - на последующих заданиях. Учащиеся среднего уровня подготовленности могут ответить правильно только на половину заданий теста и, наконец, только самые знающие в состоянии дать правильный ответ на задания самого высокого уровня трудности, расположенные в конце теста. Трудность задания может определяться двояко: а) умозрительно, на основе предполагаемого числа и характера умственных операций, необходимых для

успешного выполнения заданий, и б) после эмпирического опробывания заданий, с подсчетом доли неправильных ответов. В классической теории тестов многие годы рассматривались только эмпирические показатели трудности. В новых вариантах психологических и педагогических теорий тестов больше внимание стало уделяться характеру умственной деятельности учащихся в процессе выполнения тестовых заданий различных форм.

Показатель трудности задания рассматривается как важный системно-, и одновременно, структурообразующий фактор теста. К этому можно добавить еще один критерий - это критерий логической определенности тестового задания. Его можно сформулировать словами, близкими к формулировке Х.Карри: задание является определенным, если на него можно ответить утвердительно или отрицательно, и если существует эффективный процесс для нахождения такого ответа.

Ответ на задание педагогического теста представляет собой краткое суждение, связанное по содержанию и по форме с содержанием задания. Каждому заданию ставятся в соответствие ответы правильные и неправильные. Критерии правильности заранее определяются авторами теста. Вероятность правильного ответа на любое задание зависит от соотношения уровня знаний испытуемого и уровня трудности задания. При наличии сопоставимых шкал, эту вероятность выражают значением от нуля до единицы. Оценка ответов по степени их правильности проводится довольно редко, но при необходимости создаются задания с такими ответами, которые правильны в различной степени. Инструкция для испытуемых в таких случаях может быть такой: “Обведите кружком номер (нажимайте на клавишу с номером) наиболее правильного ответа!”

Посредством тестирования чаще других признаков проверяются знания, умения, навыки и представления. С точки зрения педагогических измерений полезно ввести два основных показателя качества знаний - уровень и структура знаний. Они оцениваются посредством регистрации

оценок, как за знание, так и за незнание всех требуемых компонентов проверяемого материала. Для объективизации этого процесса все компоненты должны быть одинаковы. Одинаковыми являются и правила выставления оценок испытуемым. Эти условия открывают дорогу для объективного сравнения индивидуальных структур знания и не знания.

Уровень 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.

Если тестовый балл ниже требуемого уровня (измеряемого критерия), то проявленные при этом знания, умения, навыки и представления указывают на докритериальный уровень подготовленности испытуемого. Этот уровень является самым распространенным и массовым. Для многих он оказывается вполне преодолемым по мере изучения ими наук и овладения мастерством. Однако некоторые, преимущественно в силу недостатка способностей и прилежания, так и остаются на этом уровне. В России на докритериальном уровне оказывается примерно 80 % выпускников школ. Если исходить из примерно 20% абитуриентов.

В истории науки и искусств известно немало случаев, когда имевшийся в соответствующие времена критериальный уровень требований к знаниям (или мастерству) мешал общественному признанию отдельных талантов. Поэтому признание нередко приходило только по мере изменения критериев оценки. Чаще всего после смерти авторов многих бессмертных произведений в сфере науки, музыки, живописи и др. Яркий пример такого рода - изменение критериев оценки творчества Ван Гога, творчества многих художников-импрессионистов, формалистов и др. Известны и такие, например, факты. Будущий лауреат Нобелевской премии Альберт Эйнштейн летом 1895 года не смог сдать экзамены в Цюрихский политехникум, а

всемирно известный пианист С. Рихтер дважды отчислялся из консерватории за общую неуспеваемость.

Каждое учебное заведение должно стремиться, в первую очередь, к формированию правильных индивидуальных структур знаний, в которых не было бы пробелов (разрывов в знаниях), и на этой основе повышать уровень подготовки. Уровень знаний в значительной степени зависит от личных усилий и способностей учащихся, 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.

Как справедливо отмечал М. Минский, человек не может хорошо учиться, если разрывы для него между известным и неизвестным слишком велики. Уровень знаний в значительной степени зависит от личных усилий и способностей учащихся. В то время как структура знаний зависит от правильной организации учебного процесса, от индивидуализации обучения, от мастерства педагога, от объективности контроля. В общем, от всего того, чего обычно у нас не хватает.

В тестовом задании внимание педагога привлекает, в первую очередь, содержание и форма. Содержание определяется как отображение фрагмента учебной дисциплины в тестовой форме, форма - как способ связи, упорядочения элементов задания. Содержание теста существует, сохраняется и передается в одной из четырех основных форм заданий. Вне тестовых форм ни тест, ни его содержание не существуют.

1.4. Теоретические основы тестирования

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 считалась буржуазной,

все цели которой считались “реакционными”. И хотя такие суждения 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 measurement). Поскольку общего названия на русском языке пока нет, эту науку можно назвать 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.

Педагогическая тестология призвана 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. Педагогическая тестология – это прикладная методическая теория научной педагогики. Ключевыми понятиями тестологии, как одна из методических теорий, являются измерение, тест, содержание и форма заданий, надежность и валидность результатов измерения. Кроме того, в тестологии используются такие понятия статистической науки, как выборочная и генеральная совокупность, средние показатели, вариация, корреляция, регрессия и др.

В теории любой науки, важную роль играют исходные понятия. Это понятие педагогическое (учебное) задание, которое 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. Формы нетестовых заданий здесь не рассматриваются. 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.

1.5. Виды тестов

1.5.1. Традиционные тесты

Существуют два основных вида тестов: традиционные и нетрадиционные. В статье “Начала теории” (УШ № 28, июль /99) уже упоминалось понятие “традиционный тест”. Там оно определялось как система заданий возрастающей трудности, специфической формы; система, позволяющая качественно и эффективно измерить уровень и оценить структуру подготовленности учащихся. Теперь пришло время рассмотреть это понятие глубже.

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.

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.

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 автору пока 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.

Например, тест по физике измеряет знания, умения, навыки и представления испытуемых в данной науке. Одна из трудностей такого измерения заключается в том, что физическое знание изрядно сопряжено с математическим. Поэтому в тесте по физике экспертно устанавливается уровень математических знаний, используемых при решении физических заданий. Превышение принятого уровня приводит к смещению результатов; по мере превышения последние все больше начинают зависеть не столько от знания физики, сколько от знания другой науки, математики. Другой важный аспект - стремление некоторых авторов включать в тесты не столько проверку знаний, сколько умение решать физические задачи, вовлекая, тем самым, интеллектуальный компонент в измерение физической подготовленности.

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.

1.5.2. Нетрадиционные тесты

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 в области 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), еще 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.

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. Отобранный таким образом, после согласования, состав экспертов в зарубежной литературе нередко называется 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.

Адаптивные тесты

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” (1). 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 5–10 минут. Адаптивное тестирование позволяет обеспечить компьютерную выдачу заданий на оптимальном, примерно 50%-ном 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 с заранее известными параметрами трудности и дифференцирующей способностью 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” (Киев, 1870 г.) можно прочитать 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”.

Критериально-ориентированные тесты

Это весьма условное и в принципе неправильное название группы тестов, получивших у нас некоторое распространение и признание. К сожалению, была сделана даже попытка ввести это название в текст наших законов об аттестации и о стандартах. По существу же мы имеем дело не столько с тестами названного рода, сколько с интерпретацией тестовых результатов

Если главной задачей является стремление выяснить, – какие элементы содержания учебной дисциплины усвоены тем или иным испытуемым, то это случай предметно-педагогического подхода к интерпретации результатов тестирования. При этом определяется: что из генеральной совокупности заданий (по англ. Domain) испытуемый знает и чего не знает. Интерпретация результатов ведется педагогами, на языке учебной дисциплины.

Вывод выстраивается вдоль логической цепочки: содержание учебной дисциплины □ генеральная совокупность задания для измерения знаний □ тест как выборка заданий из этой совокупности, ответы испытуемого □ вероятностный вывод о его знаниях учебной дисциплины. При ориентации на такие тесты требуется большое число заданий и достаточно полное

определение содержания изучаемой дисциплины. Интерпретация результатов ведется педагогами – предметниками.

Споры ведутся вокруг двух главных вопросов:

1) правильности содержания теста, что означает безошибочность формулировок его заданий, предметно-научную обоснованность, допустимость теста для проверки интересующих знаний в данной группе испытуемых. При аргументации в пользу того или иного теста педагоги-предметники опираются на понятийный аппарат, язык, принципы и вообще на знания преподаваемой ими учебной дисциплины. В таких случаях говорят о тестах с содержательно-ориентированной интерпретацией результатов. Это так называемый случай *Domain Referenced Testing*, что можно перевести как соотнесение знаний по результатам теста со знаниями, полный перечень которых представлен в генеральной совокупности (*domain*);

2) обоснованности оценки знаний по всему учебному предмету, на основе результатов тестирования испытуемых по небольшой выборке заданий теста; выборке из потенциально или реально существующей генеральной совокупности всех заданий, которые можно было бы дать испытуемым для уверенной и обоснованной оценки. Фактически это вопрос обоснования точности индуктивного вывода о знании большого числа вопросов на основе ответов по малому числу заданий теста.

Второй вид тестов связан с ориентацией на такие конкретные цели и задачи, как, например, проверка уровня усвоения сравнительно короткого перечня требуемых знаний, умений и навыков, выступающих в качестве заданного стандарта или критерия усвоения. Например, для аттестации выпускников образовательных учреждений важно иметь такие задания, которые позволяют делать вывод о минимально допустимой компетентности слушателей. За рубежом их так и называют: *Minimum Competency Tests*. При проверке минимально допустимого уровня знаний содержание заданий носит принципиально облегченный характер. Поскольку такие задания должны

выполнять все слушатели, допущенные учебным заведением к аттестации, здесь трудно говорить о тестах как методе объективного и эффективного измерения разного уровня подготовленности испытуемых, в строгом смысле понятия “тест”. Этот подход выработан для органов управления образованием, стоящих перед необходимостью в короткое время проверить состояние образования в большом количестве учебных заведений и не позволять последним опускаться ниже предельно допустимого уровня требований.

В западной литературе в таких случаях говорится о тестах с критериально-ориентированной интерпретацией результатов. Вывод выстраивается вдоль логической цепочки: задания □ ответы □ выводы о соответствии испытуемого заданному критерию. Под критериально-ориентированной интерпретацией имеется в виду сравнение содержания аттестационных материалов с результатами тестирования и вывод: что из заданного стандарта, в смысле требований, и на каком уровне, реально усвоено.

При критериально-ориентированной интерпретации требуется несколько меньшее число заданий, посредством которых определяется – что испытуемый знает и чего не знает из заданного стандарта. Иначе говоря, здесь ответы оцениваются относительно не всей области (Domain) требуемых знаний, а только из области, ограниченной конкретным стандартом или уровнем (критерием) знаний. Как и случае с Domain Referenced Testing, интерпретация результатов ведется на языке учебной дисциплины, но уже преимущественно работниками органов управления образованием и теми педагогами, на мнения которых управленцы опираются при аттестации.

Используемые при этом “тесты” не соответствуют подлинно тестовым требованиям, предъявляемым к традиционным и адаптивным тестам. При критериально-ориентированной интерпретации для диагностики заранее заданного уровня подготовленности используются, по существу, не тесты, в

традиционном понимании этого метода, а совокупности заданий в тестовой и в иной другой форме, не более того. Слово одно, но смысл другой.

Тесты с критериально-ориентированной интерпретацией нередко противопоставляются тестам с так называемой нормативно-ориентированной интерпретацией результатов. На самом деле, последние – это традиционные тесты, некоторые из которых имеют параллельные варианты.

1.6. Определение принципов разработки содержания теста для различных слоев потребителей

В предыдущем пункте содержание теста определялось как оптимальное отображение учебного материала в системе тестовых заданий. Такое отображение открывает возможность нетрадиционного построения учебного процесса, в котором тестовые материалы используются не только для контроля, но и для обучения. Например, применение заданий в компьютерных контрольно-обучающих программах позволяет учащимся самостоятельно обнаруживать пробелы в знаниях и принимать меры для их ликвидации. В таких случаях можно говорить об обучающем потенциале тестовых заданий.

При разработке теста каждый автор старается подобрать свою систему заданий, вследствие чего по одной и той же учебной дисциплине может оказаться несколько различных тестов, с неодинаковым охватом учебных тем, с несовпадающим числом заданий, и с разными баллами испытуемых. Возникает естественный вопрос, – какой из тестов лучше измеряет интересующие знания? Ответ на этот вопрос дают эксперты. Лучшим часто считается тест, у которого содержание шире и оно охватывает более глубокие уровни знаний. Такое содержание определяется с опорой на ряд педагогических принципов.

Первый принцип разработки содержания теста - соответствие содержания теста целям тестирования - уже был рассмотрен в предыдущей

статье. Остается добавить, что для оценки знаний слушателей образовательных учреждений на соответствие требованиям минимума содержания учебных программ, в средства оценки придется включать немало сравнительно легких заданий. Но тогда ни о каком тесте, в нормативном понимании этого слова, говорить не придется. Возникнет и ряд таких спорных вопросов, как определение минимально “проходного” балла (или процента правильного выполнения числа заданий, как было задумано в проекте закона об аттестации учащихся). По мнению автора, слова “минимальное содержание” (если они утвердятся в официальном лексиконе, вопреки тому, что такие слова вообще противоречат сути гуманитарно-ориентированной образовательной практики, декларируемой министерством общего и профессионального образования) можно будет соотнести с минимально положительной оценкой. Это будет оценка “три”. В то время как хорошие и отличные оценки должны выставляться за знания более широкие и глубокие, нежели минимально требуемые в образовательном стандарте.

Отсутствие четко сформулированных, общественно одобряемых целей, согласованных с методами и организационными формами их достижения, заводят тестирование в России не столько в кризисное состояние (после которого могли бы ещё остаться надежды на выздоровление), сколько в тупик, выход из которого бывает в одном направлении – назад, к исходной точке. И это уже было в истории тестов советского периода.

Второй принцип - определение значимости проверяемых знаний в общей системе проверяемых знаний. Принцип значимости указывает на необходимость включения в тест только тех элементов учебной программы, которые можно отнести к наиболее важным, без которых знания становятся несущественными, фрагментарными, состоящими из второстепенных элементов. Поскольку именно ключевые элементы формируют основу подлинных знаний, в тест в первую очередь включают, конечно, их. Отбор значимых элементов проводится с опорой на мнения педагогов-экспертов,

имеющих личный опыт преподавания и проверки знаний по интересующей учебной дисциплине.

Третий принцип - взаимосвязь содержания и формы. При разработке тестов может оказаться уместной некоторая ассоциация с созданием произведений искусства. Настоящий тест так же можно охарактеризовать как результат соединения содержания заданий с наиболее подходящей для них формой. Одновременно форма является способом существования и сохранения содержания задания. Вне тестовых форм ни тест, ни его содержание не существуют.

Начинающие разработчики тестов часто допускают ошибки несоответствия формы задания его содержанию. Главное средство предупреждения таких ошибок - овладение классификацией видов знаний, которая будет опубликована в ближайших выпусках газеты. Не всякое содержание поддается выражению в форме тестового задания. Доказательства, обширные вычисления, многословные описания не поддаются представлению в тестовой форме. Из этого вытекает важный вывод - тестовую форму нельзя рассматривать как универсальную, пригодную на все случаи проверки знаний. Да и вряд ли в педагогике есть вообще какая-либо одна форма, которая могла бы выразить все богатство изучаемых знаний. Вместе с тем, тестирование является одной из наиболее объективных и технологичных форм проведения массового контроля знаний. В этом смысле сравниться с тестированием ни один из известных методов проверки знаний не может.

Содержание контроля по каждой учебной дисциплине рассматривается сквозь призму подходящей формы. При этом анализ содержания теста относится к учебной дисциплине, в то время как поиск наилучших форм является предметом науки о тестах, т.е. тестологии. В этой науке имеется несколько довольно развитых и конкурирующих теорий. Сейчас стало ясно,

что нет абсолютно никаких оснований считать какую-либо одну теорию тестов современной, чем другие теории.

Четвертый принцип - содержательная правильность тестовых заданий. В тест включается только то содержание учебной дисциплины, которое является объективно истинным и поддается аргументации. Хотя в педагогической литературе последних лет справедливо обосновывается правомерность выделения двух сложившихся методов познания мира - рационального, естественно научного и интуитивно-образного (2), в содержание тестового контроля знаний наметилась традиция включать те элементы знаний, которые поддаются преимущественно рациональной, нежели интуитивно-образной аргументации. Каждое задание теста опирается, как правило, на факт, правило, теорему, норму, закон или на апробированный в практике метод. Спорные точки зрения, вполне приемлемые в науке, не рекомендуются включать в содержание тестовых заданий. В отличие от вопросно-ответной формы заданий, где даются ответы, правильные в различной степени, тестовые задания обычно требуют определенных ответов, признаваемых в качестве точных и бесспорных.

Пятый принцип - репрезентативность содержания учебной дисциплины в содержании теста. При разработке теста обращается внимание на полноту и достаточность числа заданий для аргументированного вывода о знаниях. В самом деле, проверить знания учащихся можно по пяти-шести заданиям. Но где уверенность, что учащиеся знают остальные элементы содержания учебной дисциплины? Путь к появлению такой уверенности лежит в наиболее полном отображении необходимого знания в заданиях теста. Репрезентативность не означает обязательного включения в тест всех значимых элементов содержания или строго пропорционального включения в тест заданий по различным темам. Ведь многие из них явно связаны между собой в общей структуре знаний, включены один в другой, полностью или частично. Кроме того, много элементов в структуре знаний иерархически

соподчинены, а потому из соображений экономии, нет смысла включать в тест задания из всех изучавшихся тем. Задания, которые выявляют знания таких элементов, можно назвать содержательно-структурными заданиями теста.

Стремление к повышению валидности тестовых результатов путем расширения числа тем учебной дисциплины и увеличения числа заданий в тесте нельзя, в принципе, признать правильным. Нет таких тестов, содержание которых вбирало бы в себя все содержание учебной дисциплины. При создании теста обычно ставится задача отобразить в нем то основное, что отражает идею измерения уровня и структуры подготовленности учащихся, как результат образовательной деятельности. Можно сказать так: содержание учебной дисциплины всегда шире содержания тестовых заданий. Некоторые из них заметно связаны между собой, иногда перекрывают друг друга по содержанию и потому могут замещаться, что позволяет уменьшить общее число заданий в тесте.

Репрезентативность заданий связана с числом заданий. В общем случае, чем больше заданий, тем репрезентативнее могут оказаться результаты. Число заданий традиционного теста обычно бывает не меньше тридцати. Банк заданий адаптивного теста содержит большее число заданий. Число заданий в тесте зависит, во-первых, от содержания проверяемого материала: чем больше объем проверяемых знаний, тем большее обычно требуется число заданий. Во-вторых, от вида тестов; интегративные тесты требуют меньшего числа заданий, в силу того, что для правильного решения каждого задания надо обладать знаниями различных учебных дисциплин. Следовательно, тест, состоящий из интегративных заданий, охватывает большее число проверяемых учебных элементов. И в-третьих, точность педагогических измерений зависит от числа заданий; в традиционном тесте точность измерения растет по мере увеличения числа заданий.

Шестой принцип - соответствие содержания теста уровню современного состояния науки. Этот принцип вытекает из естественной необходимости проверять знания школьников не на устаревшем, а на современном учебном и контрольном материале. Трудность реализации этого принципа заключается в опосредованности связи содержания теста с уровнем развития науки. Здесь полезно напомнить, что учебный предмет представляет своеобразную проекцию научного знания в плоскость усвоения.

Седьмой принцип - комплексность и сбалансированность содержания теста. Тест, разработанный для итогового контроля знаний, не может состоять из материалов только одной темы, даже если эта тема является самой ключевой в учебной дисциплине. Необходимо искать задания, комплексно отображающие основные, если не все, темы учебного курса. В то же время существует стремление сбалансировано отобразить в тесте основной теоретический материал - понятия, законы и закономерности, гипотезы, факты, вместе с методами научной и практической деятельности, с умениями эффективно решать типовые задания. Особенно важен этот принцип для разработки интегративных тестов, рассмотренных в статье “Нетрадиционные тесты”.

Восьмой принцип - системность содержания. Это означает формулирование такого содержания тестовых заданий, которое отвечало бы требованиям системности при проверке знаний. Помимо подбора заданий с системным содержанием важно иметь задания, связанные между собой общей структурой знаний. Это возможно в тех случаях, когда каждое задание проверяет преимущественно свою часть в общей системе знаний. Идеальный вариант при разработке теста – когда задания не пересекаются по содержанию и по статистическим показателям между собой, но все имеют положительные корреляции с каким-либо внешним критерием знаний испытуемых.

Девятый принцип - вариативность содержания. После первого применения теста его содержание становится известным испытуемым. И если есть условия для передачи информации о содержании заданий другим учащимся, то это почти всегда делается. Испытанным в практике способом защиты тестовых результатов от возможных искажений такого рода является создание множества вариантов заданий одного и того же теста. Соответственно тесты, состоящие из вариантов заданий, называются параллельными, если при этом выполняются еще и некоторые статистические условия. Например, они должны быть примерно равны по трудности, иметь сходные показатели вариации тестовых баллов испытуемых. Кроме того, содержание теста не может оставаться неизменным, независимым от нового содержания учебной дисциплины и от новых учебников.

Обобщенная матрица спецификаций теста приведена на рис. 1.4.

Пример распределения тестовых заданий по теме: «Базовые понятия языка Pascal» приведен на рис 1.5. **(желательно заменить на раздел информатики)**

Элементы содержания.	Объекты (цели контроля)					
	<i>Распознавать основные термины - 30%</i>	<i>Идентифицировать факты - 20%</i>	<i>Идентиф. Принципы, понятия, законы, закономер др. – 30%</i>	<i>Оценить информацию – 10%</i>	<i>Прим-ть принципы и обобщ-я к новым услов. - 10%</i>	<i>Кол-во заданий в %</i>
Разд. 1	<i>Список терминов</i>	<i>Список фактов</i>	<i>Переч-ние принцип., понятий, законов, основных связей, др</i>	<i>Анализ информац. Решение стандарт. Задач, выявить ошибки</i>	<i>Определ. желаем. пропорции. Предвидеть последствия, др.</i>	<i>- ?%</i>
Разд. 2						<i>- ?%</i>

Рис. 1.4. Предлагаемая матрица спецификаций теста

Рис. 1.5. Предлагаемая матрица теста для дисциплины Информатика

1.7. Результаты исследования подходов к определению трудности теста. Определение форм тестовых заданий для различных слоев населения.

Если педагогический тест определить кратко как систему заданий возрастающей трудности, то станет понятно, что трудность заданий является важнейшим, скажем так, тестобразующим показателем. Немало руководителей образовательных учреждений считают, что их учителя в состоянии “придумать” за короткое время могут сколько угодно “тестов” На самом же деле можно придумать сколько угодно заданий в тестовой форме (а это ещё не тесты). Их нельзя включать в настоящий тест до тех пор, пока не

станет известной мера трудности, проверяемая опытным путем. Из этого требования становится понятной обязательность предварительной эмпирической проверки каждого задания, до начала тестирования. В процессе проверки многие задания (обычно больше половины) не выдерживают предъявляемых к ним требований и потому не включаются в тест. Первое требование к тестовым заданиям: в тесте задания должны различаться по уровню трудности, что вытекает из данного ранее определения теста и рассматриваемого принципа.

Имеются различия в лексике трех как бы “незаметно” введенных здесь основных понятий теории тестов: педагогического теста, задания в тестовой форме и тестового задания.

Требования ко второму понятию лучше ввести сейчас, сделав это хотя бы кратко перечислив их, для того чтобы не отвлекаться от основной темы статьи. К заданиям в тестовой форме предъявляются следующие требования:

- правильность содержания
- логическая форма высказывания;
- правильность формы;
- краткость;
- наличие определенного места для ответов;
- правильность расположения элементов задания;
- одинаковость правил оценки ответов;
- одинаковость инструкции для всех испытуемых;
- адекватность инструкции форме и содержанию задания.

Подробное истолкование этих требований последует позже, а сейчас хотелось бы обратить внимание на то, что здесь нет требования известной трудности задания, в то время как к тесту и к тестовому заданию такое требование предъявляется. Из размышления над этим и ранее опубликованным материалом можно сделать два вывода:

- *Первый - что в тесте нет места заданиям с неизвестной мерой трудности.*
- *И второй – что не все предлагаемые задания в тестовой форме могут стать тестовыми заданиями: это разные понятия.*

В первом понятии к самым существенным относятся требования формы и содержания. К тестовым же заданиям в первую очередь предъявляется требование известной трудности, то, что явно не требуется у заданий в тестовой форме. Задания имеют шанс стать тестовыми только после опытной, скажем строже, эмпирической проверки меры их трудности, на типичных группах испытуемых.

Показатель трудности теста и тестовых заданий является содержательным и формальным одновременно. Содержательным показателем, потому что в хорошем тесте трудность может зависеть только от содержания и от уровня подготовленности самих испытуемых, в то время как в плохом тесте на результаты начинают заметно влиять форма заданий (особенно если она не адекватна содержанию), плохая организация тестирования, если имеются возможности списывания, утечки информации. Особого упоминания в этой связи заслуживает спорная практика нацеленной подготовки к централизованному тестированию.

Формальная составляющая сторона показателя трудности возникает при рассмотрении тестирования как процесса противоборства каждого испытуемого с каждым предлагаемым ему заданием. Получаемый при этом исход полезно рассматривать как результат такого противоборства. При упрощенном истолковании каждого случая противоборства испытуемого с очередным заданием обычно рассматриваются только два исхода: победа испытуемого при правильном решении задания, где он получает один балл, или поражение, за что даётся ноль баллов. Оценка результата противоборства зависит от соотношения уровня знания тестируемого к уровню трудности

задания, от избранной единицы измерения знаний и от заранее принятого правила (конвенции) - что считать "победой" испытуемого и допустима ли ничья, если говорить языком спорта.

Принцип возрастающей трудности используется при изложении содержания многих учебников и пособий, особенно по тем учебным дисциплинам, которые построены по кумулятивному принципу, что означает: знание последующих элементов курса в явном виде зависит от знания предыдущих учебных элементов. Такое построение присуще учебникам по математике, логике, иностранным языкам, статистике, техническим и многим другим наукам. В них ранее изученные понятия активно используются в последующих темах. Поэтому изучать такие дисциплины нужно только с самого начала, и без пробелов.

Большинство авторов, особенно зарубежных, не делают различий между понятиями "трудность" и "сложность". Многие разработчики тестов - то же. Однако есть работы, в которых эти понятия определяются различно. Например, А.Н.Захаров и А.М.Матюшкин отмечают, что степень трудности учебного задания не совпадает с его сложностью. Степень сложности учебного материала характеризуется реальной (объективной) насыщенностью учебного задания и формой его изложения, а степень трудности всегда предполагает соотнесение подлежащего усвоению учебного материала с ранее усвоенным учебным материалом и интеллектуальными возможностями учащихся.

Л.Н. Ланда объясняет трудность учебной задачи тем, что учащиеся часто не знают тех операций, которые надо производить, чтобы найти решение. Если систему операций для решения некоторого класса задач назвать методом решения, то по его мнению, трудность связана с незнанием метода, с незнанием, как нужно думать в процессе решения, как и в какой последовательности надо действовать с условиями задачи. Возникающие затруднения объясняются тем, что педагог часто старается дать знания о

содержании изучаемого и значительно меньше заботится о том, как надо думать, рассуждать (там же). Подобное истолкование пересекается с идеей о связи сложности задания с числом операций, которые необходимо совершить для достижения успеха. Эти определения трудности и сложности являются, по большей части, психологическими; они полезны при психологическом анализе содержания тестовых заданий.

Традиционной мерой трудности каждого задания долгие годы была доля правильных ответов в группе испытуемых, изображаемая символом p_j где индекс j указывает на номер интересующего задания (1, 2 и т. д.). Например, если правильные ответы испытуемых на третье задание теста оценивать одним баллом, а неправильные - нулем, то значение показателя p_3 можно найти из элементарного отношения

$$p_3 = R_3 / N ,$$

где R_3 означает число правильных ответов на данное задание, а N – общее число испытуемых в группе. Общая формула расчета доли правильных ответов на любое задание (j) имеет соответственно вид

$$p_j = R_j / N (1).$$

Показатель p_j долго использовался в качестве меры трудности в так называемой классической теории тестов (3). Позже была осознана содержащаяся в ней смысловая неточность: ведь увеличение значения p_j указывает не на возрастание трудности, а наоборот, на возрастание легкости, если можно использовать такое слово. Поэтому в последние годы с показателем трудности заданий стали ассоциировать противоположную статистику - долю неправильных ответов (q_j). Эта доля вычисляется из отношения числа неправильных ответов (W_j – от англ слова Wrong - неправильный) к числу испытуемых (N):

$$q_j = W_j / N (2)$$

Естественным образом принимается, что $p_j + q_j = 1$. В классической теории тестов многие годы рассматривались только эмпирические

показатели трудности. В новых вариантах психологических и педагогических теорий тестов больше внимание стало уделяться характеру умственной деятельности учащихся в процессе выполнения тестовых заданий различных форм (4).

Содержание теста не может быть только легким, средним или трудным. Здесь в полной мере проявляется известная мысль о зависимости результатов применяемого метода. Легкие задания теста создают только видимость наличия знаний у учащихся, потому что ими проверяются минимальные знания. В этой связи можно заметить, что ориентация федерального органа управления образованием на проверку минимального уровня знаний не дает, и не может, даже по определению, дать представление о реальном уровне знаний, т.е. дать ту информацию, которая давно уже нужна обществу и органам управления. Искажает результаты тестирования и подбор заведомо трудных заданий, в результате чего у большинства школьников оказываются заниженные баллы. Ориентация на трудные задания нередко рассматривается как средство усиления мотивации к учебе. Однако это средство действует неоднозначно. Одних трудные задания могут подтолкнуть к учебе, других - оттолкнуть от нее. Подобная ориентация искажает результаты и в итоге, снижает качество педагогического измерения. Если тест построен строго из заданий возрастающей трудности, то этим открывается путь к созданию одной из самых интересных шкал измерения - шкалы Л. Гутмана.

При определении теста уже отмечалось, что все задания теста, хотелось бы подчеркнуть, независимо от содержания тем, разделов и от учебных дисциплин, располагаются в порядке возрастающей трудности. Распространенная, до недавнего времени, рекомендация включать в тест больше заданий средней трудности, оправдана с точки зрения определения надежности измерения по формулам т.н. классической теории тестов. Существующие в этой теории методы оценки надежности теста дают снижение надежности при включении в тест заметно легких и трудных

заданий. В то же время увлечение заданиями одной только средней трудности приводит к серьезной деформации содержания теста: последний теряет способность нормально отображать содержание изучаемой дисциплины, в которой всегда есть легкий и трудный материал. Таким образом, в погоне за теоретически высокой надежностью теряется содержательная валидность теста. Стремление же поднять валидность теста нередко сопровождается снижением его надежности.

Если тестируется слабая, по подготовленности, группа учащихся, то оказывается, что трудные задания теста просто не работают, потому что ни один учащийся не может правильно на них ответить. Такие задания из дальнейшей обработки данных изымаются. В адаптивных контролирующих системах они не предлагаются. Содержание теста для слабых учащихся будет заметно отличаться от содержания теста для сильных учащихся. У последних, наоборот, не работают легкие задания, так как все знающие испытуемые на легкие задания отвечают правильно. Таким образом, содержание традиционного теста существенным образом варьирует в зависимости от уровня подготовленности тех групп учащихся, на измерение знаний которых нацелен тест.

Оптимальное отображение содержания учебного материала в тестовые задания требуемого уровня трудности предполагает возможность выбора подходящей формы. Содержание теста выражается в одной из четырех основных форм заданий. Это: 1) задания с выбором одного или нескольких правильных ответов из числа предложенных; 2) задания открытой формы, где ответ испытуемый дописывает сам, в отведенном для этого месте; 3) задания на установление соответствия, и 4) задания на установление правильной последовательности действий.

1.8. Разработка логических требований к содержанию теста

Тестовые задания отличаются от нетестовых не только по содержанию, но и по стилю построения предложения. Из тестового задания полностью устраняется двусмысленность, когда, например, подлежащее в именительном падеже легко путается с прямым дополнением в винительном падеже (типа “мать любит дочь”). Кроме того, иногда тестовое задание ошибочно отождествляется с загадкой. Хотя в обоих случаях ставится задача найти правильный ответ, отмеченного сходства недостаточно для отождествления загадки и тестового задания. Есть признаки, которые явно отделяют их. Для загадки наиболее важным, специфически отличительным признаком является метафоричность, образность загадки. Известно, что метафоричным выражение становится тогда, когда оно употребляется не в прямом, а в переносном значении. Тестовое задание, напротив, всегда автологично: в нем слова используются только в их прямом, непосредственном значении. Автологический стиль - это тот, о котором В.Маяковский писал: "Ищем речи точной и нагой". Тестовое задание формулируется из точных терминов и никогда не содержит метафоры, но лишнего слова и лишнего знака. Пример метафорично сформулированного задания в открытой тестовой форме: “Под сигналом в информатике понимается_____”.

Логические принципы определения содержания педагогического теста. В логике известны такие законы правильного мышления как законы тождества, непротиворечия, исключенного третьего и достаточного основания. Это позволяет выделить такие основные свойства правильного мышления, как определенность, непротиворечивость, обоснованность. Применительно к тестовой теории и практики эти общие свойства правильного мышления приобретают функции специфических регулятивов тестовой деятельности, вследствие чего приобретают значение принципов. Рассмотрим их подробнее.

1.8.1. Определенность содержания теста

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.

Например, в физических расчетах используется немало математических знаний и потому в систему физического знания обычно включается та математика, которая используется при решении физических задач. Неудача в математических расчетах порождает неудачу при ответах на задания физического теста. Отрицательный балл ставится, соответственно, за незнание физики, хотя испытуемый допустил ошибки математического толка. Если в такой тест включено много таких заданий, которые для правильного решения требуют не столько физических знаний, сколько умений выполнять усложненные расчеты, то это может быть примером неточно определенного содержания теста по физике. Чем меньше пересечение знаний одной учебной дисциплины со знаниями другой, тем определеннее выражается в тесте содержание учебной дисциплины. Определенность содержания требуется и во всех других тестах. В гетерогенном тесте это достигается посредством явного выделения заданий одной учебной дисциплины в отдельную шкалу. При этом нередко встречаются задания, хорошо работающие не только на одну, но и на две, три и даже большее число шкал.

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. Рассмотрим пример задания открытой формы, где испытуемому предлагается, на месте прочерка, дополнить утверждение своим ответом:

ТОЧКА ЕСТЬ ТО, ЧТО НЕ ИМЕЕТ _____ . В сознании возникает вопрос: "Не имеет чего? Ответ, по Евклиду – "точка не имеет частей", но это как раз случай неудачного определения содержания, как задания, так и самого понятия точки. Как известно, точка не имеет, например, цвета, вкуса, запаха и многого другого, и не только частей.

Локальные вычислительные сети не могут быть объединены с помощью _____ . Как известно, ЛВС не могут быть объединены очень многими вещами.

Немалую роль в формировании теста как системы играет требование логической правильности заданий, включенных в тест. Как отмечается в литературе, логическая правильность в формулировании тестовых заданий достигается при соблюдении следующих условий:

-соразмерности объема определяющего понятия объему определяемого. Известные в литературе примеры:

1. ЧЕТЫРЕХУГОЛЬНИК, У КОТОРОГО ВСЕ СТОРОНЫ РАВНЫ, НАЗЫВАЕТСЯ _____ (ответ - ромб).

2. ПРЯМОУГОЛЬНИК, У КОТОРОГО ВСЕ СТОРОНЫ РАВНЫ, НАЗЫВАЕТСЯ _____ (ответ - квадрат). В случае если на первое задание дается ответ "квадрат", допускается несоразмерность: объем определяющего - четырехугольника - больше объема определяемого - квадрата;.

- отсутствия тавтологии. Примеры: СТОИМОСТЬ ТОВАРА ОПРЕДЕЛЯЕТСЯ СТОИМОСТЬЮ _____ (ответ - труда). СТОИМОСТЬ ТРУДА ОПРЕДЕЛЯЕТСЯ СТОИМОСТЬЮ _____ (ответ - товара). Как уже отмечалось, в этих примерах

обнаруживает себя нарушение правил логики; одна стоимость определяется через другую, которая в свою очередь нуждается в определении (2);

- утвердительной формы тестового задания. Примеры:

Под сигналом в информатике понимается:

:- информационный поток, распространяемый в канале связи

:+ форма представления информации, предназначенная для передачи

по каналам связи

:- процесс, выполняющий обработку данных для нужд пользователя

:- электрический разряд в цепи

Информацию, существенную и важную в настоящий момент, называют:

:- коммуникативной

:+ актуальной

:- достоверной

:- объективной

В приведенных примерах ответы подобраны по определенному основанию; в них соответственно перечислены лады и элементы росписи. Отсутствие общего основания приводит к логической противоречивости содержания задания и ответов. Например, в задании:

Информатика и программирование- это...

o неравнозначные понятия

o равнозначные понятия

o непересекающиеся понятия

o умение пользоваться программным обеспечением

Четвёртый ответ не соответствует содержанию остальных ответов и потому его можно признать некорректным.

1.8.2. Непротиворечивость содержания заданий

Непротиворечивость содержания заданий требует, чтобы относительно одной и той же мысли не возникали суждения, одновременно утверждающие

и отрицающие ее. Недопустимо существование двух исключаящих 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. Такие 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 измерений (но об этом в статьях, где будут рассматриваться вопросы надежности тестов).

1.8.3. Обоснованность

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.

В дополнение к трем перечисленным свойствам логически правильного мышления отмечу еще одно требование к тестовым заданиям, которое является столь же формальным, сколь и содержательным. Это требование краткости тестовых заданий. Краткость обеспечивается тщательным подбором слов, символов, графиков, позволяющих минимумом средств добиваться максимума ясности смыслового содержания задания. Исключаются повторы, малопонятные, редко употребляемые, а также неизвестные для слушателей символы, иностранные слова, затрудняющие

восприятие смысла. Примерное количество слов в задании - пять-девять, но, в общем, чем меньше, тем лучше. Например:

Программа The Vat позволяет...

- o загружать веб-страницы
- o загружать и редактировать электронную почту
- o архивировать электронную почту

Хорошо, когда задания содержат не более одного придаточного предложения. Как отмечал известный лингвист А.М.Пешковский, точность и легкость понимания растут по мере уменьшения словесного состава фразы и увеличения ее бессловесной подпочвы. Чем меньше слов, тем меньше недоразумений.

Хороший способ достижения краткости задания - это спросить о чем-нибудь одном. Часто встречающиеся случаи утяжеления заданий требованиями что-то найти, решить и затем еще и объяснить отрицательно сказываются на качестве задания и теста в целом, хотя с педагогической точки зрения легко понять причину формулирования такого рода заданий. Если в таких заданиях ученик что-то решает без учителя, потом что-то ему объясняет, то возникает сопряжение объективного метода с субъективным, плюс к этому - затруднения с оценками при решении вопроса какой выставить балл. В противоположность этому, одно из важных требований теста – иметь заранее разработанные правила выставления баллов без участия учителя.

1.9. Композиция заданий в тестовой форме

Понятие "композиция" означает произведение, структуру, состав, соединение – все это как результат творческого процесса, в котором создатель стремится к некоторому совершенству. При этом самыми существенными факторами успеха оказываются воля, оригинальность идеи, мастерство исполнения, терпение. Плюс бесконечное внимание к так

называемым мелочам, без которых талантливое произведение не создать; как говорил Микельанжело Буанаротти, от мелочей зависит совершенство, а совершенство – это уже не мелочь. Композиция заданий в тестовой форме образует такое структурное соединение элементов, которое позволяет выразить содержание и форму каждого задания в гармоничной целостности. Именно форма и содержание являются главными компонентами процесса создания тестовых заданий. Этот процесс можно также называть разработкой, написанием, подготовкой или какими то другими словами. В западной литературе, кроме того, нередко используется понятие “конструирование”. Учитывая преимущественно техническую сферу употребления данного слова, автор этой публикации склоняется к использованию понятия “композиция заданий в тестовой форме”, вкладывая в него идею лучшего, на текущий момент, соединения содержания заданий с наиболее подходящей формой. Это и есть самый существенный признак профессионально созданных заданий. Цель композиции - создание таких заданий, которые можно было бы включить в тест и использовать как в традиционно организованном учебном процессе, так и в автоматизированных системах контроля знаний.

В процессе композиции к заданиям в тестовой форме предъявляются следующие требования:

- логическая форма высказывания;
- правильность формы;
- краткость;
- наличие определенного места для ответов;
- правильность расположения элементов задания;
- одинаковость правил оценки ответов;
- одинаковость инструкции для всех испытуемых;
- адекватность инструкции форме и содержанию задания.

Логическая форма высказывания является универсальной формой четкого выражения мысли человеком, способным проявить знания по изучаемой учебной дисциплине. Эта форма во многих случаях заменяет вопросы. Учебные вопросы многословны и порождают ответы, полные и неполные, правильные и неправильные, разные по форме, содержанию и по структуре, вследствие чего оценка таких ответов требует обязательного участия преподавателя и сопровождается некоторой долей субъективизма. Вопросы и ответы на них иногда бывают столь неопределенными и многословными, что для выявления их истинности требуются большие затраты интеллектуальной энергии, в то время как технологичная методика тестирования предполагает четкую и быструю дифференцируемость ответов. В этом смысле традиционные вопросы и ответы не технологичны; их не рекомендуется включать в тест. То же относится и к некоторым задачам, имеющим тяжеловесные формулировки.

Логическое преимущество задания в тестовой форме заключается в возможности его естественного превращения, после ответа слушателя, в форму истинного или ложного высказывания. Переход к высказываниям позволяет создавать задания по-новому, опираясь при этом на ряд методических принципов, рассматриваемых далее.

Правильная форма заданий - это средство упорядочения и эффективной организации содержания теста. Организация предполагает анализ содержания учебной дисциплины, классификацию учебного материала, установление меж тематических и меж предметных связей, укрупнение дидактических единиц, представление этих единиц через элементы композиции заданий. Форма заданий правильная, если она позволяет точно выразить содержание, понятна для всех испытуемых, исключает возможность появления ошибочных ответов по формальным признакам.

Технологическое преимущество заданий тестовой формы проявляется в их соответствии требованиям автоматизации рутинных компонентов

обучения и контроля знаний. Если в каждом задании автоматически ставить, в зависимости от ответа, самый простой вариант оценки, 1 или 0, то легко видеть преимущества, вытекающие из применения такой оценки, основанной на двужначной логике. Это позволяет быстро регистрировать ответы и объективно их оценивать по заранее разработанным правилам, применяемым ко всем одинаково, без исключений. Задания в тестовой форме легко вводятся в компьютер, компактно проецируются на экран монитора, хорошо различаются по форме и смыслу.

Семантическое преимущество заданий заключается в лучшем понимании их смысла и значения. Это связано, во-первых, со словесным составом задания в тестовой форме: смысл тестового утверждения улавливается всегда лучше, чем смысл вопроса. В тестовых утверждениях нет ни одного лишнего слова и даже знака, в то время как вопрос требует ряда дополнительных слов и знаков для выражения требуемого смысла, значения и интонации.

Если к заданиям даются готовые ответы на выбор (обычно один правильный и остальные неправильные), то такие задания лучше называть заданиями с выбором одного правильного ответа. Это первая, по счету, форма. Логической основой этой формы является закон исключенного третьего, впервые сформулированный Аристотелем. Выбор правильного ответа дает истинное суждение, а выбор неправильного - ложное суждение. Третьего не дано. Из этого закона следует методическое правило: в каждом задании с выбором одного правильного ответа последний должен быть, что придает однозначность замыслу самого задания и не допускает противоречивых толкований у испытуемых.

Другой вариант заданий этой же, первой формы, где имеется уже не один, а несколько правильных ответов. Это задания с выбором нескольких правильных ответов.

И еще один, третий вариант заданий этой же формы используется для проверки сопоставительных знаний- с выбором одного, наиболее правильного ответа из некоторого числа ответов, правильных в разной степени. Каждый вариант заданий первой формы можно разделить на такие виды, как задания с двумя, тремя, четырьмя, пятью и более ответами.

Следование закону исключенного третьего налагает логический запрет на применение таких ответов, как “правильного ответа нет”, “все ответы правильные” или “все ответы неправильные”, которые в практике все еще встречаются. В настоящей работе этот запрет соблюдается.

Во второй форме задание сформулировано так, что готового ответа нет; каждому испытуемому во время тестирования ответ приходится вписывать самому, в отведенном для этого месте. Такие задания можно назвать заданиями открытой формы. После дополнения задания определенным ответом получается истинное или ложное высказывание.

Задания, где элементам одного множества требуется поставить в соответствие элементы другого множества, можно назвать заданиями на установление соответствия. Это третья форма. В настоящем издании появились новые варианты заданий с большим числом множеств сопоставляемых элементов.

И, наконец, в тех случаях, когда требуется установить правильную последовательность вычислений, действий, шагов, операций, терминов в определениях используются задания на установление правильной последовательности. Это четвертая форма заданий. Каждая из перечисленных форм позволяет проверить специфические виды знаний, а также соответствующие им контрольные материалы. Выбор форм зависит от цели тестирования и содержания теста, от технических возможностей и уровня подготовленности преподавателей в области теории и методики тестового контроля знаний.

Основные требования к композиции

Краткость заданий в тестовой форме вытекает из преимуществ логической формы высказывания, Поэтому задания всегда короче задач и вопросов. Краткость обеспечивается тщательным подбором слов, символов, графиков, позволяющих минимумом средств добиваться ясности смысла содержания задания. Исключаются повторы, малопонятные, редко употребляемые, а также неизвестные для учащихся символы, иностранные слова, затрудняющие восприятие смысла. Хорошо, когда задания содержат не более одного придаточного предложения. Для достижения краткости задания лучше спрашивать о чем-нибудь одном.

Наличие определенного места для ответов является одним из внешних признаков задания в тестовой форме. В заданиях с выбором ответов - это код (цифры или буквы) требуемого ответа, из числа прилагаемых к каждому заданию. В заданиях открытой формы ответ пишется вместо прочерка, следуемого за текстом задания. В заданиях на установлении соответствия ответы или пишутся в специально отведенной для этого строке ниже текста каждого задания, или фиксируется с помощью так называемой "мыши" при компьютерном тестировании. И, наконец, в заданиях на установлении правильной последовательности испытуемый ставит ранги в специально отведенном для этого месте.

Правильность расположения элементов задания является требованием, помогающим испытуемым не тратить время на определение места для ответов и быстрее зафиксировать свое решение.

Одинаковость правил оценки ответов в рамках принятой формы является важным средством организации тестирования. Ни одному испытуемому не дается никаких преимуществ перед другим, все отвечают на одни и те же задания, всем дается одинаковое время. Правила оценки определяются заранее и абсолютно одинаково применяются ко всем испытуемым. Понятно, что все это делается для уменьшения ошибок измерения, вызванных нарушениями стандартных условий тестирования. В

смысле стандартизации условия тестирования напоминают требования к проведению эксперимента. Именно об этом писал один из первых создателей тестового метода (2).

Ответ на задание представляет собой краткое суждение, связанное по содержанию и по форме с содержанием задания. Критерии правильности ответов заранее определяются автором задания. Оценка ответов по степени их правильности в практике проводится редко, но при необходимости создаются задания с такими ответами, которые правильны в различной степени (3).

Адекватность инструкции форме и содержанию задания означает взаимное соответствие перечисленных компонентов, необходимое для выполнения задания своей функции. Условие адекватности позволяет довести до сознания испытуемых все требования, заложенные в содержании задания. Несоответствие вызывает ошибку понимания смысла задания, и соответственно, приводит к ошибочным ответам там, где таковым не место.

Полезно дать еще одно определение задания в тестовой форме: это варьирующая по элементам содержания и по трудности единица контрольного материала, сформулированная в утвердительной форме предложения с неизвестным. Подстановка правильного ответа вместо неизвестного компонента превращает задание в истинное высказывание; подстановка неправильного ответа приводит к образованию ложного высказывания, что свидетельствует о незнании слушателем данного учебного материала.

Овладение формой является необходимым, но недостаточным условием создания полноценных тестов. Форма придает заданиям лишь структурную целостность и определенность, внешнюю организованность. Задания в тестовой форме только внешне похожи на тестовые задания, а это недостаточно для их включения в тест; нужна еще проверка так называемых тестообразующих свойств заданий.

Обобщенные требования к составлению тестовых заданий приведены на рис. 1.6.

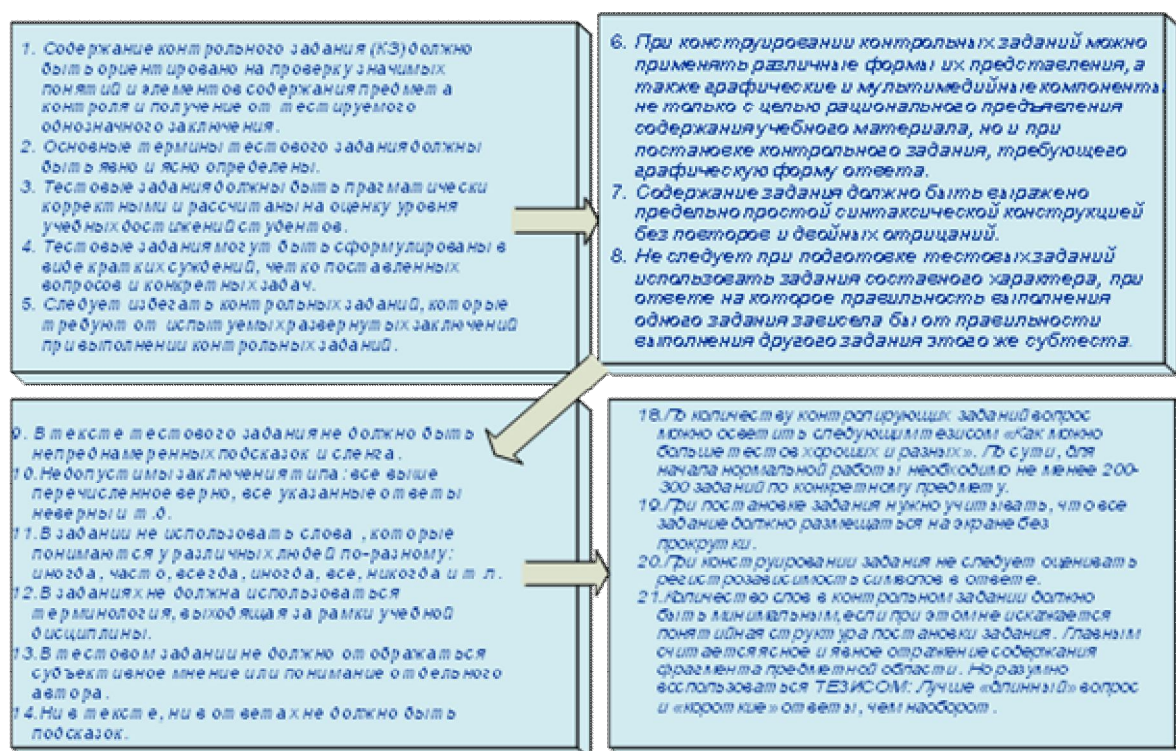


Рис. 1.6. Обобщенные требования к составлению тестовых заданий

2. Логико-информационная структура АПИМ.

Общепринятые требования к качеству АПИМ (в отношении достоверности, согласованности, полноты и т. п.), а также условия их использования в БД, сетях и вычислительных приложениях, определяют целую совокупность характеристик, отражающих структуру этого типа данных.

Фонд АПИМ включает как первичные (сырые) данные, полученные непосредственно от разработчиков тестовых заданий, так и вторичные, прошедшие качественную экспертизу и статистическую обработку, а также и данные полученные по результатам тестирования слушателей в образовательных учреждениях которые можно считать результатами эксперимента. При этом полученные в результате обработки вторичные (рекомендованные) данные по объему и структуре могут заметно отличаться от первичных экспериментальных данных.

Для целей хранения и распространения применяются многообразные формы представления данных. Первичные данные задаются, как правило, в табличной, реже, в графической форме; вторичные – в табличной форме, в виде аналитических функций (например, уравнений состояния) или программных кодов.

Как первичные, так и вторичные данные по свойствам объектов имеют некоторую логическую структуру, строго соответствующую модели для объектов данного класса (в простейшем случае структура может быть представлена в виде таблицы). Однако, для широкого класса справочных данных наблюдаются вариации логической структуры применительно к отдельным наборам данных, например, применительно к разным объектам или разным свойствам. Такая особенность данных позволяет отнести их к, так называемым, слабоструктурированным (или полуструктурированным) данным, в соответствии с типологией, принятой в теории БД. Изменчивость логической структуры данных означает, что типовая форма в виде перечня

атрибутов (свойств), приписываемых объекту (веществу, материалу) может заметно меняться в зависимости от класса объекта, его изученности, диапазона параметров и т. д. Специалистам по БД эта проблема знакома на примере временных рядов, когда приходится отслеживать изменения не только показателей, но и классификаторов, номенклатуры и т. п., с тем, чтобы обеспечить совместное использование данных с изменяющейся во времени структурой.

4. Данные по свойствам в скрытом виде содержат «знания» в виде математической модели, определяющей зависимость свойств от параметров состояния, фазы, вещества (или класса веществ), и «в идеале» оценки достоверности модели. Выделенные в явном виде, эти знания позволяют: (1) провести верификацию данных на предмет их достоверности и согласованности; (2) выполнить экстраполяцию и интерполяцию по параметрам; (3) расширить номенклатуру свойств, доступных для пользователя; (4) построить (или выбрать из predetermined множества) математическую модель, то есть, функциональную зависимость свойства (или комплекса свойств) от параметров состояния. Особое значение в технологиях подготовки АПИМ имеет верификация согласованности данных. Так называемая, совместная обработка данных (*multiproperty or simultaneous analysis*), когда модель восстанавливается по данным различной природы, что дает одновременный контроль достоверности всей совокупности данных и наиболее надежную оценку параметров модели, стала в настоящее время одной из основных технологий при подготовке АПИМ. Анализ согласованности данных включает критический анализ экспериментальных методов, полноты учета погрешностей и т. п.

Заключенные в данных знания существенно меняют организацию БД или вычислительной системы, используемой как потребителями, так и экспертами, отвечающими за наполнение фондов. Возможность эффективной свертки исходной информации позволяет сократить объем и унифицировать

структуру вторичных данных, возложив решение пользовательских задач на приложения.

Активное использование знаний при подготовке АПИМ предполагает работу эксперта с целым множеством моделей, заранее выбранных на основе определенных представлений. Множественность моделей (структурная оптимизация), во-первых, усиливает значимость данных, структура которых может меняться от одного экземпляра или набора данных к другому (плохо структурированные данные – см. выше). С другой стороны, в задачу БД входит теперь манипуляция не только данными, но и знаниями, то есть подбор моделей, адекватных поставленной задаче: типу объекта, свойства, диапазону параметров и проч. От БД, как основного элемента информационной технологии, происходит постепенный переход к базе знаний, способной кодировать накопленные знания о предметной области и продуцировать новые знания.

Сущность знаний, заключенных в первичных данных, означает также, что при выборе определенного свойства для определенного объекта эти данные должны быть согласованы как с известными закономерностями, так и с другими данными, например, по другим свойствам того же объекта, или по тем же свойствам, но других объектов.

Предлагаемая модель АПИМ приведена на рис. 2.1.

Модель аттестационного педагогического измерительного материала по дисциплине



Рис. 2.1. Предлагаемая модель АПИМ

2.1. Требования, связанные с динамическим характером АПИМ.

Как уже отмечалось, динамический характер справочных данных связан с реальной практикой их перманентного уточнения и пересмотра, вызванного расширением массива исходных данных, апробацией новых моделей (например, уравнений состояния), методик обработки и согласования и т. п. Отсюда следует целесообразность применения в работах по АПИМ, по крайней мере, двух новых (но уже опробованных в сфере бизнеса) информационных технологий: (1) электронные публикации в виде динамических документов (используются также термины виртуальный или динамически компоуемый документ); (2) электронный документооборот, система, обеспечивающая автоматизацию и протоколирование всех рабочих стадий, повторяемых экспертами по мере появления новых экспериментальных данных или привлечения новых математических моделей, методов обработки и проч.

Сам по себе переход к электронным документам (размещаемым на CD или на сервере), как основному средству публикации данных, имеет ряд преимуществ, в сравнении с традиционной формой печатного издания: простота издательской технологии, включение в документ графических и вычислительных модулей, интеграция с БД или электронными таблицами, отсутствие ограничений на объем текстового и табличного материала. Поставка электронного документа как источника справочной информации открывает ряд уникальных возможностей, например:

- появляется возможность непрерывного обеспечения пользователей новыми версиями тестов, появляющихся при включении в обработку новых данных;
- появляется возможность представления системы справочных данных в виде динамических сетевых документов, автоматически обновляемых при расширении или ревизии фонда первичных данных.

Переход от обычных (статических) документов, сохраняющихся в том виде, как они были созданы, к динамическим электронным документам - принципиально новый шаг во всей технологии подготовки и распространения справочных данных. Динамический документ определяют как совокупность объектов (таких как текст, графика, таблица, программный модуль и др.), которая динамически компоуется в единый документ по запросу пользователя в процессе его интерактивной работы с документом. Основная особенность такого документа - распределенный характер информации и ее динамическая сборка в момент запроса со стороны пользователя, хотя пользователь при работе с таким документом непосредственно не наблюдает особенностей его структуры и компоновки. Технология работы с документом, подлежащему динамической сборке при обращении к нему, порождает целый ряд преимуществ, как для пользователя, так и для эксперта, отвечающего за его изготовление:

- для пользователя доступна более «свежая» информация, поскольку при каждом обращении заново генерируется содержание документа;

- для создателя документа легче обновлять содержание, так как обновление любого фрагмента приводит к обновлению всех документов, содержащих указатель на этот фрагмент;

- информация, составляющая динамический электронный документ, является распределенной, то есть может храниться на разных (в том числе, удаленных) серверах;

- реализуется главный принцип современной технологии программирования - однократное хранение при многократном использовании, что экономит дорогостоящие ресурсы памяти.

В итоге можно обеспечить перманентную генерацию новых версий тестов по мере появления новых тестовых вопросов.

Публикация новых данных ставит и другую задачу – повторное проведение экспертизы и обработки всей совокупности данных с проверкой их согласованности и достоверности. Типичной является ситуация, когда новые данные требуют от эксперта пересмотреть отношение к старым. Абстрагируясь от научной стороны вопроса, можно сказать, что такие процедуры всегда связаны с необходимостью одновременной работы с большим числом документов, в которых фиксируются как сами данные, так и детали, связанные с их обработкой. Сама процедура ревизии и обработки данных достаточно трудоемка и включает много звеньев, так что желательно для последующей работы фиксировать не только исходные данные, но и весь протокол рабочего процесса, приведшего к новым оценкам. Если не зафиксировать весь объем этой промежуточной информации, эксперту придется повторять громоздкую процедуру анализа и обработки без учета уже найденных ранее решений. Надо учитывать также необходимость согласования работ на отдельных звеньях технологической цепочки, требования по координации работы различных экспертов, в том числе,

удаленных друг от друга и проч. Справиться со всем комплексом подобных проблем, тяжесть которых нарастает по мере увеличения масштабов справочно-аналитической деятельности, может помочь нашедшая в последние годы применение система электронного документооборота. Под термином электронный документооборот понимается формализованная совокупность процедур сбора, хранения, передачи и обработки документов в их жизненном цикле. Основу систем электронного документооборота, который стал внедряться как система ведения архивно-библиотечных средств, процедур поддержки решений и др., является интеграция различных продуктов и технологий в рамках единой среды с постепенным вытеснением «бумажного документа».

Системы электронного документооборота заменяют традиционное ведение БД, позволяя автоматизировать все стадии работы с документом: его ввод, индексирование и каталогизацию, отслеживание рабочего процесса (workflow) движения документа между разными экспертами, процедуры аналитической обработки и, наконец, генерацию новых документов с их распространением по потребителям информации.

При этом система следует алгоритмам, регламентирующим движение документа, автоматически фиксируя его модификации после работы каждого исполнителя.

Из всего спектра задач по манипуляции документами, решаемых системами электронного документооборота, технологии подготовки АПИМ наиболее близка задача управления потоком работ (workflow). В основном, она включает, так называемую, систему маршрутизации и контроля исполнения. Разработано несколько подходов к маршрутизации рабочего процесса: документо-ориентированный (маршрутизируется документ как основной объект, а остальные параметры маршрутизации ассоциируются), работо-ориентированный (основным объектом является работа, к которой прикрепляют разнообразный список документов и приложений), жесткая

маршрутизация документов по заранее определенным маршрутам с контролем исполнения, свободная маршрутизация с назначением маршрута по результатам движения документов и ряд др.

Поддержание и хранение протоколов движения документов, если они включают наборы данных и параметры тестов, необходимы для оценки результатов работы эксперта, при пересмотрах ключевых величин и многих других задач. Хранение и использование всей информации о процедурах предыдущего этапа позволяет облегчить и формализовать последующие работы по ревизии и обновлению данных.

Подобные протоколы могут помочь также в решении проблемы сохранения исходной и промежуточной информации при смене поколений научных работников.

2.2. Требования по работе с плохо алгоритмируемыми задачами.

Переход от хорошо алгоритмируемых задач, с которых начиналось развитие вычислительной техники (математическая физика, теория управления, задачи оптимизации и проч.), к задачам плохо формализуемым, которые лежат в основе бизнеса и большинства дисциплин, изучающих природу и человека (науки о Земле, медицина, психология, общественные науки) в последние годы становится одной из основных тенденций в развитии информационных и вычислительных технологий. Работа со АПИМ, понимаемая в широком смысле, относится именно к плохо формализуемым задачам, поскольку в этой деятельности приходится использовать большие массивы неполной и искаженной количественной информации, качественные оценки и т.п. Работа со АПИМ является столь сложной и многозвенной именно из-за принципиально плохой формализуемости многих этапов, связанных с оценкой тестов, их проверкой на полноту, согласованность и достоверность.

3. Разработка модели оценивания при аттестационно-педагогических измерениях. Оценка эффективности предложений

3.1. Разработка требований к модели аттестационного педагогического измерителя.

По ГОС высшего профессионального образования в редакции 2000 г. каждый вуз обязан разработать собственные профессиональные образовательные программы, включающие учебный план, рабочие программы дисциплин, контрольные материалы по дисциплинам. В целях обеспечения разработки контрольных материалов рабочие программы по дисциплинам должны содержать перечень дидактических единиц учебного материала, подлежащих контролю при аттестации. Дидактическими единицами являются законы и закономерности, характерные для конкретной дисциплины. В первом приближении в качестве дидактических единиц могут быть приняты разделы дисциплин, приведенные в ГОС.

Опыт показал, что оптимальное количество контролируемых дидактических единиц в одном контрольном материале от 4 до 12. Для контроля освоения каждой дидактической единицы используется набор заданий (3-7) одинаковой трудности, уровень которой установлен ГОС. Поскольку создание заданий, адекватных по трудности требованиям ГОС, является прерогативой эксперта-разработчика контрольного материала, появляется вероятность создания разнотрудных заданий. В связи с этим решение об освоении дидактических единиц принимается и когда экзаменуемый не выполнил всех заданий из предъявленного набора. Обычно критерий освоения дидактических единиц составляет 60-70% выполненных заданий из набора.

Исходя из результатов анализа, проведенного выше, модель аттестационно-педагогического измерителя должна отвечать следующим требованиям:

1. Для адекватного контроля освоения содержания ГОС аттестационно-педагогический измеритель должен оценивать каждый элемент структуры

(дидактическую единицу, раздел дисциплины) на заданном уровне освоения. Оценка каждого раздела должна быть проведена в бинарной шкале "освоен — не освоен".

2. Для повышения надежности оценки освоения каждого раздела дисциплины (дидактической единицы), что аттестационно-педагогический измеритель должен содержать несколько заданий для контроля каждого раздела. Таким образом, такой измеритель для отдельной дисциплины должен строиться как совокупность групп заданий с относительно небольшим числом заданий (3-7) одинаковой трудности.

3. Для каждой группы заданий должен быть задан критерий освоения раздела дисциплины (норма). Обычно для измерителей такого рода нормой является 60-70% правильно выполненных заданий группы.

3.2. Предлагаемая модель методики обработки результатов аттестационно-педагогического измерения

Результат аттестационно-педагогического измерения по дисциплине в целом для каждого слушателя будет представлять собой сумму зачтенных групп заданий (освоенных разделов дисциплины). Эту сумму необходимо нормировать на общее количество групп заданий (разделов дисциплины), чтобы полученный результат интерпретировался как доля освоенного программного материала (разделов дисциплины) на уровне требований ГОС по данной дисциплине.

При таком способе обработки результаты опроса всех слушателей будут распределены по шкале «Доля освоенных разделов дисциплины». В соответствии с требованиями закона «Об образовании» [1] в этом ряду следует выделить долю слушателей, полностью освоивших дисциплину на уровне требований ГОС.

Для более подробного анализа состояния уровня подготовки слушателей по данной дисциплине результаты опроса представим графиком,

построенным в координатах, где ординатой является процент участников опроса, изменяющийся от 0 до 100%, а абсциссой - процент освоенных ими разделов дисциплины, также изменяющийся от 0 до 100%. Вид распределения доли слушателей от доли освоенных ими разделов дисциплины «Информатика» на примере группы представлен на рис. 3.1-3.2.

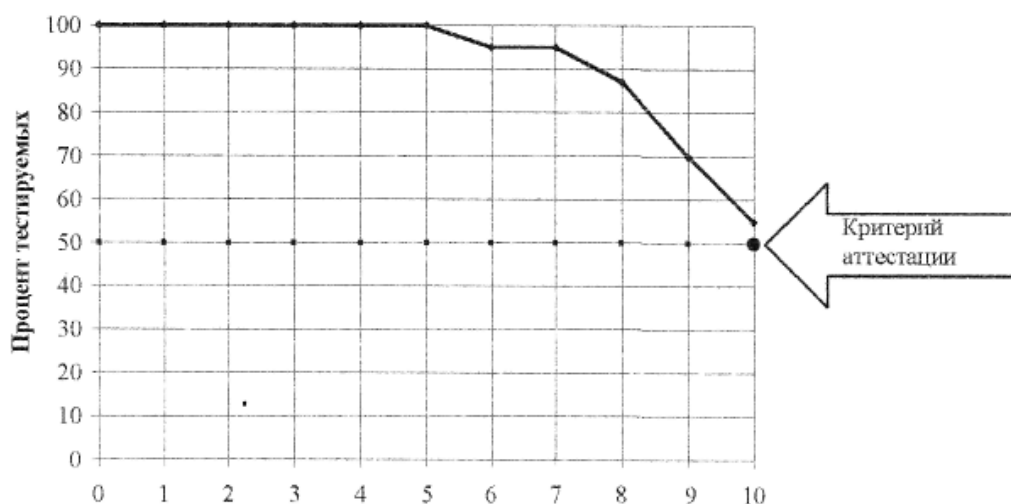


Рис. 3.1. Результаты тестирования группы по дисциплине «Информатика»

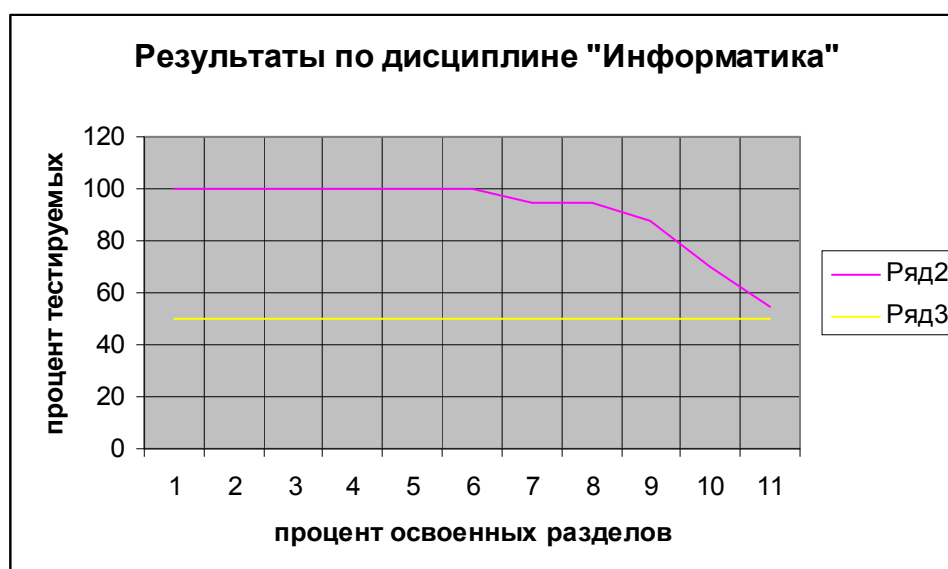


Рис. 3.2. Вид распределения доли слушателей от доли освоенных ими разделов дисциплины

Жирная ломаная линия на приведенном рисунке иллюстрирует результаты контроля с помощью предлагаемой модели аттестационно-педагогического измерителя. Точка с наименованием "Критерий аттестации" соответствует наличию критерия — 50% слушателей, полностью освоивших дисциплину на уровне требований ГОС. На этом примере доля слушателей, освоивших дисциплину в полном объеме, выше критерия аттестации, и следовательно, учебное заведение гарантирует качество обучения по данной дисциплине определенной ООП на уровне государственных аттестационных требований. Построение таких же зависимостей для других ООП позволит принять решение об аттестации (не аттестации) по данной дисциплине.

3.3. Математическая модель оценки результатов освоения дисциплины

Опишем математическую модель для построения графика освоения дисциплины (см. рис. 3.1) и вытекающие отсюда требования к уровню трудности заданий в АПИМ.

Модель расчёта вероятности решения групп заданий

На первом этапе моделирования представим АПИМ как совокупность независимых разделов дисциплины, которые имеют одинаковый коэффициент решаемости (вероятность решения). Необходимо получить значение вероятности решения совокупности всех разделов (групп заданий) АПИМ. Для построения графика освоения следует рассчитать вероятности решения не только всех разделов АПИМ, а на один, на два или на три раздела меньше общего числа.

Такая модель может быть описана формулой Бернулли для расчета вероятности не менее k благоприятных событий из общего числа n

независимых событий, причем вероятность благоприятных событий одинакова и равна p .

Формула Бернулли (или биномиального распределения) имеет вид:

$$P_n^k = C_n^k \cdot p^k \cdot (1-p)^{(n-k)}, \quad (1)$$

где $C_n^k = \frac{n!}{k!(n-k)!}$.

Первая задача по моделированию графика освоения дисциплины — по результату выполнения всего АПИМ, т.е. по крайней правой точке графика (см. рис.1). Требуется получить такое значение вероятности решения раздела, при котором график должен давать 50% вероятности выполнения совокупности всех разделов АПИМ, т.е. чтобы не менее 50% слушателей получали зачет по всем разделам АПИМ. Это является по сути обратной задачей, т.к. надо определить исходную вероятность p решения раздела при известном значении полной вероятности выполнения всех n разделов из n . Результаты расчета приведены в таблице 3.1 и на рисунке 3.3.

Введем следующие сокращения:

- коэффициент верного выполнения всего АПИМ - КВВ АПИМ,
- коэффициент верного выполнения одного раздела АПИМ - КВВ раздела (группы заданий) АПИМ (табл. 3.1).

Таблица 3.1

Вероятность освоения одного раздела дисциплины в зависимости от числа разделов (групп заданий) в АПИМ дисциплины (при условии освоения всей дисциплины с вероятностью 0,5)

Число групп заданий (разделов в АПИМ)	4	5	6	7	8	9	10	11	12
Вероятность верного выполнения 1 группы заданий (раздела) АПИМ	0,841	0,871	0,891	0,905	0,917	0,926	0,933	0,939	0,944

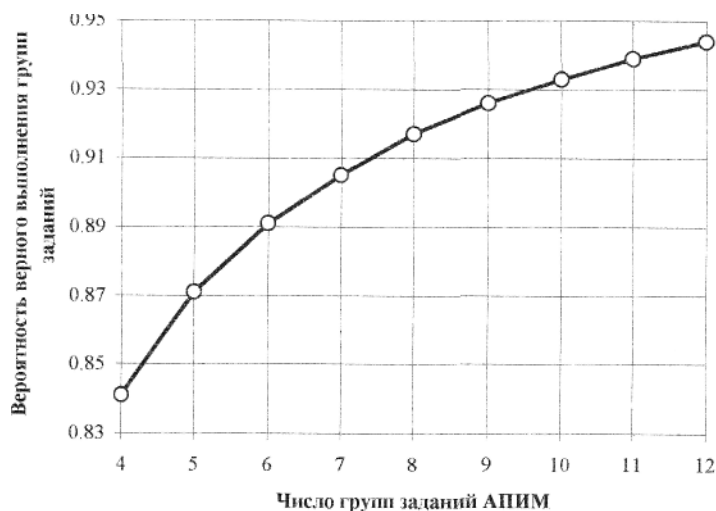


Рис.3.3. Моделирование зависимости КВВ раздела (группы заданий) от числа разделов в АПИМ по дисциплине при принятом значении КВВ АПИМ, равном 0,5

Как видно из приведенных данных, значения КВВ раздела АПИМ, достаточно велики, т.е. практически все слушатели (около 90%) должны получить зачет по каждому разделу дисциплины. С одной стороны, такие высокие значения могут быть оправданы достаточной легкостью контрольных заданий, поскольку требования ГОС относятся к минимально достаточному уровню освоения дисциплин. С другой стороны, как правило, освоение каждого раздела дисциплины проверяется группой (набором из 3-7) заданий с установленным критерием зачета (требуется выполнить только часть заданий группы). Поэтому для группы заданий необходимо рассчитать вероятность верного выполнения одного задания. Обозначим эту величину как коэффициент верного выполнения задания по теме раздела дисциплины — КВВ тематического задания АПИМ.

Эта задача вновь является "обратной", поскольку известно значение КВВ раздела АПИМ, а необходимо найти значение КВВ тематического задания АПИМ. Напомним, что значения КВВ разделов АПИМ приведены в табл. 3.1.

Для группы из трех заданий с критерием зачета "верное выполнение не менее двух заданий" формула Бернулли будет содержать два слагаемых:

$$P_3(k \geq 2) = P_3^2 + P_3^3 \quad (2)$$

Для группы из 5-ти заданий с критерием зачета "выполнение не менее 3-х заданий" сумма будет состоять из 3-х слагаемых:

$$P_5(k \geq 3) = P_5^3 + P_5^4 + P_5^5 \quad (3)$$

Аналогичная формула получается и для раздела из семи заданий.

Результаты расчетов приведены в табл. 3.2 и представлены на рис. 3.3

Таблица 3.2

Число групп заданий (разделов) АПИМ	КВВ раздела АПИМ	КВВ тематического задания АПИМ при числе заданий в группе		
		Три (критерий зачёта 2, т. е. 67%)	Пять (критерий зачёта 3, т. е. 60%)	Семь (критерий зачёта 4, т. е. 57%)
4	0,841	0,748	0,703	0,676
5	0,871	0,775	0,727	0,698
6	0,891	0,795	0,745	0,714
7	0,905	0,81	0,758	0,726
8	0,917	0,823	0,771	0,737
9	0,926	0,834	0,781	0,746
10	0,933	0,842	0,789	0,754
11	0,939	0,85	0,796	-
12	0,944	0,856	0,803	-

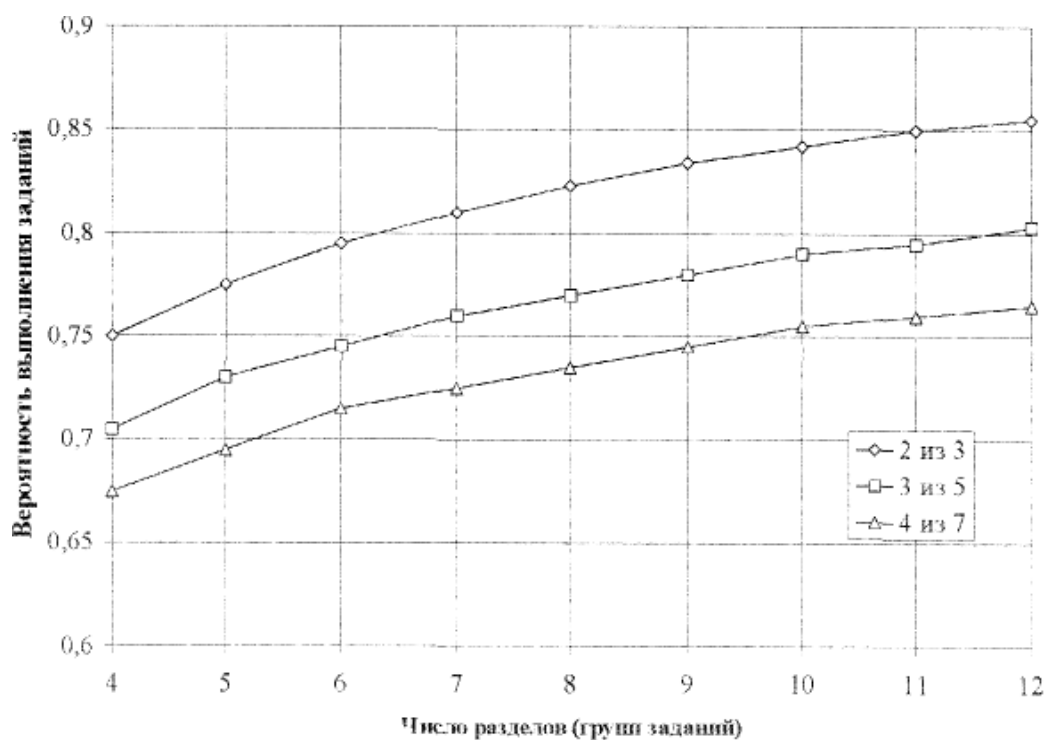


Рис. 3.3. Моделирование зависимости КВВ тематических заданий от числа групп заданий АПИМ при различных количествах заданий в группе (3, 5, 7 и соответственно критериям зачета 2, 3, 4) при постоянном значении КВВ АПИМ равном 0,5

Как видно из табл. 3.2 и рис. 3.3, при использовании групп из пяти заданий вероятность правильного выполнения отдельного задания лежит в диапазоне от 0,7 до 0,8, что является вполне приемлемым для конструирования измерителей по оценке минимальных требований к уровню подготовки слушателей. Иными словами, при разработке заданий для АПИМ с такой структурой их уровень трудности должен находиться в пределах от 0,7 до 0,8.

При построении АПИМ из групп по 7 заданий с критерием зачета не менее 4 значения уровня трудности заданий становится еще меньше.

Следует иметь в виду, что данные таблицы 2 относятся к средним значениям КВВ тематических заданий АПИМ, т.е. отдельные задания могут иметь другие значения коэффициентов трудности. Это дает возможность сравнивать эмпирические данные с предлагаемой моделью расчета

результатов контроля. Анализ полученных результатов позволяет сделать вывод о том, что математическая модель оценки освоения дисциплины с помощью АПИМ позволяет:

- определить ориентировочный уровень трудности тематических заданий при разработке АПИМ с различным числом групп заданий;
- определить средний уровень КВВ раздела АПИМ для различных значений КВВ АПИМ в целом;

" выработать статистические критерии сертификационных требований к АПИМ.

3.4. Сравнение модельных и эмпирических результатов аттестационных педагогических измерений

В качестве примера для сравнения модельных и эмпирических результатов освоения дисциплины с помощью АПИМ был взят апробированный на 111 слушателях макет АПИМ по информатике для экономических специальностей. По результатам апробации были выделены для контроля 10 разделов и установлены критерии их освоения для получения оценки "зачет".

По результатам апробации были рассчитаны экспериментальные коэффициенты решаемости разделов с соответствующими критериями освоения (рис. 3.4). По средним значениям коэффициентов решаемости заданий в группах был проведен теоретический расчет вероятности решения групп заданий, результаты которого также показаны на рис. 3.4.

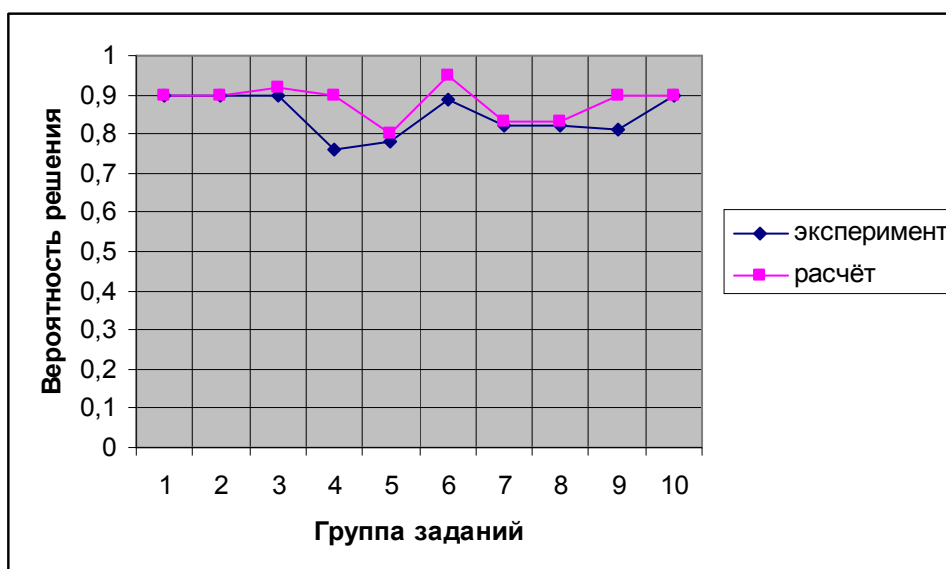


Рис 3.4. Экспериментальная и расчетная вероятность решения групп заданий для АПИМ по информатике

Как видно из рисунка 4, для 7 из 10 групп заданий получено хорошее согласие расчетных и экспериментальных значений.

В соответствии с математической моделью для АПИМ из 10 групп заданий, среднее значение вероятности решения групп заданий должно быть равно 0,93. По данным апробации получается значение 0,85, что приводит к меньшему проценту слушателей, освоивших все разделы АПИМ. Сравнение математической модели графика освоения АПИМ и экспериментальной зависимости приведено на рис. 3.5, причем математическая модель графика рассчитана как для экспериментальных значений вероятности решения разделов, так и для расчетных (теоретических), которые приведены на рисунке 4. Как видно из рис. 3.5, реальная экспериментальная кривая дает большее значение доли слушателей, полностью освоивших дисциплину на уровне требований ГОС.

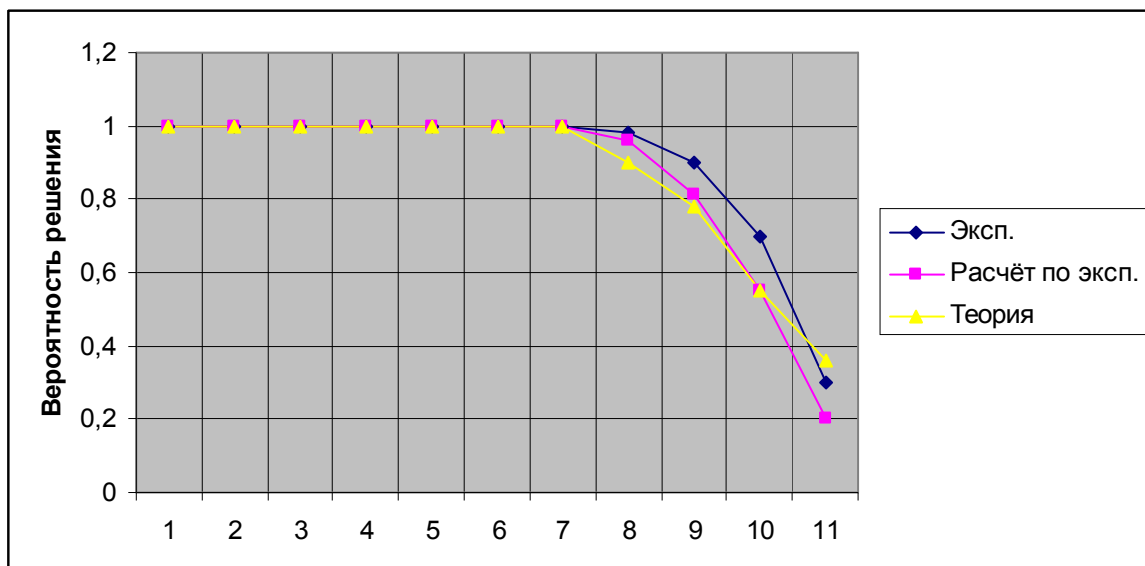


Рис. 3.5. Расчетные графики и экспериментальные значения результатов оценки освоения дисциплины с помощью АПИМ по математике