

Министерство образования Российской Федерации

Санкт-Петербургский государственный политехнический университет

Б.С. ИВАНОВ

**ОСНОВЫ ПЕДАГОГИЧЕСКОЙ ДИАГНОСТИКИ
И МОНИТОРИНГ ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ
В ТЕХНИЧЕСКОМ ВУЗЕ**

МОНОГРАФИЯ

Санкт-Петербург - 2003

УДК 159.92

Печатается по решению Редакционно-издательского Совета СПбГПУ.

Рецензент: д-р пед. наук, проф. А.И. Сурьгин

Иванов Б.С.

Основы педагогической диагностики и мониторинг образовательной деятельности в техническом вузе

Монография.- СПб.: Изд-во СПбГПУ, 2003. - 120 с.

В монографии освещены методологические аспекты разработки, применения и обработки результатов педагогических и психологических тестов. Материалы монографии предназначены для слушателей ФПКП и преподавателей, приступающих к разработке тестов, применяющих или собирающихся применить тестирование в своей педагогической практике, а также для студентов, аспирантов и стажёров, избравших педагогику в качестве специальности основного или дополнительного образования.

Оглавление

	с.
Введение	4
1. Основы диагностики	10
2. Методологические аспекты использования тестовых технологий....	13
3. Таксономические показатели диагностирования.....	26
4. Статистические параметры тестовых заданий.....	42
5. Диагностические показатели теста	51
6. Диагностическая ценность признаков и обследования.....	64
7. Образовательный мониторинг.....	73
8. Психологическое тестирование	82
9. Анкетирование	105
10. Система компьютерного тестирования.....	107
Заключение	110
Литература	113

Введение

Эффективное управление образовательным процессом невозможно без получения достоверных данных о его результатах. Тенденции развития современной отечественной системы высшего профессионального образования, её интеграция в международное образовательное пространство обуславливают более широкое использование достижений информационных технологий при решении задач контроля и обеспечения необходимого уровня качества образования. В образовании, как и в любой другой любой сфере деятельности, стремятся достичь максимально возможных результатов в условиях имеющихся ресурсов, которые, как правило, бывают ограниченными.

Под качеством образования можно понимать меру соответствия фактически достигнутых результатов предварительно планируемому. Гарантии качества образования без измерений его результатов носят во многом декларативный характер. Свидетельством высокого качества учебного процесса не может служить количество выставленных преподавателем отличных или хороших оценок, так как, во-первых, в образовательных учреждениях отсутствуют единые критерии для оценивания учебных достижений учащихся, а во-вторых, преподавателю не этично самому оценивать результаты своей профессиональной деятельности. Ведь фактически оценка - это не только характеристика уровня подготовки учащегося, но и показатель результативности работы самого преподавателя.

Педагогическое инструментальное средство для решения обозначенных проблем – это тестирование. Многие преподаватели, да и не только они, к примеру, студенты, абитуриенты, их родители, руководители образовательных учреждений, не совсем верно представляют себе, в чём состоит его сущность. Это порождает негативное отношение к тестированию как к чему-то неприятному, формальному, чуждому, непонятному, неприемлемому. Такое отношение к тестированию является следствием недостаточно хорошего представления о

нём, как об инструментальном средстве образовательного процесса. Одни преподаватели видят в тестировании панацею от всех недостатков образовательного процесса, другие – способ заставить учеников трудиться с большей отдачей, третьи – отдают дань моде или указаниям свыше и, проводя тестирование, вообще не задумываются над тем, что они делают. Отсутствие положительного результата объясняется ими «недостатками» тестирования: дескать «не способно тестирование заменить живое общение», «не может тестирование научить творчески мыслить» и т.п. Подобные выводы можно сравнить с сетованиями неквалифицированного врача на то, что «градусник, сколько его ни ставь, не способен вылечить больного» или, что «никакие показания ртутного столбика градусника не заменят руку врача, положенную на лоб больного».

Здесь необходимо чёткое разграничение понятий. Можно ли «обидеть» человека, сообщив ему о наличии у него какой-либо болезни, или что повышенная температура может указывать на начало у него какого-либо заболевания? Каждый заинтересован в хорошем состоянии своего здоровья и не будет скрывать от врача информацию, способствующую установлению правильного диагноза. Диагностирование – это ещё не лечение болезни, но необходимая процедура для того, чтобы лечение было правильным и давало положительный результат. Если рассматривать тестирование как диагностическую процедуру, то снимаются многие опасения, связанные с его использованием в образовательном процессе. Необоснованное ожидание «чуда», которое должно произойти при введении тестирования, убережёт от разочарований.

Любой человек хотел бы иметь подробную и достоверную информацию о состоянии своего здоровья. У нас с вами не бывает отрицательного отношения к флюорографическому обследованию, к сдаче различного рода анализов, измерению давления, проверке зрения и слуха. В сущности, задачи педагогического тестирования те же – на ранней стадии выявить возможные проблемы и совместно с преподавателем победить «болезнь незнания». В конце концов, не знать что-либо не зазорно, плохо лишь не хотеть знать больше и лучше.

Проблемы контроля в образовании целесообразно рассматривать в контексте их исторического развития /2/. Свидетельства об осуществлении контроля результатов обучения сохранились до нас в исторических документах, возраст которых насчитывает тысячелетия. Уже в середине III тысячелетия до н. э. в Древнем Вавилоне проводились испытания выпускников в школах, где готовились писцы. Писец был одной из центральных фигур месопотамского общества: он умел измерять поля, делить имущество, петь, играть на музыкальных инструментах. Во время экзаменационных испытаний проверяли его умения писать, разбираться в тканях, металлах, растениях, а также знание им всех четырех арифметических действий /17/.

В эпоху Среднего и Нового царств в Египте уже существовали школы, учась в которых «студенты» должны были знать законы и уложения, историю и географию, основы техники, основы грамматики и начала рисунка /43/. Образование позволяло египетским чиновникам выполнять самые разнообразные обязанности и с большой лёгкостью переходить с одной должности на другую: так один из них по имени Уна был сначала стражником, потом судьёй, затем занимался поиском камней для изготовления памятников, строил корабли, очищал каналы, а когда разразилась война, стал начальником штаба. Обучение и опрос вели чиновники и жрецы. Так писец Хори, экзаменуя своих учеников и желая уличить их в невежестве, задавал вопросы: «Каков рацион войска на марше? Сколько нужно кирпичей для постройки платформы заданных размеров? Сколько нужно людей для перевозки обелиска? Как поставить на пьедестал колоссальную статую? Как организовать военную экспедицию?» И под конец ещё ряд вопросов по географии Сирии. Учебная программа, по тем временам, - весьма насыщенная, а постановка и содержание вопросов свидетельствуют, что опрос по форме был довольно близок к современному трактованию тестирования. В методических указаниях писца Аменемопе мы читаем: «Пиши рукой своей, читай устами своими, советуйся со знающими больше, чем ты. Упражняйся в этом деле, и найдёшь ты его [полезным] после [наступления] старости.

Счастливы пища искусный в деле своём... Будь настойчив в обучении, и ты овладеешь им [своим делом]. Не проводи ни одного дня в безделье, иначе будут бить тебя. Уши юноши на спине его и он внемлет, когда его бьют. Пусть внемлет сердце твоё тому, что сказал я, - будет это полезно тебе... Будь настойчив в получении советов! Не ленись! Пиши! Не предавайся пресыщению [письмом], отдай ты сердце и внимай словам! Найдёшь ты их полезными» /перевод М.А. Коростовцева/.

В Древнем Египте искусству жрецов обучали только тех, кто выдерживал систему определенных испытаний. Вначале кандидат проходил процедуру, которую сейчас можно было бы назвать собеседованием. При этом выясняли биографические данные, уровень образованности, оценивали внешность, умение вести беседу. Затем проверяли умения трудиться, слушать, молчать. Проводили испытания огнем, водой и угрозой смерти. Тем, кто не уверен, что выдержит все тяготы длительного образования, предлагалось подумать - с какой стороны окончательно закрыть за собой дверь в храм - с внутренней или внешней /76/. Эту суровую систему испытаний и отбора успешно преодолел в молодые годы Пифагор. Вернувшись после учебы в Грецию, он основал свою школу, допуск в которую открывал только после серии различных испытаний, похожих на те, которые выдержал он сам /12/.

Использование письменных контрольных работ и экзаменов в странах Востока уже в те годы считалось вполне нормальным и естественным делом, однако этого нельзя сказать о странах Запада. Ценность письменных работ быстрее других оценил Орден Иезуитов, увидевший в них средство повышения эффективности учебных занятий. Используя свое влияние, Орден распространил практику применения письменных работ во многих европейских странах (т.н. «иезуитские методы» в образовании), где он имел влияние. Ответная реакция на это в первую очередь проявилась в Америке. В 1762 г. там состоялось открытое выступление студентов Йельского университета против использова-

ния письменной формы экзаменов, из-за которых им приходится много "зубрить".

В 1884 г. в США вышла первая книга с тестовыми материалами, содержащая задания и ответы к ним с оценками по пятибалльной шкале. В этой книге содержались задания по математике, истории, грамматике, навигации, давались примерные тексты для сочинений вместе с методом количественной оценки сочинений. Это был первый в истории случай использования в педагогике простейших статистических расчетов. Быстро распространявшаяся в США, такая форма контроля знаний имела не только много сторонников, но и противников. Известно, например, о школьном инспекторе из Чикаго, запретившем в 1881г. проводить письменный контроль и обязавшем осуществлять перевод учащихся в очередной класс не на основе проверок, а опираясь исключительно на мнения учителей и директоров школ. Он полагал, что никто лучше учителя не сможет оценивать способности ученика, и потому экзамены и всякие другие формы контроля являются издевательством над здравым смыслом. Так было положено начало общественному движению против какого-либо контроля в школе. Пик такого движения в США пришелся на конец XIX века (1880-е годы), после чего вновь стали появляться предпосылки для формирования благожелательного отношения к контролю знаний.

Потребность в создании общественно-государственной системы контроля знаний была осознана в США в 1885 году. В Нью-Йорке начал действовать экзаменационный совет - один из немногих советов, которые смогли разработать методы объективного контроля знаний. Начало общегосударственной системы оценки знаний положило создание в 1900 году комитета по проверке знаний абитуриентов колледжей в США.

В России конца XIX - начала XX века впервые ставится вопрос разработки педагогических тестов и их применения. "Проверка знаний, - писал В.Я. Уланов, - представляется мне по опыту самым скучным, бесполезным и даже вред-

ным занятием. Она терпима, как необходимое зло, только там, где от учителя требуют ежемесячных отметок как средства контроля не только над учениками, но и над учителем" /70/.

Как отмечал позднее В.А. Сухомлинский /68/: «... Следует иметь в виду некоторые подводные камни в самой логике педагогического процесса: обучение проникнуто постоянной повседневной проверкой (контролем), ежечасным сравнением успехов одного ученика с успехами другого. За всем этим таится опасность разочарования, неуверенности в своих силах, замкнутости, равнодушия, озлобления, то есть таких душевных сдвигов, которые приводят к огрублению души, утрате чуткости к ... слову и красоте».

Но кораблю образовательной системы нужны не только ветер и паруса, необходимы также лодья, руль и надёжный рулевой. «В теории управления, - отмечает один из известнейших педагогов В.А. Якунин /78/, - контроль рассматривается как важнейшее, относительно самостоятельное и замыкающее звено в управленческом цикле. Контроль является механизмом выявления и оценки результатов произведённого действия. Его основное назначение состоит в обеспечении обратной связи, осведомляющей о соответствии фактических результатов функционирования системы её конечным целям. Если бы реальные результаты всегда совпадали с нормативами, контроль вряд ли когда-либо был уместен».

Эффективность образовательного процесса и функционирования вуза в целом, обуславливается, прежде всего, качеством образования. Стремление управлять качеством образовательного процесса обуславливает необходимость сопоставления его результатов, так как управление возможно только при условии наличия различий. Обеспечение качества образования, подтверждённое фактической статистикой тестовых испытаний, должно являться одной из стратегических целей высшего учебного заведения.

1. Основы диагностики

Под диагностикой понимается наука о распознавании действительного состояния любых объектов в условиях ограниченной или избыточной информации. В процессе диагностирования устанавливается диагноз, т.е. с той или иной степенью достоверности определяется состояние объекта диагностирования. Наблюдая за изменением состояния можно прогнозировать развитие и поведение объектов диагностирования. Наиболее широко методы диагностирования применяются при определении уровня владения учебным материалом (педагогическая диагностика), типа и состояния личности (психодиагностика), состояния больного (медицинская диагностика), состояния технических объектов или систем (техническая диагностика).

Основная цель педагогического диагностирования – получение диагностической информации. Её интерпретация и использование – это отдельные проблемы и в большинстве случаев не должны являться компетенцией преподавателей. Для принятия верных решений информация о реальном положении дел должна быть объективной, информативной и контролепригодной.

Объектами диагностирования может быть сама образовательная система и (или) её элементы: субъекты и объекты образовательного процесса, содержание учебных дисциплин, успеваемость и т.д. Разработка системы педагогической диагностики предусматривает поиск наиболее эффективных методов решения следующих задач:

- создание информационных моделей контентного анализа учебного материала изучаемых дисциплин;
- таксономический анализ, выделение уровней усвоения учебного материала, их семантический анализ;
- определение статистических показателей тестовых заданий, определяющих их диагностические свойства;

- исследование онтологии и семантики тестовых заданий с позиций интеррогативной (эпистемической) логики, построение математических моделей тестовых заданий;
- определение диагностических свойств теста как композиции тестовых заданий, оценивание показателей ценности и полноты диагностического обследования;
- разработка программно-методического обеспечения информационной среды педагогического и психологического диагностирования.

Теоретическим фундаментом для решения основной задачи диагностики принято считать /7/ общую теорию распознавания образов (состояний или диагнозов объектов диагностики). Определение диагнозов зависит от особенностей задачи и целей исследования. В большинстве случаев диагнозы (классы) устанавливаются заранее, и в этих условиях задачу распознавания часто называют задачей классификации, т.е. отнесения объекта диагностирования к определённом классу. Алгоритмы распознавания частично основываются на диагностических моделях, устанавливающих связь между состояниями контролируемого объекта и их отображением в пространстве диагностических признаков. Например, диагноз «успеваемость» может быть связан с набором диагностических признаков: уровень начальной подготовки, посещаемость занятий, состояние здоровья, уровень компьютерной грамотности и т.д. Решение диагностической задачи всегда связано с риском ложного срабатывания (ЛС) или пропуска цели (ПЦ). Эти термины в диагностике появились из радиолокации, где впервые стали применяться методы теории статистических решений. Важной частью проблемы распознавания являются правила принятия решений (решающие правила).

Вторым важным направлением диагностики является теория контролеспособности. Контролеспособностью называется свойство обеспечивать достоверную оценку диагноза и создаётся принятой системой диагностики. Весьма значительной задачей теории контролеспособности является изучение средств и

методов получения диагностической информации. Методы создания автоматизированных систем контроля, в том числе тестового, составляют одно из направлений контролеспособности. Контролеспособность связана также с разработкой алгоритмов диагностических тестов, минимизацией параметров процесса установления диагноза. Так как диагностика предполагает обработку большого объема информации, то принятие решения (распознавание) обычно осуществляют с помощью компьютеров.

На рис. 1 показана структура диагностики. Она характеризуется двумя взаимосвязанными направлениями: теорией распознавания и теорией контролеспособности.

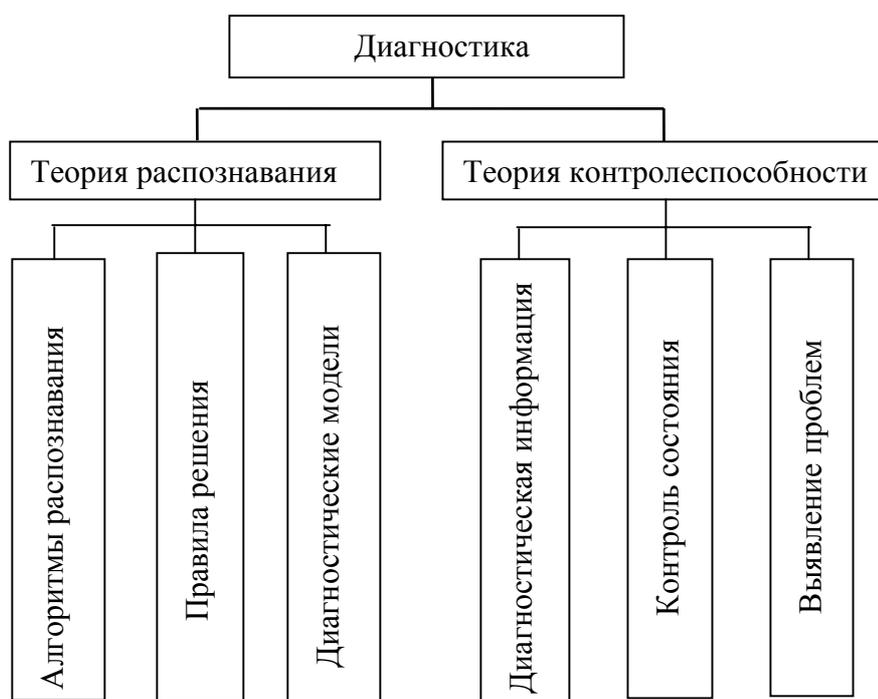


Рис. 1 Структура методов диагностики

Состояние объекта диагностирования описывается совокупностью определяющих его параметров (признаков). Это множество определяющих параметров может быть различным, в первую очередь, в связи с самой задачей распознавания. Роль диагностических признаков при педагогическом диагностирова-

нии играют тестовые задания, которые должны отвечать указанным выше определённым требованиям (информативность, контролепригодность и др.) Совокупность последовательных действий в процессе распознавания представляет собой алгоритм распознавания.

Различают системы функционального и тестового диагностирования. В первом случае анализируются параметры объекта диагностирования в его нормальном рабочем состоянии. Во втором случае объект подвергается специальному воздействию, и анализу подвергается его отклик на это воздействие. Педагогическое и психологическое тестирование обычно относятся ко второму виду диагностирования.

2. Методологические аспекты использования тестовых технологий

Составление и выполнение тестовых заданий связано с моделью будущей профессиональной деятельности специалиста. При этом под моделью понимается такая мысленно представляемая система, которая, отображая объект исследования, способная замещать его так, что её изучение дает нам новую информацию об этом объекте. Существуют и получили признание несколько моделей педагогического диагностирования. Л.С. Выготским предложена модель «круга диагностического исследования», имеющая структуру, схожую с циклом Деминга при управлении качеством и содержащая следующие этапы:

- эмпирический: определение объекта и целей диагностирования, выделение и фиксация исходных симптомов, особенностей состояния;
- этиологический: определение причин, вызывающих те или иные симптомы и особенности;
- типологический: определение уровня развития и формулирование предположения о пути развития, формулируется гипотеза о причинах выявленных проблем;

- диагностический: получение данных о сущности, данные соотносятся с типологией проблемных областей;
- прогностический: определение наиболее вероятного пути развития, формулируются предложения о преодолении выявленных проблем;
- корректировочный: определение корректирующего воздействия, наблюдение за эффективностью воздействия.

Модель Л.С. Выготского получена на основе теоретического обобщения идей педагогической диагностики. Существует также подход к педагогическому диагностированию Н.В. Кузьминой, появившийся в результате развития практических методов, в основе которых лежат следующие этапы:

- определение цели диагностирования;
- анализ ситуационной модели;
- определение форм, методов и средств диагностирования;
- определение взаимодействия преподавателя и студента в ходе диагностирования;
- контроль и оценка действий студентов;
- анализ педагогической деятельности преподавателя по результатам диагностического обследования;
- анализ познавательной деятельности студентов по результатам обследования;
- установление причин и степени отклонений от эталона (норматива);
- определение корректирующего воздействия.

Существует довольно много различных подходов, но какова бы ни была модель, она представляет собой компромисс между всеобщностью, точностью и реалистичностью. Усиление одной из этих компонент неизбежно ведёт к ослаблению двух других. Следствием этого является то, что при имеющихся реальных физических ограничениях на объём обрабатываемой информации, результат получается тем определённое, чем уже область приложения этого результата. Под моделирование здесь понимается представление о том, как ком-

пактно в символической форме могут быть представлены прежние, обычно расплывчато задаваемые знания и как записать алгоритмическую компоненту так, чтобы она находилась в каком-то разумном соответствии с закладываемыми в модель предпосылками. Искусство составления моделей в значительной степени определяется тем чувством меры, которое помогает уравнивать имеющиеся потребности с реальными возможностями.

В диагностировании с использованием тестовых технологий дискуссионными до сих пор являются вопросы о том, что собой представляют тестовые задания, является ли их составление искусством или технологией, может ли быть построена математическая модель тестового задания. Проблеме семантики «вопроса» как ключевого понятия (не обязательно для тестового задания) посвящены ряд исследований в области интеррогативной (эротетической) логики /72, 45/. Согласно предлагаемой концепции всякий вопрос непременно состоит из двух составляющих: утверждающей, т.е. вносящей некоторые знания и делающей вопрос возможным (так называемая предпосылка вопроса), и собственно вопрошающей. Предпосылка может быть верной, ошибочной или недостаточной для постановки данного вопроса, что можно установить, только привлекая какую-то другую, внешнюю по отношению к данному вопросу информацию. Вопрошающая часть вопроса не может быть ни истинной, ни ложной, она может быть только уместной или неуместной, хотя существует точка зрения, согласно которой сам вопрос есть уже некое смысловое предложение, детерминантом которого является его ответ. Отсюда следует, например, что вопросы могут быть неуместными или даже невозможными. К числу последних, например, могут быть отнесены вопросы: «Зачем нужна математика?» или «Когда и где появилась педагогика?», т.к. любые возможные ответы для этих вопросов будут выглядеть как нелепость. Наука всегда проблемна, то есть начинается с постановки вопросов. Можно конечно вести наблюдения или даже ставить эксперименты, не задавая вопросов, но вряд ли такую деятельность можно считать научной. Отсутствие во-

просов объясняется отсутствием теории для того, чтобы сделать содержательной предпосылку вопроса. В этом случае эксперимент превращается в слепой поиск, так как отсутствует язык, на котором можно сформулировать теоретические представления, основанные на опыте прошлого. Таким образом, особенностью научного метода является стремление *получить ответ на четко поставленный вопрос*, сформулированный на основе прежних знаний. Знания же всегда условны и изменчивы, наука развивается диалектически, поэтому корректность вопроса имеет временные рамки.

Математика представляет собой метаязык, позволяющий задавать вопросы в чрезвычайно компактной форме, используя для этого абстрактно-символическую форму записи. Любая математическая модель представляет собой хорошо поставленный вопрос. Его предпосылка – это чёткое разделение зависимых и независимых переменных, ответственных за состояние объекта моделирования, и аналитическая запись самой модели. Вопрошающая часть – это задание набора параметров модели, подлежащего численному оцениванию. Изменение предпосылки влечёт за собой изменение вопрошающей компоненты.

Итак, получить результат от тестирования, т.е. «что-то» увидеть после него, можно только тогда, когда на это «что-то» направлено внимание. Проблема или вопрос всегда существуют, но зачастую они остаются на глубоко интуитивном, иногда даже невербализированном уровне. Успешность формализации вопроса, видимо, не связана с уровнем математической подготовленности. Физику это может быть так же трудно, как, скажем, педагогу или врачу. Дело здесь может быть в стиле мышления, в развитии тех его компонентов, которые связаны с видением проблем. Тренинг и опыт разработки большого числа тестов могут привести к некоторому прогрессу. Каждую научную гипотезу, особенно если она задана математически, можно рассматривать как вопрос. В некотором смысле предпосылкой такой модели является утверждение о дискретности нашего языка и непрерывности сознания. Вопрошающая часть вопроса

направлена на выражение многообразия нашего речевого представления конкретной понятийной модели.

Наше сознание непрерывно, но тестовые задания, уровни владения учебным материалом, учебные элементы - дискретны. Это даёт возможность взглянуть на проблему с точки зрения общих статистических критериев, так как имеется некоторая особенность в методах обработки и интерпретации результатов экспериментов с дискретными факторами. Для моделей с непрерывными переменными находят параметры моделей с помощью метода наименьших квадратов, а далее, если необходимо, проверяется гипотеза об их незначимости. Для моделей же с дискретными переменными в статистике принято сначала проверять гипотезы о незначимости определённых групп параметров с помощью методов дисперсного анализа, а далее, если такого рода гипотезы отвергаются, находить методом наименьших квадратов оценки параметров при наличии ограничений на некоторые комбинации параметров.

В.В. Налимов /45/, исходя из соображений достаточно общего характера показал, что экспериментальные данные никогда не могут подтвердить правомерность той или иной модели, можно лишь показать, что результаты наблюдений не противоречат некоторой рассматриваемой гипотезе. Известный английский философ К. Поппер в 60-х годах прошлого века предложил даже отказаться от термина «проверка гипотез», заменив его термином «опровержимость гипотез». Если в результате исследования данная гипотеза не опровергается, то это ещё не означает, что гипотеза может быть принята, поскольку эти же результаты наблюдений могут не противоречить и множеству других невысказанных гипотез. В практике диагностирования это может означать, к примеру, что гипотезе «Знания студента соответствуют требуемому уровню владения учебным материалом» следует предпочесть гипотезу «Знания студента не противоречат требуемому уровню владения учебным материалом», имея в виду, что они могут при этом не противоречить и некоторым другим из выделенных уровней владения учебным материалом.

Основой принятия решений часто служит сопоставление результатов диагностирования с некоторым эталоном. В технических вузах для большинства учебных предметов роль такого эталона исполняют государственные образовательные стандарты.

Диагностическим измерительным материалом являются педагогические тесты, представляющие собой набор специальных заданий, способных решить задачи диагностирования. К. Ингенкампом /28/ предлагается следующее определение тестирования: «Тестирование - это метод педагогической диагностики, с помощью которого выборка поведения, отражающая предпосылки и результаты учебного процесса, должна по возможности максимально отвечать принципам сопоставимости, объективности, надежности и валидности измерений, должна пройти обработку и интерпретацию и быть готовой к использованию в педагогической практике». Для того чтобы педагогические тесты могли выполнять указанные функции, сами тесты должны разрабатываться с учетом выше-названных исходных положений и отвечать определенным требованиям /5/:

Содержательная валидность (от valid (англ.) – правильный) - соответствие теста содержанию и объему полученной в процессе обучения информации. Валидность в общем случае означает диагностирование именно того свойства, для которого тест разработан. Высокая валидность теста означает, что из него убраны все элементы, которые не относятся к диагностируемой области. Валидность теста позволяет производить накопление аргументов для интерпретации результатов тестирования.

Функциональная валидность - соответствие теста контролируемому уровню усвоения. Процедура диагностирования должна быть предназначена для выявления интересующего нас диагноза. Обычно это – владение учебным материалом на заданном уровне. Если диагнозов несколько, то в рамках одного тестирования статистическая достоверность определения каждого из диагнозов будет снижаться.

Определенность теста. Это требование означает, что задания в тесте должны быть сформулированы в максимально простой, корректной и лаконичной форме, не допускающей различного толкования. По статистике более половины тестовых заданий обычно, к сожалению, таковы, что студенту не понятно, что именно преподаватель хочет от него, каких действий ожидает. Читая задания теста, учащийся должен хорошо представлять, какую деятельность он должен выполнить, какие знания и в каком объеме продемонстрировать, чтобы его ответ соответствовал заранее составленному эталону. При невыполнении требований определенности теста может возникнуть ситуация, когда преподаватель квалифицирует знания учащегося как отрицательные, в то время как «виновником» неправильного ответа является он сам, предложивший неопределенное задание.

Простота теста. Это требование может быть иначе сформулировано, как необходимость иметь в заданиях теста возможно более четкую и прямолинейную формулировку на строго определённый и известный тестируемому вид деятельности, соответствующий диагностируемому уровню владения учебным материалом. При этом преподаватель будет иметь возможность четко определить, достиг ли учащийся заданного уровня усвоения учебного материала.

Однозначность теста. Это требование обеспечивается наличием в тесте эталона, определяющего систему измерения и оценки качества выполненной тестируемым деятельности по тесту. Только наличие эталона обеспечивает отсутствие противоречий в оценках различных экспертов успешности выполнения теста учащимися.

Надежность теста. Требование надежности теста состоит в устойчивости получаемых результатов при проведении тестирования одного и того же испытуемого. Для оценки теста на соответствие этому требованию производят повторное тестирование с помощью различных тестов данного вида, либо при достаточно большом объёме теста получают необходимые статистические оценки в зависимости от успешности выполнения его отдельных частей. Если

корреляция оценки удовлетворяет принятому уровню надежности, тест признается надежным, на этом же уровне.

Дифференцирующая способность теста. Тест как педагогический измерительный материал должен иметь максимально возможную разрешающую способность, т.е. разделять студентов, имеющих более высокие и менее высокие показатели учебной успеваемости. Вместо балльной системы в последнее время всё чаще используют так называемые рейтинговые системы.

Другие требования. Различными исследователями-тестологами предложено ещё довольно много требований к тестам, представляющихся менее значимыми, чем перечисленные выше.

Широкий круг решаемых с помощью тестирования задач, их разнородность и внутренняя противоречивость затрудняют принятие определённой и однозначной процедуры проведения, а также методики обработки результатов. Большинство известных систем тестирования по существу таковыми не являются и, как правило, реализуют опрос с помощью некоторого произвольного набора контрольных заданий, не обладающего диагностическими свойствами и решающего лишь некоторые частные задачи.

Проблема усугубляется в случаях необходимости осуществления тестирования посредством компьютеров в системах дистанционного обучения и удалённого доступа, в том числе и с помощью сетевых технологий. Невозможность использования в большинстве случаев онлайн-процедур накладывает определённые требования на программное обеспечение тестирования. Компьютерная программа должна не только по возможности моделировать ситуацию обычного экзамена и при этом выдавать оперативную и эффективную оценку достижений, проверяемых в ходе тестирования.

Необходимо отметить, что существует несколько различных подходов к подбору тестовых заданий, проведению тестирования и интерпретации его результатов. Различие в подходах обусловлено различиями целей процедуры педагогического диагностирования. Различают /16/ нормативно-

ориентированный, критериально-ориентированный и личностно-ориентированный подходы, предъявляющие определённые требования к форме и содержанию тестовых заданий, их композиции в тесте. Сравнительная характеристика трёх видов тестирования приведена в табл. 1.

Таблица 1

Сравнительная характеристика нормативно-ориентированного, критериально-ориентированного и личностно-ориентированного тестирований

Нормативно-ориентированное тестирование	Критериально-ориентированное тестирование	Личностно-ориентированное тестирование
Тестовые задания должны быть однородны	Допускается разнообразие форм и типов тестовых заданий	Возможно большее разнообразие форм и типов тестовых заданий
Исключаются очень лёгкие и очень трудные задания	Содержат набор заданий, выполнение которых являлось целью обучения	Задания, выполненные ранее неправильно, должны быть в тесте при его повторном прохождении
Допускается обзорный характер учебного материала тестовых заданий	Имеется спецификация учебных элементов дисциплины	Проводится сопоставление результатов по выявленным проблемным областям

Нормативное или нормативно-ориентированное тестирование применяется в тех случаях, когда необходимо получить сравнительные данные по выборке тестируемых. Например, отобрать лучших среди желающих поступить в вуз можно с помощью нормативно-ориентированного тестирования. При этом, образно говоря, с помощью теста мы выстраиваем тестируемых в одну колонну в затылок друг другу. Нормативное тестирование - идеальное средство решения подобной задачи: все находятся в равных условиях, обеспечена объективность и оперативность получения результатов. При этом виде тестирования следует исключать из теста разнородность заданий с тем, чтобы одним не достались более трудные задания, чем другим. Оценки при нормативном тестировании, как проходные баллы на вступительных экзаменах, большого диагностического значения не имеют. Для сопоставимости и различения результатов вместо сис-

темы оценок более приемлема рейтинговая система как обладающая лучшей дифференцирующей способностью. Методы статистики в сочетании с компьютеризацией опроса гарантируют высокие показатели дифференцируемости, объективности и надежности процедуры тестирования.

Критериальное или критериально-ориентированное тестирование - более совершенная система диагностики, однако существенно повышает планку требований к набору тестовых заданий. Здесь целью диагностирования является определение уровня владения учебным материалом. Такое тестирование обычно проводится как итоговое при завершении изучения раздела, дисциплины или цикла дисциплин. В идеальном случае для каждого уровня имеется свой тест, состоящий из десятков (сотен) тестовых заданий, выполнение которых должно дать однозначный ответ на вопрос, соответствует ли подготовленность тестируемого установленным на данном уровне критериальным требованиям или как правильнее, не противоречит ли подготовленность студента критериальным требованиям данного уровня.

Личностно-ориентированное тестирование производит сравнение текущих результатов тестируемого с его прошлыми результатами, оценивая, таким образом, динамику овладения учебным материалом. При этом появляется возможность повысить надёжность диагностирования проблемных областей специальным подбором заданий для повторных попыток прохождения теста. Данный вид тестирования позволяет организовать систему мониторинга учебной деятельности, однако существенно увеличивает объём хранимой и обрабатываемой статистической информации.

Отдельно необходимо остановиться на использовании тестового контроля для диагностирования готовности удовлетворить требования Государственных образовательных стандартов, которые регламентируют лишь минимально необходимый уровень подготовки, требующийся для выдачи сертификационного документа (аттестата о среднем или диплома о высшем образовании). Здесь следует отметить, что ГОСы являются не идеалом, которого необходимо дос-

тичь, а тем минимумом, который каждый студент обязан освоить (принцип минимально необходимой достаточности). Разумеется, что этим минимумом не должны исчерпываться цели обучения. При этом появляется следующая методологическая проблема: если тест как модель результата обучения соответствует объёму и содержанию государственного образовательного стандарта, который в свою очередь, отражает лишь необходимый обязательный минимум владения учебным материалом, то не должно ли из этого следовать, например, что для получения документа об образовании по результатам тестирования требуется 100%-ный положительный итог выполнения теста. Не будут ли нормативы в 60-70% правильно выполненных заданий, обычно принимаемые при тестировании, свидетельствовать об уровне владения обязательным минимумом также на уровне 60-70%, и достаточно ли будет таких «достижений» для получения документов, подтверждающих образование? ГОС также не может требовать от выпускников школ и вузов быть творческими личностями. В силу неопределённости этого требования (творческая личность) смысл документа об образовании был бы во многом утрачен. Невыполнение учащимися теста, составленного на основе требований Государственного образовательного стандарта, может свидетельствовать, например, о следующем:

- требования образовательного стандарта не соответствуют реальному состоянию и возможностям образовательной системы;
- качество обучения в конкретном образовательном учреждении не соответствует требованиям государственного образовательного стандарта;
- задания теста не соответствуют содержанию образовательного стандарта;
- методика педагогических измерений при тестировании не отражает объективные показатели учебных достижений.

Тестовый контроль отличается от других методов контроля (устные и письменные экзамены, зачеты, контрольные работы и т.п.) тем, что он представляет собой специально подготовленный контрольный набор заданий, по-

звolyающий надежно и адекватно количественно оценить знания обучающихся посредством статистических методов.

Основными достоинствами применения тестового контроля являются:

- объективность результатов проверки, так как наличие заранее определенного эталона ответа (ответов) каждый раз приводит к одному и тому же результату;
- повышение эффективности контролирующей деятельности со стороны преподавателя за счет увеличения её частоты и регулярности;
- возможность автоматизации проверки знаний учащихся, в том числе с использованием компьютерных технологий;
- возможность использования в системах дистанционного и открытого образования.

Применение тестов в традиционных формах обучения не исключает использование других видов контроля знаний, тем более что тестовый контроль имеет и свои недостатки, в частности, не способствует развитию навыков делового общения, коммуникабельности, аргументации своей позиции (в устной или письменной форме). По поводу последнего стоит заметить, что развитие навыков аргументации следует производить не во время испытаний (нормативное время экзамена студента составляет порядка 20 минут), а во время обучения. Контроль же уровня развития вышеуказанных навыков вряд ли следует производить преподавателям-предметникам, не являющимся специалистами по данной проблематике.

Все вышеуказанные преимущества тестового контроля могут быть достигнуты лишь при использовании теории педагогических тестов, которая сложилась на стыке диагностики, педагогики, психологии и математической статистики. Проведение тестирования может иметь следующие цели:

- определение степени соответствия результатов обучения декларированным целям (как по выборке, так и по отдельным объектам обследования);

- сопоставление индивидуальных показателей уровня обученности с показателями для выборки (генеральной совокупности);
- аттестация (вступительные или выпускные экзамены);
- определение проблемных областей дисциплины (разделы, которые оказались недостаточно хорошо изложены в лекционном курсе, слабо закреплены на упражнениях, по которым отсутствуют необходимые методические указания и т.д.);
- создание базы тестовых заданий, обладающих хорошими диагностическими свойствами;
- выработка критериев сопоставления результатов тестирования с принятой в вузах балльной системой оценок.

Целью, определяющей сущность тестирования, является диагностирование, т.е. установление диагноза (состояния, характеризующего уровень обученности) тестируемого. При этом педагогическое тестирование представляет собой процедуру получения статистической информации, на основании которой могут быть приняты или отвергнуты гипотезы о достижении тестируемым определенного уровня владения учебным материалом. Статистическая информация должна быть диагностичной (информативной, контролепригодной) и значимой для решения поставленной задачи.

Состояние объекта диагностирования описывается совокупностью (множеством) определяющих его параметров (признаков). При этом множество определяющих параметров может быть различным, в первую очередь, в связи с самой задачей распознавания. Распознавание состояния объекта диагностики - отнесение его к одному из возможных классов (диагнозов). Число диагнозов зависит от особенностей задачи и целей исследования. Наиболее часто предлагается провести выбор одного из двух диагнозов, например, "сильный студент" - "слабый студент" (дифференциальная диагностика или дихотомия). В других случаях необходимо более подробно охарактеризовать ситуацию, установить балльную оценку или уровень владения учебным материалом. В большинстве

случаев диагностики (классы) устанавливаются заранее и в этих условиях задачу распознавания часто называют задачей классификации.

Педагогической практикой выработаны следующие этапы разработки тестов:

- составление спецификации (графа) учебных элементов по данной дисциплине (разделу, теме);
- определение объектов контроля и выделение учебных элементов, по которым будут составлены тесты;
- составление тестов в первом (рабочем) варианте;
- экспертно-редакционная проверка и корректировка тестов;
- экспериментальная проверка;
- анализ результатов экспериментальной проверки и корректировка заданий и эталонов;
- определение паспортных выходных характеристик тестов;
- публикация теста (в случае сетевых решений - это установка его на Web-сервере).

3. Таксономические показатели диагностирования

Как отмечалось выше, цель диагностирования – сопоставление реального состояния объекта диагностики с некоторым эталоном, который может (должен) быть определён и описан в системе диагностических признаков до проведения тестирования и фактически должен являться целью обучения.

Успешность достижения цели (конечного состояния) прямо зависит от знания основных промежуточных состояний и их последовательности. Следовательно, программа управления качеством должна обеспечить прохождение формируемых видов ментальной (познавательной) деятельности обучаемых через основные качественные этапы процесса обучения. При этом необходимо четкое представление о функциональной роли каждого этапа в процессе усвое-

ния знаний. А для этого, в свою очередь, необходим анализ содержательной информации, предлагаемой обучаемым для усвоения, то есть количественное измерение и качественное описание содержания учебного материала.

При реализации этой достаточно очевидной идеи обычно сталкиваются с проблемой определения способа однозначного и определённого перевода результатов обучения на язык действий, выполняемых в ходе тестирования. Эта проблема решается двумя основными путями:

- построением четкой системы целей, внутри которой выделены их категории и последовательные уровни (иерархия), - такие системы получили название педагогических таксономий (от греч. taxis - расположение по порядку и nomos - закон);
- построением системы диагностических признаков, позволяющих адекватно и достоверно диагностировать достижение выделенного таксономического уровня.

Таксономия обозначает такую классификацию и систематизацию объектов, которая построена на основе их естественной взаимосвязи и использует для описания объектов категории, расположенные последовательно, по нарастающей сложности, т.е. по иерархии.

Впервые задача построения такой схемы педагогических целей была поставлена в США. После окончания второй мировой войны группа педагогов и психологов, входивших в Комитет по приему экзаменов в колледже под руководством Б. Блума, провела исследование с тем, чтобы разработать общие способы и правила четкой формулировки и упорядочения педагогических целей. В 1956 г. вышла в свет первая часть «Таксономии», содержащая описание целей в познавательной (когнитивной) области. Эта система целей получила широкую международную известность. Ее используют при планировании обучения и оценке его результатов, она служит надежным инструментом при опытной проверке новых курсов. В последующие годы Д. Кратволом и другими учеными была создана вторая часть «Таксономии» (в аффективной или эмоционально-

ценностной области). В дальнейшем продолжалась разработка и уточнение таксономий педагогических целей, были созданы несколько вариантов когнитивных, психомоторных, а также аффективная и операционно-деятельностная таксономии. Когнитивные цели легче объективизировать, представить в виде образцов деятельности. Поэтому, хотя таксономия аффективных целей применяется в педагогической диагностике, собственно технологические разработки ведутся для целей другого типа - прежде всего когнитивных.

В когнитивную (познавательную) область входят цели от запоминания и воспроизведения изученного материала до решения проблем, в ходе которого необходимо переосмыслить имеющиеся знания, строить их новые сочетания с предварительно изученными идеями, методами, процедурами (способами действий), включая создание нового. Сюда относится большинство целей обучения, выдвигаемых в программах, учебниках и решаемых в повседневной педагогической практике. Использование четкой, упорядоченной, иерархической классификации целей важно, прежде всего, по следующим причинам:

- концентрация усилия на главном; пользуясь таксономией, преподаватель не только выделяет и конструирует цели, но и упорядочивает их, определяя первоочередные задачи, порядок и перспективы дальнейшей работы;
- ясность и гласность в совместной работе преподавателя и студентов; конкретные учебные цели дают преподавателю возможность разъяснить студентам ориентиры в их общей учебной работе, обсудить их, сделать ясными для понимания;
- создание эталонов оценки результатов обучения; обращение к четким формулировкам целей, которые выражены через результаты деятельности, поддается более надежной и объективной оценке.

Таксономические модели относятся к так называемым классификационным моделям, которые подразделяются на модели логических классификаций (подобно библиотечной классификации УДК) и модели численной таксономии (кластер-анализ, метод главных компонент и др.). Их характерной особенностью

стью является то, что они описывают наблюдаемые явления вне причинно-следственных связей. Последнее обстоятельство и делает их наиболее удобными для использования: предпосылки предельно просты, наблюдаемые переменные не делятся на зависимые и независимые, они лишь специфицируются (перечисляются) и для них задается некая метрика. Процедура классифицирования при этом ограничивается поиском иерархии таксонов в императивно заданной метрике.

На сегодняшний день существуют большое число таксономий. Рассмотрим основные категории учебных целей и примеры обобщённых типов учебной деятельности, выраженных через деятельность обучаемых для некоторых существующих в настоящее время таксономий.

В.П. Беспалько /5/ выделяет следующие уровни: узнавание и различение, репродуктивный, продуктивный, творческий или в более поздней редакции /5/: ученический, алгоритмический, эвристический, творческий.

Система Б. Блума в когнитивной области также является довольно распространенной. В ней выделяются следующие уровни: узнавание и различение, понимание, применение, анализ, синтез, оценка.

В.П. Симонов /76/ выделяет следующие пять уровней или степеней обученности: различение, запоминание, понимание, элементарные умение и навыки, перенос.

Таксономические уровни с одной стороны выступают в качестве когнитивных целей обучения, с другой – являются объектами диагностики. Формулировка тестовых заданий должна соответствовать диагностируемому уровню владения учебным материалом. Форма и содержание диагностических признаков в виде тестовых заданий должно соответствовать возможностям обучаемого (тестируемого) на данном уровне владения учебным материалом, поэтому раскрытие этих категорий должно быть рассмотрено более детально. Дадим

краткую обобщённую характеристику указанных выше таксономических уровней и рекомендации по формированию тестовых заданий.

Узнавание и различение - уровень знакомства с учебными элементами, принятие решения и осуществление действий базируется на повторном восприятии ранее изучавшейся информации. Это процесс восстановления запечатлённого в памяти мысленного образа некоторых объектов в результате повторных воздействий этих объектов на человека. Процесс осуществляется тем быстрее и адекватнее, чем существеннее и ярче признаки, по которым сравнивается объект, воздействующий на человека в данный момент, и мысленный образ этого объекта, хранящийся в памяти. Установление различия и сходства (различение), также является одним из первых признаков познания. Эти категории устанавливают признаки того, чем один объект отличается от другого, а также нечто самостоятельное, относительно устойчивое, что у объектов совпадает, объединяет их в классификационные группы. Чтобы узнать объект, надо определить то, чем он отличается от других объектов и чем он сходен с другими объектами. Основными существенными мыслительными операциями являются альтернативные суждения типа «Да/Нет». «Самая низкая степень владения знаниями – возможность только узнавания в дальнейшем: человек может лишь констатировать, что знания были получены раньше, но не воспроизвести их тем или иным образом» /55 т.2, с.119/. Эта категория обозначает владение на уровне «знакомости» понятийным аппаратом, глоссарием или тезаурусом (словарь специальных терминов): обучаемый узнаёт и различает употребляемые термины, его суждения - односторонние и категоричные, речь немногословна. Достаточно хорошей формой тестовых заданий закрытого типа здесь могут быть задания по выбору из двух или более альтернатив: «Укажите, используется ли в таких-то целях такой-то объект», «Укажите, является ли такое-то свойство атрибутом такого-то объекта», «Укажите, может ли такой-то объект совершать такие-то действия», «Укажите, можно ли с таким-то объектом совершать такие-то действия», «Укажите, характерен ли такой-то признак для такого-то объек-

та» с эталонами ответов «Да/Нет». Переходить к более сложным заданиям следует лишь после того, как возможности вышеприведенных заданий с односложными ответами будут исчерпаны и ими будет охвачено большинство учебных элементов нашей предметной области. Более сложными заданиями для данного уровня усвоения будут те, в которых содержится уже более двух эталонов ответов: «Укажите, может ли иметь место то-то и то-то» с эталонами «всегда, никогда, иногда», «справедливо утверждение А, справедливо утверждение В, справедливы оба утверждения А и В, не справедливо ни одно из утверждений А или В» и т.д. В заданиях этого типа, отдавая предпочтение одному эталону из нескольких (трёх - пяти), мы одновременно сопоставляем большее число понятий (учебных элементов). Эталоны, в том числе и неверные, должны быть подобраны таким образом, чтобы их правильность или неправильность не была очевидной, т.е. неправильный эталон ответа должен становиться правильным в несколько иной формулировке тестового задания. Здесь можно также формулировать и задания с отрицанием: «Укажите, каким свойством из нижеперечисленных, не может обладать такой-то объект», «Укажите ситуацию из нижеперечисленных, в которой такой-то объект не может применяться по назначению» и т.д.

Запоминание - на этой ступени обученности значительно увеличивается объём изученной и диагностируемой информации. Н.И. Кондаков приводит следующее определение этого термина: «Запоминание – образование временных связей между группами нервных клеток головного мозга, способных по истечении времени вновь актуализироваться. В основе запоминания лежит смысловое содержание отображённых в сознании объектов и связь с активной деятельностью человека, с решением им практических задач, требующих воспроизведения образов» /35/. Педагогам дошкольного образования часто приходится наблюдать малышей с упоением читающих выученные стишки, несмотря на то, что значение многих слов в заученном тексте им пока ещё не понятно. Этот этап своего развития им, так или иначе, необходимо пройти. Не стоит пытаться

«обмануть природу» и требовать на этом этапе запоминания через понимание и объяснения. Несостоятельность такого подхода выявится очень скоро: у детей можно отбить охоту к обучению, они отстанут в развитии. Преподаватели, «жалая» студента, «разгружая» его от этой мыслительной деятельности и не контролируя её, пожнут горькие плоды разочарования в конце цикла обучения. Здесь основной метод обучения - так называемая «зубрёжка», основной метод диагностики - контроль воспроизведения информации в той или иной форме. Умение «сообразить» от студента на данном уровне пока не требуется. В ответах оценивается не способность к рассуждениям, а широта фактологического описания. Вспомним модные телевикторины «Как стать миллионером», «Слабое звено» или «Своя игра». «Сообразить» в них практически невозможно, зато можно «вспомнить» или проявить интуицию (т.е. извлечь нужную информацию из подсознания, где она хранилась как бы «временно забытой»). Таким требованиям в наилучшем отношении удовлетворяют так называемые «конструктивные» задания или задания на ввод информации в строку ответа: «Напишите название объекта, имеющего такое-то определение», «Напишите название объекта, имеющего такие-то свойства», «Напишите название объекта, способного выполнять такие-то действия», «Напишите название свойства, общего для такого-то и такого-то объектов», «Напишите название объекта, показанного на рисунке» и т.д. В конструктивных тестовых заданиях нежелательны, а часто и просто невозможны задания с отрицанием. Зато, здесь широко могут использоваться и задания по выбору: «Укажите все свойства из нижеперечисленных, которые присущи такому-то объекту», «Укажите недостатки из нижеперечисленных, характерные для такого-то объекта», «Укажите из нижеперечисленных области возможного использования такого-то объекта». Задания на упорядочение: «Расположите такие-то величины по убыванию», «Приведите правильный порядок проведения такой-то лабораторной работы» и т.д. Количество учебных элементов, к которым студент вынужден обращаться, здесь столь велико (если это не так, то ценность этого этапа существенно девальвируется!), что в попыт-

ках удержать их в памяти студент самостоятельно будет пытаться их как-то упорядочивать, классифицировать, систематизировать. При этом образуются структурно-логические и причинно-следственные связи учебных элементов. Для выявления уровня представлений о таких взаимосвязях возможны следующие задания: «Укажите, может ли событие А являться следствием события В», «Укажите, всегда ли наступление события А влечёт за собой событие В» и т.д. Особенно эффективны такие задания при составлении тестов по многофакторным лабораторным исследованиям (лабораторным работам). Необходимость запомнить, интерпретировать и объяснить для себя наблюдаемые явления побуждает обращаться к обобщающим формульным представлениям. Формулы запоминаются, если они «выстраданы» размышлениями на тему «А почему это так?» В результате появляются предпосылки того, что тестируемый воспроизводит употребляемые термины, помнит конкретные факты, помнит методы и процедуры, воспроизводит основные понятия, воспроизводит правила и принципы. Нет другого пути эффективного запоминания большего объёма информации, кроме осознания взаимосвязи суждений, отношений, преобразований. Знание основывается на фактах и является методом усвоения информации или владения теорией. Эта категория знания включает: факты, терминологию, способы представления понятий и явлений, тенденции развития, хронологию, последовательность событий, классификацию, критерии, методологию, общие и абстрактные понятия, теорию, принципы обобщения. Эта категория обозначает запоминание и воспроизведение изученного материала. Речь может идти о различных видах содержания - от конкретных фактов до целостных теорий. Общая черта этой категории - припоминание соответствующих сведений. Здесь, как нигде более, необходимо чувство меры. «Зубрёжка, школярство, - отмечал В.А. Сухомлинский, - страшные враги школы. В Павлыше учителя точно разграничивают, что надо запомнить на всю жизнь, а что надо только понять и усвоить без зубрёжки – в упражнениях». И ещё: «Почему мои дети так легко запоминали буквы, научились читать и писать? Потому, что перед ними не ставилась

цель сделать это. Потому, что каждая буква для ребёнка была воплощением яркого образа, вызвавшего чувство восхищения. Если бы я каждый день давал малышам-дошкольникам «порцию знаний» - показывал букву и требовал запомнить её – ничего бы не вышло. Это, конечно, не означает, что надо прятать от ребёнка цель. Учить следует так, чтобы дети не думали о цели, - это облегчит умственный труд»/68/. Впрочем, во многом здесь идёт речь о методологии обучения, а не диагностирования.

Репродуктивный уровень - характеризует виды деятельности без опоры на помощь или подсказку, воссоздание, воспроизведение находящихся в памяти образов. (От лат. re – приставка, означающая возобновление или повторение действия, productio – производство, произведение). Это может быть воспроизведение по памяти учебного материала, решение типовых задач. Обучаемый использует понятия и принципы в стандартных ситуациях, применяет законы, теории в конкретных практических случаях, демонстрирует правильное применение метода или процедуры.

Понимание - иногда этот уровень понимают, как достижение студентом способности к рассуждениям по аналогии с чем-то уже известным /65/. Для того, чтобы учащийся что-либо понял, мы должны уметь объяснить ему это. А для этого мы сами должны это очень хорошо понимать. Проблема определения понятия понимания не проста. Д. Данин приводит диалог В. Гейзенберга с Н. Бором:

«Гейзенберг: - Если внутренняя структура атома так недоступна наглядному описанию, как вы говорите, и если взаправду у нас для неё нет языка, как мы можем надеяться хоть когда-нибудь достигнуть понимания атома?»

После минутного колебания Бор сказал:

- Думаю, мы всё же достигнем этого. Но по дороге нам придётся узнать, что реально означает само слово ПОНИМАНИЕ...» /13/

Понимание имеет две стороны – субъективную и объективную. Субъективная сторона заключается в том, что понимание для человека всегда эмоцио-

нальное переживание, по большей части положительное, связанное с преодолением некоторого барьера, в избавлении от неопределённости, неясности. Объективная сторона понимания состоит во включении нового знания в систему имеющихся знаний, что предполагает наличие следующих аспектов: определение причинно-следственных связей, связывающих данное знание с имеющимися; способность к обобщениям, т.е. представлениям о том, что данное знание является более общим или частным знанием имеющихся знаний; изучение происхождения знания, этапов его развития. Понимание интерпретируется, как способность продуцировать умозаключения. Под умозаключением Н.И. Кондаков понимает «формы мышления или логического действия, в результате которого из одного или нескольких известных нам и определённым образом связанных суждений получается новое суждение, в котором содержится новое знание» /30/. Например, из двух посылок «Все жидкости упруги» и «Вода – жидкость» следует вывод «Вода упруга». Ни из одного из исходных суждений, взятых в отдельности, это новое знание почерпнуть невозможно.

Показателем способности понимать изученное может служить преобразование (транспонирование) материала из одной формы выражения в другую, «перевод» его с одного «языка» на другой (например, из словесной формы - в математическую). На этом уровне обучаемый понимает факты, правила и принципы, интерпретирует словесный материал, интерпретирует схемы, графики, диаграммы, преобразует словесный материал в математические выражения, предположительно описывает будущие последствия, вытекающие из имеющихся данных. Д.П. Горский отмечает: «Знаковое выражение имеет смысл, значение, является понимаемым, если по отношению к нему могут быть сформулированы соответствующие правила введения и исключения» /14/. В качестве показателя понимания может также выступать интерпретация материала учеником (объяснение, краткое изложение) или же предположение о дальнейшем ходе явлений, событий (предсказание последствий, результатов). Признаком понимания является способность индивидуума критически оценить некоторые

ранее неизвестные ему утверждения из данной области знаний, способность правильно предсказать ещё неизвестные реальные факты, причём не только качественно, но и количественно. Такие учебные результаты превосходят простое запоминание материала. Понимание рассматривается как знание, которое позволяет вступить в коммуникацию и пользоваться имеющейся информацией. Понимание - более высокая ступень усвоения, чем простое знание. Оно обнаруживается в интерпретации смысла текста, в умении схватить основной смысл раздела, перевести математические символы в вербальные, использовать полученные данные для определения следствий и т.д. На данном уровне при обучении запоминание учебного материала происходит через рассуждения. Материал хорошо запоминается, если он укладывается в привычную схему или, напротив, почему-либо противоречит ей. И тогда приходит потребность корректировки уже самой схемы (ещё раз целесообразно напомнить, что схемы формируются в сознании студента на предыдущем этапе, а не сообщаются ему в готовом виде на данном этапе). Для студента становится уже неприятной «зубрёжка», необходимость выучить что-либо, не понимая сути. Одна из важнейших целей познания в естественных и технических науках является выход на практику, на предметные действия, способность применить знание к действиям в мире вещей является необходимым признаком действительного понимания. Тестовые задания должны отражать возросший уровень владения учебным материалом. Их составление становится ещё более трудным занятием. Здесь возможен любой вид тестовых заданий, но вид деятельности при их выполнении должен предусматривать рассуждения, логические умозаключения. Например: «Закон Ома однозначно связывает между собой напряжение, ток и сопротивление электрической цепи. Величина тока определяется величиной сопротивления. Укажите, справедливо ли обратное: величина сопротивления определяется проходящим через него током», «Укажите, что следует исключить из приведенного ниже списка: декабрь, июнь, январь, февраль» (без указания критерия!), «Проведите аналогию: Плутон - Харон, Земля - ...» (Харон - спутник Плутона!) и

т.д. И ещё раз необходимо обратить внимание на требование однозначности ответов на задания тестов. Для составления заданий на этом уровне полезно на листе бумаги нанести граф структурно-логических связей учебных элементов изучаемой предметной области и отметить на нём наиболее устойчивые и значимые связи.

Продуктивный уровень - способность решать нетиповые задачи на применение знаний в условиях, близких к реальной практической деятельности. Знания трансформируются в знания-умения. Продуктивными умозаключениями Л.В. Рутковский (1859-1920), известный русский логик, называет такие, когда «усмотрев в предмете наблюдения какой-либо признак или вообще получив какое-либо определение данного предмета, мы без дальнейшего опыта приписываем этому предмету новое определение в силу того, что ему присуще данное в основном определении суждение» /30/. Предложенный Л.В. Рутковским термин «продукция» имеет тот же латинский корень «duc», на основе которого существуют термины традукция, индукция или дедукция, служащие для обозначения типов умозаключений. Приставка «про», означающее поступательное движение вперёд, характеризует специфический оттенок нового умозаключения, при котором наша мысль как бы прогрессирует, идёт вперёд в деле изучения предмета со стороны приложимых к нему определений, переходя от познанного уже определения к ещё непознанному. Эта категория обозначает умение использовать изученный материал в конкретных условиях и новых ситуациях. Сюда входит применение приемов, методов, понятий, законов, принципов, теорий. Соответствующие результаты обучения требуют более высокого уровня владения материалом, чем понимание. Применение рассматривается, как умение применять информацию (правила, методы, общие понятия) в новой ситуации без подсказки. Получение численных оценок некоторых величин путём решения типовых задач - одно (но далеко не единственное!) из возможных приложений полученных знаний. Иногда это выбор единственно верного алгоритма действий, иногда - выбор предпочтений при учёте большого числа про-

тиводействующих факторов, но чаще всё-таки - решение задач. При их формулировании необходимо указывать размерность величины, являющейся ответом. Полезно также вместо условных обозначений величин, входящих в формулы, давать их полное наименование: «Определите величину напряжений изгиба (в МПа) балки при следующих исходных данных: изгибающий момент - такой-то, диаметр сечения балки - такой-то». При этом достигается дополнительная компиляция символьных выражений, которые представляют собой формулы, в смысловые понятия и обратно.

Ученический уровень - характеризуется деятельностью по узнаванию с подсказками в виде эталонов готовых ответов; сюда же относится знание классификационных признаков, свойств и способность осуществлять на основе этой информации различение объектов изучения.

Алгоритмический уровень – готовность к выполнению алгоритмических действий по решению типовых задач, т.е. задач, требующих применения готовых (не преобразованных) алгоритмов. На этом же уровне требуется воспроизведение по памяти объектов изучения, их свойств и классификационных признаков.

Творческий (эвристический) уровень - характеризует умение ориентироваться, принимать решения в новых, проблемных ситуациях, когда задана цель каких-либо действий, но отсутствует готовый алгоритм. Для достижения результата требуется конструирование алгоритма. В решении таких проблем должна проявляться научная эрудиция, умение выдвигать и проверять обоснованные гипотезы, применять соответствующие методики исследования, прогнозировать возможные исходы и трудности. Цель обучения задаётся в предельно общей форме. Требуется установить возможность её осуществления, проанализировать ситуацию, предложить способ решения на основе преобразования известных алгоритмов действий. Теория решения изобретательских задач (т.н. ТРИЗ), методы поискового проектирования, эвристические методы нахождения решений. Можно ещё долго продолжать этот список, т.к. множество пре-

подавателей избрали сферой своей профессиональной деятельности обучение методологии совершения открытий или, по крайней мере, развитию творческих потенций. Тест по методологии творчества безусловно возможен, однако у автора данной работы нет сведений об успешном использовании тестирования для диагностики таланта или хотя бы творческих способностей. Частично проблема заключается в том, что при тестировании в качестве обязательного требования выступает наличие эталона решения. Для таксономического уровня, характеризующего творчество, требования простоты и однозначности осуществить не слишком просто. Проблемы здесь, вероятно, носят временный характер, однако оставим вопрос о диагностировании творчества до лучших времён и вне рамок данной работы.

Анализ - обучаемый выделяет скрытые (неявные) предположения, видит ошибки и упущения в логике рассуждения, проводит различия между фактами и следствиями, оценивает значимость данных. Эта категория обозначает умение разбить материал на составляющие так, чтобы ясно выступала его структура. Сюда относятся вычленение частей целого, выявление взаимосвязей между ними, осознание принципов организации целого. Учебные результаты характеризуются при этом более высоким интеллектуальным уровнем, чем понимание и применение, поскольку требуют осознания, как содержания учебного материала, так и его внутреннего строения. Анализ рассматривается как знание, позволяющее делить информацию на части и устанавливать взаимозависимость между ними. Анализ включает умение определить элементы, составляющие данное содержание, выяснить зависимость между отдельными частями и элементами.

Синтез – обучаемый на этом уровне пишет небольшое творческое сочинение, предлагает план проведения эксперимента, использует знания из разных областей, чтобы составить план решения той или иной проблемы. Эта категория обозначает также умение комбинировать элементы, чтобы получить целое, обладающее новизной. Таким новым продуктом может быть сообщение (вы-

ступление, доклад), план действий или совокупность обобщенных связей (схемы для упорядочения имеющихся сведений). Соответствующие учебные результаты предполагают деятельность творческого характера с акцентом на создание новых схем и структур. Синтез - рассматривается как знание, позволяющее реорганизовать информацию из разных источников и на этой основе создать новый образец. Синтез предполагает творческую переработку информации, в результате чего вырабатывается общий план действия, создается новое целое, разрабатывается информация, объясняющая явление или событие.

Оценка - обучаемый оценивает логику построения материала в виде письменного текста, оценивает соответствие вывода имеющимся данным, оценивает значимость того или иного продукта деятельности; исходя из внутренних критериев, оценивает значимость того или иного результата деятельности, исходя из внешних критериев. Эта категория обозначает умение оценивать значение того или иного материала (утверждения, художественного произведения, исследовательских данных) для конкретной цели. Суждения обучаемого должны основываться на четких критериях. Критерии могут быть как внутренними (структурными, логическими), так и внешними (соответствие намеченной цели). Критерии могут определяться самим обучаемым или же задаваться ему извне (например, преподавателем). Данная категория предполагает достижение учебных результатов по всем предшествующим категориям плюс оценочные суждения, основанные на обозначенных критериях. Оценка позволяет судить о ценности какой-либо идеи, метода, материала. Это новый шаг в овладении знаниями, характеризующийся проникновением в суть предмета, явления.

В ряде работ, содержащих таксономические исследования, приводится попытка сопоставить степени обученности не только с качественной, но и количественной стороны. При допущении линейного закона по указанным уровням «Педагогическая энциклопедия» предлагает «багаж знаний» соотносить как последовательный ряд нечётных чисел, т.е. $1 : 3 : 5 : 7 : 9$, что в процентах соответствует 4%, 12%, 20%, 28% и 36%. На первый взгляд это может показаться

довольно привлекательным, так как даже без выхода на уровень творчества (пятый, наиболее высокий уровень) положительный результат от учебной деятельности составит 64% или около двух третей. Сомнение же вызывает результат сопоставления веса отдельных уровней, например, 4-го и 5-го уровня. Их различие оценивается как 9/7 или 1,29, то есть при достижении умения (навыка) решения типовых задач переход к «творчеству» возможен в результате прироста «багажа знаний» порядка 30%.

Другой, более обоснованный на наш взгляд, подход предлагается классиками диалектического материализма. Так Ф. Энгельс отмечал /77/: «...наука движется вперёд пропорционально массе знаний, унаследованных ею от предшествующего поколения, следовательно, при самых обыкновенных условиях она ... растёт в геометрической прогрессии». Применяя закон геометрической прогрессии для той же пятиуровневой таксономии, мы получим следующие показатели наполнения различных уровней: 1%, 2,8%, 8%, 23% и 65%, что в сумме составит также около 100%. Здесь уже достижение даже 4-го уровня обученности выглядит не слишком значимым. Ещё более значительная разница будет наблюдаться при исключении 5-го уровня: 1%, 4%, 18% и 77% (оценивание без выхода на уровень «творчество»).

Данные прохождения ряда тестов обработаны автором методом математического моделирования, т.е. интерпретировались по-разному в зависимости от сравниваемых таксономий. Анализ результатов позволяет сделать выводы о том, что для сильных студентов (прошедших тест с принятым нормативом) выбор таксономии не является существенным фактором, а для гетерогенных тестов (тестов с неоднородными по диагностируемыми уровням усвоения и видам деятельности тестовыми заданиями) выбор таксономии оказывается существенным фактором.

Для уменьшения неопределённости оценивания, связанной с выбором таксономии желательно, чтобы в одном тесте встречались тестовые задания не более чем 2-х уровней (в крайнем случае - 3-х уровней).

4. Статистические параметры тестовых заданий

Статистический анализ тестовых заданий проводят, основываясь на статистике их выполнения студентами. Одной из основных задач разработчика теста является составление таких заданий, которые были бы способны отделить сильных студентов от слабых на учебном материале данной дисциплины и на диагностируемом уровне владения этим учебным материалом. Попробуем оценить тестовое задание не с точки зрения его содержательной части, а с позиций успешности или не успешности его выполнения сильными (успешно прошедшими тест в целом) и слабыми (не выполнившими критериальные требования) студентами. После ряда попыток прохождения теста некоторой группой студентов (так называемое пробное или пилотное тестирование) каждое тестовое задание приобретает некоторую статистику. Для получения статистических показателей необходимо рассмотреть различные сочетания событий, связанных с успешностью или неуспешностью выполнения как отдельного тестового задания (каждого из заданий теста), так и теста в целом. Эти события зависимые и совместные, следовательно, к ним может быть применён байесовский подход, в рамках которого возможны следующие ситуации: 1) тестируемый выполнил данное тестовое задание и успешно сдал тест (сильный студент); 2) тестируемый не выполнил данное тестовое задание, но успешно сдал тест (сильный студент); 3) тестируемый выполнил данное тестовое задание, но не прошёл тест (слабый студент); 4) тестируемый не выполнил данное тестовое задание и не прошёл тест в целом (слабый студент).

Обозначим N_1 - количество сильных студентов, выполнивших данное тестовое задание, N_2 - количество сильных студентов, не выполнивших данное тестовое задание; N_3 - количество слабых студентов, выполнивших данное тестовое задание; N_4 - количество слабых студентов, не выполнивших данное тестовое задание.

$N = N_1 + N_2 + N_3 + N_4$ - суммарное количество попыток выполнения данного тестового задания.

Простейшая диагностическая характеристика тестового задания, которая может быть получена пока ещё вне рамок байесовского подхода, называется коэффициентом решаемости и является статистической оценкой вероятности его выполнения произвольным (случайным, т.е. он может оказаться как сильным, так и слабым) студентом. Эта оценка, определяемая как отношение числа успешных выполнений данного тестового задания к общему числу попыток

$$K_p = \frac{N_1 + N_3}{N}, \quad (1)$$

не позволяет отделить сильных студентов от слабых, так как задание успешно выполнялось как сильными студентами, так и слабыми, может быть, лишь с различной степенью успешности.

При принятии гипотезы о том, что выборка студентов, приступающих к тестированию, относится к той же генеральной совокупности, для которой получены ранее значения коэффициентов решаемости, можно прогнозировать успешность выполнения ими тех или иных тестовых заданий и по этому критерию проводить их (тестовых заданий) предварительный отбор.

Тестовым заданием с высокими диагностическими свойствами является такое, которое успешно выполняется сильным студентом и не выполняется слабым. Именно это свойство тестовых заданий используется для разделения студентов на сильных и слабых в зависимости от успешности выполнения ими (студентами) того или иного тестового задания. Таким образом, коэффициент решаемости в определённой степени характеризует соотношение сильных и слабых студентов в контрольной выборке и может служить статистической оценкой значения априорной вероятности диагноза сильного студента.

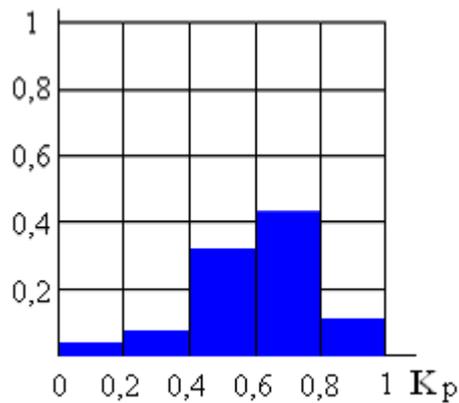


Рис. 2. Гистограмма распределения значений коэффициента решаемости тестовых заданий

На рис. 2 показана одна из типовых диаграмм (здесь и далее приводимые гистограммы являются результатом обработки статистики тестовых заданий, выполняемой с помощью программы статистической обработки результатов компьютерного тестирования) после проведения тестирования с использованием тестирующей компьютерной программы. Хорошими диагностическими свойствами обладают тесты, в которых коэффициент решаемости тестовых заданий распределён по нормальному закону.

Коэффициент решаемости - это достаточно интересный статистический показатель. Если решаемость тех же тестовых заданий для новой выборки тестируемых окажется выше, можно утверждать, что студенты из этой группы тестируемых объективно являются более сильными. Решаемость тестовых заданий также является критерием оценки их качества с позиций студентов как объектов образовательного процесса, так как каждый из них субъективно заинтересован в более успешном прохождении теста и в том, чтобы по результатам тестирования его смогли отнести к числу так называемых «сильных» студентов.

Рассмотрим произведение двух зависимых событий: невыполнение тестового задания \bar{z}_i и нахождение в диагнозе D_1 (сильный студент). Вероятность такого сложного события определяется по формуле

$$P(\bar{z}_i \cdot D_1) = P(\bar{z}_i | D_1)P(D_1) = P(D_1 | \bar{z}_i)P(\bar{z}_i). \quad (2)$$

Отсюда

$$P(D_1 | \bar{z}_i) = \frac{P(D_1)P(\bar{z}_i | D_1)}{P(\bar{z}_i)}. \quad (3)$$

Полная вероятность неуспешного выполнения тестового задания z_i (знаменатель формулы (3)) находится суперпозицией по всем диагнозам. При этом

$$P(D_1 | \bar{z}_i) = \frac{P(D_1)P(\bar{z}_i | D_1)}{P(D_1)P(\bar{z}_i | D_1) + P(D_2)P(\bar{z}_i | D_2)}. \quad (4)$$

Ошибка 1-го рода или коэффициент трудности K_T тестового задания может быть определён как статистическая оценка вероятности невыполнения тестового задания сильным студентом, определяемой по формуле (4). Пользуясь зависимостью (4), приходим к следующей, более простой формуле:

$$K_T = \frac{N_2}{N_2 + N_4}. \quad (5)$$

То есть коэффициент трудности K_T - это доля сильных студентов среди всех, не выполнивших данное тестовое задание. По Байесу коэффициент трудности - это условная вероятность события, заключающегося в том, что студент, не выполнив данное тестовое задание, в итоге успешно пройдёт весь тест.

При приёме экзаменов каждому преподавателю за свою достаточно продолжительную педагогическую практику вольно или невольно случалось совершать действия, когда объективно сильный студент оценивался ниже своих потенциальных возможностей. В этом случае говорят, что преподаватель «сре-

зал» студента на экзамене. Для этих целей у преподавателя обычно припасён арсенал так называемых «каверзных» вопросов. Такие вопросы формулируются на основе использования каких-то нюансов, которые могут ускользнуть от внимания готовящегося к экзамену, либо студенту предлагается продемонстрировать владение учебным материалом в необычной постановке, когда его здравый смысл должен подсказать ему неверный ответ. Годятся ли такие вопросы в качестве тестовых заданий или могут ли они иметь место в тесте? Во-первых, оценивание студента при тестировании должно вестись не по нюансам, а по базовым учебным элементам, а во-вторых, тест - это модель обычного экзамена, и, в каком-то понимании, ничто, свойственное ситуации реального экзамена, ему не чуждо. То есть в статистическом плане обязательно возникнут случаи, когда какое-либо тестовое задание окажется не выполненным студентом, который успешно прошел тест в целом, выполнив установленные критериальные требования, и с полным основанием может быть отнесён к сильным студентам. Такое тестовое задание в этом случае сыграет роль «вредного» преподавателя и будет в статистическом смысле не совсем хорошим. Для любого теста желательна такая статистика, при которой доля заданий с большими значениями коэффициента K_T или статистических ошибок 1-го рода невелика. Гистограмма распределения коэффициента трудности для реального теста приведена на рисунке 3.

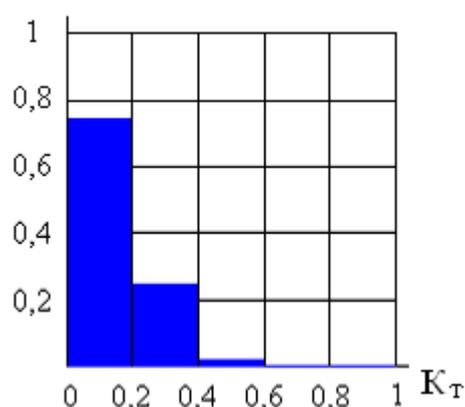


Рис. 3. Гистограмма распределения значений ошибок 1-го рода (коэффициента трудности тестового задания)

Как это видно из рассмотрения гистограммы, в данном тесте практически отсутствуют задания с высокими значениями коэффициента трудности ($K_T > 0,5$).

Преподаватель на обычном экзамене может быть не только «вредным», но и «добрым», т.е. бросающим «утопающему» студенту соломинку или спасательный круг в виде лёгких или так называемых «наводящих» вопросов. Не касаясь этической стороны вопроса (речь здесь может идти о профессиональной этике преподавателя), эта реальность экзамена также нашла отражение в соответствующем статистическом показателе тестового задания, названным автором коэффициентом интуитивности, так как, отвечая на задание, студент больше полагается на свою интуицию, чем на знания. Этот показатель наряду с коэффициентом трудности также характеризует тестовое задание не с лучшей стороны.

Рассмотрим произведение двух зависимых событий: выполнение тестового задания z_i и нахождение в диагнозе D_2 (слабый студент). Вероятность такого сложного события аналогично (2) определяется по формуле

$$P(z_i \cdot D_2) = P(z_i | D_2)P(D_2) = P(D_2 | z_i)P(z_i). \quad (6)$$

Отсюда

$$P(D_2 | z_i) = \frac{P(D_2)P(z_i | D_2)}{P(z_i)}. \quad (7)$$

Полная вероятность успешного выполнения тестового задания z_i , определяемая формулой (7) может быть определена суперпозицией по всем диагнозам. При этом

$$P(D_2 | z_i) = \frac{P(D_2)P(z_i | D_2)}{P(D_1)P(z_i | D_1) + P(D_2)P(z_i | D_2)}. \quad (8)$$

Ошибка 2-го рода или коэффициент интуитивности $K_{и}$ тестового задания может быть определён как статистическая оценка вероятности, определяемой по формуле (8), выполнения тестового задания слабым студентом или (что, то же самое) как вероятности того, что студент, выполнивший данное тестовое задание, в итоге не пройдёт тест в целом.

Байес назвал бы такой показатель условной вероятностью события, заключающегося в том, что студент, выполнивший данное тестовое задание, в итоге не пройдёт тест в целом или статистической ошибкой 2-го рода. Таким образом, правильное выполнение тестового задания в данном случае не играет положительной роли и в итоге затрудняет установление правильного диагноза: студент является слабым. Упростив зависимость (8), коэффициент интуитивности можно определить следующим образом:

$$K_{и} = \frac{N_3}{N_1 + N_3}. \quad (9)$$

Коэффициент интуитивности или статистическая оценка величины ошибки 2-го рода, таким образом, характеризует долю слабых студентов среди всех правильно выполнивших (сильных и слабых) данное тестовое задание. Успешность выполнения тестового задания слабыми студентами (диагноз D_2) не может положительно характеризовать данное тестовое задание. Чаще всего это обстоятельство имеет место, когда формулировка задания некорректна или запутана. Сильный студент, пытаясь определить правильный ответ, продуцирует при этом ложные умозаключения, а слабый студент по интуиции, особенно не вникая в суть вопроса, приходит к правильному решению. Гистограмма распределения коэффициента интуитивности приведена на рисунке 4.

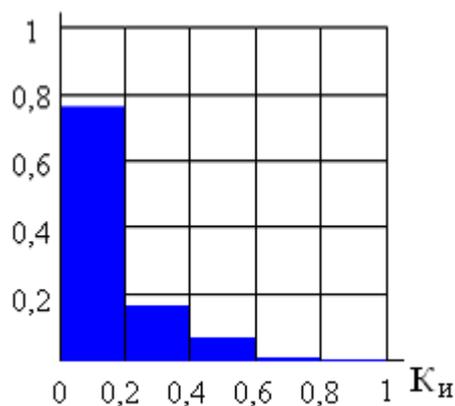


Рис. 4. Гистограмма распределения значений ошибок 2-го рода
(коэффициента интуитивности тестового задания)

Приведенные гистограммы распределений ошибок первого и второго рода характерны для правильно проводимых и хорошо организованных педагогических тестов. Любые попытки «помочь» слабому студенту, подсказав ему правильный ответ на какое-либо отдельное тестовое задание, или «запутать» сильного студента, предложив ему по ходу ответа ложную схему рассуждений, испортят гистограммы, увеличив в них число заданий с большими значениями ошибок 1-го и 2-го рода. Примерно к тем же последствиям приводит использование студентом шпаргалок при ответах на ряд (не на большинство) заданий теста. Таким образом, эти гистограммы в какой-то степени характеризуют «чистоту» проведения процедуры тестирования. Любое вмешательство преподавателя в процедуру тестирования портит статистику тестовых заданий, поэтому необходима полная индифферентность преподавателя по отношению к тестируемым во время выполнения ими теста. Целесообразно также, чтобы преподаватель, чьи студенты проходят тестирование, в аудитории не присутствовал или находился бы в качестве стороннего наблюдателя.

Коэффициенты трудности и интуитивности или ошибки 1-го и 2-го рода тестовых заданий являются соответственно парциальными статистическими оценками ошибок, отнесёнными не к совокупности заданий теста, а к данному конкретным тестовым заданиям. Уровень значимости риска ошибочного реше-

ния при ответе на тестовое задание с учётом (4) и (8) характеризует ситуацию, когда сильный студент не справляется с тестовым заданием, а слабый успешно выполняет его:

$$O_i = K_T + K_{И} = P(D_1 | \bar{z}_i) + P(D_2 | z_i). \quad (10)$$

Применительно к тестовым заданиям статистической оценкой, характеризующей способность отделять сильных студентов от слабых, может служить коэффициент дискриминантности, определяемый по формуле:

$$K_{д} = \frac{N_1 + N_4}{N}. \quad (11)$$

При анализе и отборе тестовых заданий преподавателю нужно стремиться сохранять в тесте те из них, которые имеют более высокие значения коэффициента дискриминантности, являющегося показателем качества тестового задания с позиции преподавателя. Гистограмма распределения коэффициента дискриминантности выборки тестовых заданий для теста с достаточно хорошими диагностическими свойствами приведена на рисунке 5.

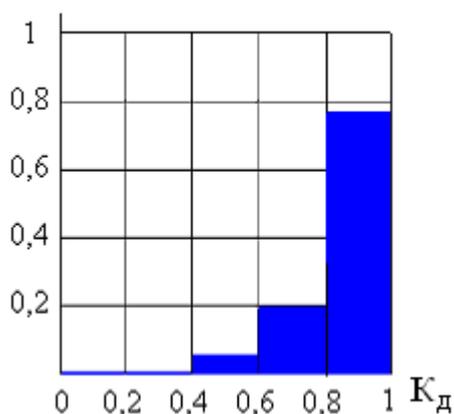


Рис. 5. Гистограмма распределения значений коэффициента дискриминантности тестового задания

При анализе и селекции тестовых заданий можно ориентироваться на все четыре показателя или, учитывая требования максимизации коэффициентов решаемости и дискриминантности и минимизации значений ошибок 1-го и 2-го рода, в качестве интегрального показателя может быть принят коэффициент качества тестового задания

$$K_K = C_1 K_P + C_2 K_D - C_3 K_T - C_4 K_{И}, \quad (12)$$

где C_1 , C_2 , C_3 и C_4 – коэффициенты, учитывающие значимость соответствующих факторов, определяемые императивно, исходя из целей проводимого диагностического обследования.

Реально в техническом вузе существуют сотни учебных предметов, по которым может существовать не один, а несколько тестов, каждый из которых состоит из нескольких десятков или сотен тестовых заданий. Предлагаемая методика оценки тестовых заданий с помощью гистограмм распределений и значений статистических параметров тестовых заданий представляет собой экспресс-метод оценки диагностических свойств без рассмотрения их содержательного наполнения.

5. Диагностические показатели теста

Диагностическим признаком при выполнении отдельного закрытого тестового задания является его результат, который может быть выражен, как правило, одним из двух символов (+/-), (1/0) или (да/нет). Такой признак имеет качественный характер и его можно рассматривать как систему, имеющую одно из двух возможных состояний. Эти состояния альтернативны, так как реализуется только одно из них. Наблюдение комплексного диагностического признака заключается в том, что наблюдается каждый из простых диагностических признаков. Таким образом, тестовые задания в совокупности образуют тест, а ре-

зультаты их выполнения представляют собой комплексный диагностический признак.

Среди методов диагностики метод, основанный на обобщённой формуле Байеса, является наиболее распространённым. К его достоинствам относят простоту и эффективность, к недостаткам - большой объём предварительной информации, "угнетение" редко встречающихся диагнозов и др. В тех случаях, когда объём статистических данных достаточно велик, его целесообразно использовать как один из наиболее надёжных и эффективных методов.

Пусть $D_1, D_2, \dots, D_j, \dots, D_m$ - диагнозы (состояния), обладающие свойствами полноты и несовместности, т.е.

$$\sum_{j=1}^m P(D_j) = 1; \quad (13)$$

$z_1, z_2, \dots, z_i, \dots, z_n$ - диагностические признаки, образующие комплексный диагностический признак Z .

Комплексный диагностический признак Z наблюдается в каждом из диагнозов D_j , если в этом диагнозе наблюдаются диагностические признаки z_1, z_2, \dots, z_n , т.е.

$$P(Z|D_j) = \prod_{i=1}^n P(z_i|D_j), \quad (14)$$

где $P(z_1|D_j), P(z_2|D_j), \dots, P(z_i|D_j), \dots, P(z_n|D_j)$ - частоты наблюдения диагностических признаков в соответствующем диагнозе.

Вероятность наблюдения комплексного диагностического признака Z во всех диагнозах определяется по формуле полной вероятности

$$P(Z) = \sum_{j=1}^m P(D_j) \prod_{i=1}^n P(z_i|D_j), \quad (15)$$

где $P(D_j)$ - априорные вероятности диагнозов D_j , определяемые по статистическим данным.

Диагноз и диагностический признак суть совместные и зависимые события, а их произведение (пересечение)

$$P(ZD_j)=P(D_j)P(Z|D_j)=P(Z)P(D_j|Z). \quad (16)$$

Подставляя (14) в (16) и учитывая (15), имеем

$$P(D_j|Z)=P(D_j) \frac{\prod_{i=1}^n P(z_i|D_j)}{\sum_{j=1}^m P(D_j) \prod_{i=1}^n P(z_i|D_j)}, \quad (17)$$

что и является обобщённой формулой Байеса для определения апостериорной вероятности диагноза при наблюдении комплексного диагностического признака.

Принимая гипотезу о том, чтобы на основании результатов тестирования отнести студента к тому или иному диагнозу, мы можем совершить ошибку 1-го или 2-го рода (при этом не путаем их со статистическими показателями тестовых заданий!), т.е. совершить ошибку в принятии решения. Таким образом, следующий этап решения задачи - выработка правил решения (см. рис.1). В методах статистических решений решающее правило выбирается исходя из некоторых условий оптимальности, например, из условия минимума риска или какого-либо другого. Возникшие в математической статистике как методы проверки статистических гипотез (Нейман, Пирсон), методы статистических решений успешно используются и в задачах диагностики.

Ошибкой 1-го рода или пропуском цели (ПЦ) является случай, когда принимается решение об отрицательном результате тестирования (диагноз D_2), но действительное состояние объекта диагностирования соответствует диагнозу D_1 (сильный студент).

Ошибка 2-го рода или ложное срабатывание (ЛС) - принятие решения о зачёте результатов тестирования (назначение диагноза D_1) слабому студенту, фактический уровень обученности которого соответствует диагнозу D_2 .

Отсутствие ошибок 1-го и 2-го рода будет иметь место тогда, когда произведена правильная оценка учебных достижений учащихся, т.е. объективно сильные из них отделены от объективно слабых. Возможные исходы в случаях принятия или отклонения гипотез приведены в табл. 2.

Таблица 2

Таблица решений при проверке гипотез

Статистическая гипотеза	Принимается	Отвергается
Верна	Правильное решение	Ошибка 1-го рода (ПЦ)
Не верна	Ошибка 2-го рода (ЛС)	Правильное решение

Обозначим события H_{12} - пропуск цели, а H_{21} - ложное срабатывание. $P(H_{12})$ и $P(H_{21})$ - соответственно вероятности ошибок 1-го и 2-го рода. Здесь первый индекс соответствует действительному состоянию, а второй - принятому. Вероятность пропуска цели $P(H_{12})$ равна вероятности произведения двух событий: студент априорно является сильным (вероятность $P(D_1)$) и по результатам тестирования он может быть оценен отрицательно (вероятность $P(D_2|Z)$).

$$P(H_{12}) = P(D_1) \cdot P(D_2|Z). \quad (18)$$

Подобным же образом находится вероятность ложного срабатывания

$$P(H_{21}) = P(D_2) \cdot P(D_1|Z). \quad (19)$$

По результатам тестирования ошибочное решение будет принято, если имеет место или пропуск цели или ложное срабатывание. Таким образом, вероятность ошибочного решения (риска) по правилу сложения вероятностей равна

$$O = P(H_{12}) + P(H_{21}) = P(D_1) \cdot P(D_2|Z) + P(D_2) \cdot P(D_1|Z). \quad (20)$$

В юриспруденции этической нормой считается, что лучше оправдать несколько виновных (ошибки 2-го рода), чем осудить одного невиновного (ошибка 1-го рода). На промышленных предприятиях для поддержания своей марки и престижа вполне обоснованно полагают, что лучше забраковать несколько годных изделий (ошибки 1-го рода), чем выпустить на продажу хотя бы одно бракованное изделие (ошибка 2-го рода).

Различная значимость статистических ошибок 1-го и 2-го рода может быть учтена в суммарном коэффициенте ошибки с соответствующими весовыми коэффициентами:

$$O_{\Sigma} = G_1 \cdot P(H_{12}) + G_2 \cdot P(H_{21}), \quad (21)$$

где G_1 и G_2 - коэффициенты, учитывающие значимость статистических ошибок 1-го рода и 2-го рода.

В образовательном процессе там, где превалирует гуманистическое начало, следует принимать $G_1 < G_2$. В тех случаях, когда определяющим критерием является качество образования, принимают $G_1 > G_2$. Компромиссным решением может служить соотношение $G_1 = G_2 = 1$.

С точки зрения лиц, отвечающих за постановку учебного процесса, преподаватель, допускающий на экзамене большое количество статистических ошибок как 1-го, так и 2-го рода, заслуживает порицания. Напротив, диагностическая деятельность преподавателя, не допускающего таких ошибок, то есть оценивающего сильного студента положительно, а слабого отрицательно, должна приветствоваться. С тех же позиций можно оценивать и тесты.

В некоторых случаях, когда требуется высокая надёжность распознавания (большая стоимость ошибок пропуска цели и ложной тревоги), целесообразно вводить так называемую зону неопределённости (зону отказа от распознава-

ния). Отказ от распознавания, разумеется, является нежелательным событием и свидетельствует, что имеющейся информации недостаточно для принятия решения и нужны дополнительные сведения.

Статистическую информацию, связанную с оценками априорной вероятности диагнозов и частотами (статистическими оценками вероятностей) наблюдения диагностических признаков в тех или иных диагнозах группируют в специальные таблицы, называемыми диагностическими матрицами (табл. 3).

Таблица 3

Форма диагностической матрицы

D_j	$P(D_j)$	$P(z_1 D_j)$	$P(z_2 D_j)$...	$P(z_n D_j)$
D_1					
D_2					
...					
D_m					

Здесь D_j - диагнозы,

$P(D_j)$ - априорные вероятности диагнозов,

$P(z_i|D_j)$ - статистические оценки вероятности наблюдения i -го диагностического признака в j -ом диагнозе.

При прохождении гомогенного теста с однородными тестовыми заданиями, соответствующими одному определённому уровню владения учебным материалом, число диагнозов не может быть большим. Скорее всего, таких диагнозов два: D_1 - сильный студент, D_2 - слабый студент. Студенты, выполнившие по результатам тестирования критериальные требования, например, набравшие не менее 80% возможного количества баллов или выполнившие не менее 80% тестовых заданий, могут быть отнесены к сильным, не выполнившие - к слабым. Количество диагностических признаков (тестовых заданий) в одном тесте может составлять несколько десятков. Ввиду значительного объёма обрабатываемой числовой информации, её анализ не выглядит наглядным и показательным, хотя при использовании компьютерных методов обработки скольконибудь существенных проблем не создаёт. Поэтому для иллюстрации данной

методики рассмотрим более простую модель теста, состоящего всего из трёх тестовых заданий. Предлагаемая методика обрабатывалась с помощью специальной компьютерной программы StatTZ, моделировавшей статистические показатели тестовых заданий в зависимости от успешности и неуспешности их выполнения, причём число заданий в программе было принято именно равным трём.

Итак, будем для определённости считать, что наш тест - гомогенный и содержит всего три тестовых задания. Результаты выполнения этих заданий показаны в табл. 4:

Таблица 4

Статистические параметры тестовых заданий примера теста

	N ₁	N ₂	N ₃	N ₄	K _P	K _T	K _И	K _Д
Задание 1	17	3	1	4	0,72	0,429	0,056	0,84
Задание 2	19	2	2	1	0,875	0,667	0,095	0,833
Задание 3	20	2	1	1	0,875	0,667	0,048	0,875

Диагностическая матрица для этих трёх тестовых заданий:

Таблица 5

Диагностическая матрица примера теста

D _i	P(D _i)	P(z ₁ D _i)	P(z ₂ D _i)	P(z ₃ D _i)
D ₁	0,863	0,85	0,905	0,909
D ₂	0,137	0,2	0,667	0,5

Здесь априорная вероятность диагноза D₁ соответствует отношению числа сильных студентов как правильно, так и неправильно выполнивших ранее эти три задания, к общему числу их выполнявших:

$$P(D_1) = (17+3+19+2+20+2)/(17+3+1+4+19+2+2+1+20+2+1+1) = 0,863.$$

Априорная вероятность диагноза D₂ соответствует отношению числа слабых студентов как правильно, так и неправильно выполнивших ранее эти три задания к общему числу их выполнявших:

$$P(D_2) = (1+4+2+1+1+1)/(17+3+1+4+19+2+2+1+20+2+1+1) = 0,137.$$

Частота наблюдения 1-го диагностического признака $P(z_1|D_1)$ соответствует частоте выполнения задания 1 сильным студентом (диагноз D_1):

$$P(z_1|D_1) = 17/(17+3) = 0,85.$$

Частота наблюдения 1-го диагностического признака $P(z_1|D_2)$ соответствует частоте выполнения задания 1 слабым студентом (диагноз D_2):

$$P(z_1|D_2) = 1/(1+4) = 0,2.$$

Аналогично заполнены остальные ячейки диагностической матрицы для заданий 2 и 3. Метод определения апостериорной (условной или опосредованной) вероятности диагноза основан на обобщённой формуле Байеса (17) для наблюдения комплексного диагностического признака, состоящего из ряда простых.

При использовании данных диагностической матрицы в соответствии с формулой Байеса для получения апостериорной вероятности диагноза необходимо произведение чисел строки, соответствующей данному диагнозу разделить на сумму таких произведений по всем строкам матрицы.

Для диагноза D_1 апостериорная вероятность

$$P(D_1 | Z) = \frac{0,863 \cdot 0,85 \cdot 0,905 \cdot 0,909}{0,863 \cdot 0,85 \cdot 0,905 \cdot 0,909 + 0,137 \cdot 0,2 \cdot 0,667 \cdot 0,5} = 0,985.$$

Для диагноза D_2 апостериорная вероятность

$$P(D_2 | Z) = \frac{0,137 \cdot 0,2 \cdot 0,667 \cdot 0,5}{0,863 \cdot 0,85 \cdot 0,905 \cdot 0,909 + 0,137 \cdot 0,2 \cdot 0,667 \cdot 0,5} = 0,015.$$

Таким образом, успешное выполнение всех трёх тестовых заданий позволило уточнить априорную вероятность диагноза $P(D_1) = 0,863$ и получить апостериорную вероятность $P(D_1 | Z) = 0,985$.

Попробуем учесть результат прохождения теста. В случае правильного выполнения всех трёх тестовых заданий студентом, который по результатам тестирования был признан сильным, таблица исходных данных примет вид:

Таблица 6

Статистические параметры тестовых заданий
при успешном выполнении первого тестового задания

	N ₁	N ₂	N ₃	N ₄	K _Р	K _Г	K _И	K _Д
Задание 1	18	3	1	5	0,731	0,429	0,053	0,846
Задание 2	20	2	3	1	0,88	0,667	0,091	0,84
Задание 3	21	2	2	1	0,88	0,667	0,045	0,88

В данном случае статистической ошибкой 1-го рода считается событие, при котором студент, априорно считавшийся сильным ($P(D_1) = 0,863$), по результатам тестирования мог быть отнесён к слабым ($P(D_2 | Z) = 0,015$). Совпадение этих двух событий (их произведение) и даёт искомую вероятность: $P(O_1) = P(D_1) \cdot P(D_2 | Z) = 0,863 \cdot 0,015 = 0,013$.

При определении ошибки 2-го рода следует учитывать, что студент изначально мог быть отнесён к слабым ($P(D_2) = 0,137$), а по результатам тестирования оказался отнесённым к сильным ($P(D_1 | Z) = 0,985$). Произведение этих событий является ошибкой 2-го рода, и её вероятность:

$$P(O_2) = P(D_2) \cdot P(D_1 | Z) = 0,137 \cdot 0,985 = 0,135.$$

Вероятность принятия правильного решения логически следует из следующего: эти три задания помогают принять правильное решение, если сильный студент ($P(D_1) = 0,863$) по результатам тестирования отнесён к сильным ($P(D_1 | Z) = 0,985$) или слабый студент ($P(D_2) = 0,137$) отнесён к слабым ($P(D_2 | Z) = 0,015$). Предлог «или» говорит нам о сумме событий и о необходимости суммировании соответствующих вероятностей:

$$P_D = P(D_1) \cdot P(D_1 | Z) + P(D_2) \cdot P(D_2 | Z) = 0,863 \cdot 0,985 + 0,137 \cdot 0,015 = 0,852.$$

Рассмотрим далее случай, когда одно или несколько тестовых заданий теста окажутся невыполненными. Пусть в условиях предыдущего примера студент не справился с первым тестовым заданием, правильно выполнив второе и третье. Следует снова воспользоваться диагностической матрицей, заменив в ней частоты наблюдения соответствующего диагностического признака (в данном случае 1-го) во всех диагнозах на статистические оценки противоположного события. Так, рассматриваемая выше диагностическая матрица при невыполнении задания z_1 примет вид:

Таблица 7

Диагностическая матрица теста при неуспешном выполнении первого тестового задания

D_i	$P(D_i)$	$1-P(z_1 D_i)$	$P(z_2 D_i)$	$P(z_3 D_i)$
D_1	0,863	0,15	0,905	0,909
D_2	0,137	0,8	0,667	0,5

При этом для диагноза D_1 апостериорная вероятность

$$P(D_1 | Z) = \frac{0,863 \cdot 0,15 \cdot 0,905 \cdot 0,909}{0,863 \cdot 0,15 \cdot 0,905 \cdot 0,909 + 0,137 \cdot 0,8 \cdot 0,667 \cdot 0,5} = 0,745.$$

Для диагноза D_2 апостериорная вероятность

$$P(D_2 | Z) = \frac{0,137 \cdot 0,8 \cdot 0,667 \cdot 0,5}{0,863 \cdot 0,15 \cdot 0,905 \cdot 0,909 + 0,137 \cdot 0,8 \cdot 0,667 \cdot 0,5} = 0,255.$$

Мы уже с вероятностью существенно меньшей, чем априорная вероятность диагноза можем отнести данного студента к числу сильных студентов, несмотря на то, что два других тестовых задания им были выполнены. При критериальном требовании считать студента сильным, если апостериорная вероятность диагноза будет более 0,8, данный студент при невыполнении им первого тестового задания должен быть отнесён к слабым. Посмотрим, как изменятся в этом

случае статистические характеристики выполнения тестовых заданий. Соответствующая таблица примет вид:

Таблица 8

Статистические параметры тестовых заданий
при неуспешном выполнении первого тестового задания

	N_1	N_2	N_3	N_4	K_P	K_T	$K_{И}$	$K_{Д}$
Задание 1	17	3	1	5	0,692	0,375	0,056	0,846
Задание 2	19	2	3	1	0,88	0,667	0,136	0,8
Задание 3	20	2	2	1	0,88	0,667	0,091	0,84

Как и следовало ожидать, показатели дискриминантности заданий 2 и 3 несколько уменьшились, так как они не помогли диагностировать слабого студента, а показатель дискриминантности задания 1 соответственно немного возрос (задание правильно подсказало нам информацию о том, что студент в итоге оказался слабым).

Показатель качества диагностирования, определённый аналогичным образом по данным примера 2 оказался равным 0,678, то есть существенно ниже.

При невыполнении всех трёх тестовых заданий этот же показатель диагностирования окажется равным 0,171. Столь низкое значение свидетельствует о том, что апостериорные «результаты» тестирования не слишком хорошо подтверждают априорные «надежды» на успешное прохождение теста.

При выполнении большего количества тестовых заданий, чем использованное в приведенных примерах, метод оценки качества теста и диагностических параметров тестовых заданий не изменяется, немного увеличивается лишь объём вычислений, что при использовании компьютерной техники не доставляет каких-либо проблем.

Проблемы доверия к результатам тестирования или оценивания уровня его надёжности относятся к ключевым. Степень доверия результатам эксперимента напрямую зависит от их воспроизводимости и может быть статистически

оценена. Применительно к тестированию это означает, что результаты двух и более тестирований студента с помощью одного и того же теста должны совпадать или быть достаточно близкими. Очень часто студенты, неудовлетворённые оценкой своих достижений, тут же обращаются с просьбой повторить тестирование. По опыту проведения тестирования можно судить, что немедленное удовлетворение такой просьбы способно значительно ухудшить первоначальный результат: сказывается утомление и, как следствие, деконцентрация внимания. Если дать студенту отдохнуть, обратиться к конспекту или учебнику, получить консультацию у преподавателя, то результат тестирования может и должен измениться в лучшую сторону.

Сомнения в диагностических возможностях тестирования вызывает также сопоставление студентом ожидаемых результатов (путём самооценки) с фактическими результатами прохождения теста. Студент А и без тестирования априори считает себя сильнее студента Б. Если результат тестирования студента Б выше, то «виноватым» называется также «плохое» тестирование.

Надежность теста может оцениваться корреляцией между повторными применениями одного и того же теста, корреляцией между данным и эквивалентным тестом, корреляцией между частями одного и того же теста и с помощью других процедур. В силу объективных условий организации и проведения тестирования повторение этой процедуры (так называемый метод «тест-ретеста») может оказаться неприемлемым. Попробуем произвести оценивание в рамках прохождения одного теста. При достаточно большом числе тестовых заданий тест можно условно разбить на две части. Только не первую половину выполняемых тестовых заданий будем сравнивать со второй, а, памятуя об утомляемости студентов к концу тестирования, лучше наблюдать за результатами выполнения нечётных и чётных заданий. У различных по силе студентов результаты выполнения этих двух групп заданий должны приблизительно совпасть. Степень совпадения двух значений, определённая по выборке для нескольких студентов, в статистике оценивается коэффициентом парной корреляции.

Для каждой из половин теста

$$r_{1/2} = \frac{\sum_{i=1}^K X_i Y_i - 2\left(\sum_{i=1}^K X_i \sum_{i=1}^K Y_i\right) / K}{\sqrt{\sum_{i=1}^K X_i^2 - \frac{2\left(\sum_{i=1}^K X_i\right)^2}{K}}} \sqrt{\sum_{i=1}^K Y_i^2 - \frac{2\left(\sum_{i=1}^K Y_i\right)^2}{K}}, \quad (22)$$

где X_i и Y_i - результаты прохождения тестирования соответственно по нечётным и чётным номерам тестовых заданий,

K - число студентов в рассматриваемой выборке.

Коэффициент корреляции, характеризующий надёжность теста в целом определяется по формуле

$$r = \frac{2r_{1/2}}{1+r_{1/2}}. \quad (23)$$

Для гомогенных тестов в качестве X_i и Y_i можно считать количество правильно выполненных нечётных и чётных заданий, для гетерогенных тестов - количество баллов, набранных соответственно за нечётные и чётные задания.

Результаты выполнения одного из тестов, содержащего 40 однородных тестовых заданий выборкой из $K = 12$ студентов, представлены табл. 9.

Таблица 9

Результаты выполнения чётных и нечётных заданий теста

Студент в списке	X_i	Y_i
1	18	17
2	14	16
3	5	6
4	17	16
5	11	11
6	8	7
7	12	11
8	20	18
9	17	19
10	19	18
11	15	15
12	16	14

Расчёты по указанным выше формулам (получены с использованием стандартных статистических функций табличного процессора Excel) приводят к следующим значениям: $r_{1/2} = 0,954$ и $r = 0,976$, что говорит о высоком уровне воспроизводимости результатов тестирования и о хорошем показателе надёжности теста.

Практика тестирования показывает, что при стохастическом назначении тестовых заданий с помощью генератора случайных чисел показатели надёжности теста в среднем оказываются несколько выше, чем в тех случаях, когда для формирования теста преподаватель осуществляет отбор тестовых заданий по своему вкусу или с использованием специализированных компьютерных программ, называемых «генераторами тестов».

6. Диагностическая ценность признаков и обследования

В диагностических обследованиях /7/, в том числе педагогических, большое значение имеет описание объекта диагностирования в системе признаков, обладающих определённой диагностической ценностью. Использование неинформативных признаков не только оказывается бесполезным, но и снижает эффективность самого процесса диагностирования, а также создаёт помехи при распознавании. Диагностические признаки должны быть не только контролепригодны, но и информативны. Последнее означает, что количественное определение диагностической ценности признаков и комплексов признаков может быть проведено на основе теории информации, при этом основной принцип состоит в следующем: диагностическая ценность признака определяется информацией, которая вносится признаком в систему состояний. Распознавание состояний системы осуществляется путём наблюдения за другой, связанной с ней системой, - системой признаков. В качестве диагностического веса реализации признака z_i для диагноза D_j принимается

$$K_{D_j}(z_i) = \log_2 \frac{P(z_i | D_j)}{P(z_i)}. \quad (24)$$

Величина $K_{D_j}(z_i)$ встречается в ряде работ по теории информации. У некоторых авторов эта величина называется ценностью информации, так как представляет собой информацию о состоянии D_j , которой обладает состояние признака z_i .

Диагностический вес отсутствия признака

$$K_{D_j}(\bar{z}_i) = \log_2 \frac{P(\bar{z}_i | D_j)}{P(\bar{z}_i)} = \log_2 \frac{1 - P(z_i | D_j)}{1 - P(z_i)}. \quad (25)$$

Равенство (24) определяет независимый диагностический вес данной реализации признака для диагноза D_j , то есть он характерен для случаев, когда реализации других диагностических признаков ещё неизвестны (при одновременном обследовании по нескольким признакам) и когда вероятность появления данной реализации признака не зависит от результатов предыдущих обследований. Возможны, тем не менее, случаи, когда значение реализации признака зависит от того, какие реализации признаков до этого имели место.

Если апостериорная вероятность диагноза увеличилась по сравнению с априорной, то диагностический вес положителен, отрицательные значения диагностического веса означают отрицание данного диагноза.

Обследование, обладающее небольшой диагностической ценностью для одного диагноза, может иметь значительную ценность для другого. Общей диагностической ценностью обследования по признаку z_i для всей системы диагнозов D_j является количество информации, вносимое обследованием в систему диагнозов. Эта величина представляет собой ожидаемое значение информации,

которое может быть внесено обследованием в установление неизвестного заранее диагноза, принадлежащего рассматриваемой системе диагнозов.

Понятие диагностического веса реализации признака применимо только по отношению к данному конкретному диагнозу как степень его подтверждения или отрицания. Учёт диагностического веса по всем реализациям признака и по всем диагнозам приводит к понятию информативной или диагностической ценности обследования. Диагностический вес той или иной реализации признака ещё не даёт представления о диагностической ценности обследования по данному признаку. Так при обследовании по простому признаку может оказаться, что его наличие не имеет диагностического веса, в то время как его отсутствие чрезвычайно важно для установления диагноза.

Диагностическая ценность обследования для простого признака (тестовые задания закрытого типа) согласно К. Шеннону:

$$K_{D_j}(z_i) = P(z_i | D_j) \log_2 \left[\frac{P(z_i | D_j)}{P(z_i)} \right] + P(\bar{z}_i | D_j) \log_2 \left[\frac{P(\bar{z}_i | D_j)}{P(\bar{z}_i)} \right]. \quad (26)$$

Учитывая очевидное соотношение $P(\bar{z}_i | D_j) = 1 - P(z_i | D_j)$, получим

$$K_{D_j}(z_i) = P(z_i | D_j) \log_2 \left[\frac{P(z_i | D_j)}{P(z_i)} \right] + [1 - P(z_i | D_j)] \log_2 \left[\frac{1 - P(z_i | D_j)}{1 - P(z_i)} \right]. \quad (27)$$

Наибольшую диагностическую ценность имеют обследования по признакам, которые часто встречаются при данном диагнозе, а вообще редко и, наоборот, по признакам, встречающимся при данном диагнозе редко, а вообще - часто. При совпадении $P(z_i | D_j)$ и $P(z_i)$ обследование не имеет никакой диагностической ценности. Диагностическая ценность обследования не может быть отрицательной величиной, так как информация, полученная при обследовании, не

может ухудшить параметры процесса распознавания действительного состояния объекта. Гистограммы распределений значения ценности диагностического обследования для каждого из диагнозов D_1 и D_2 (сильный студент и слабый студент) для одного из реально выполнявшихся тестов показаны соответственно на рис. 6 и рис. 7.

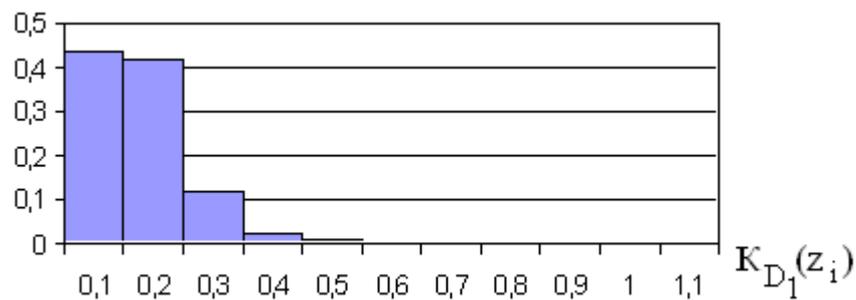


Рис.6. Гистограмма распределения значений ценности диагностического обследования для диагноза D_1

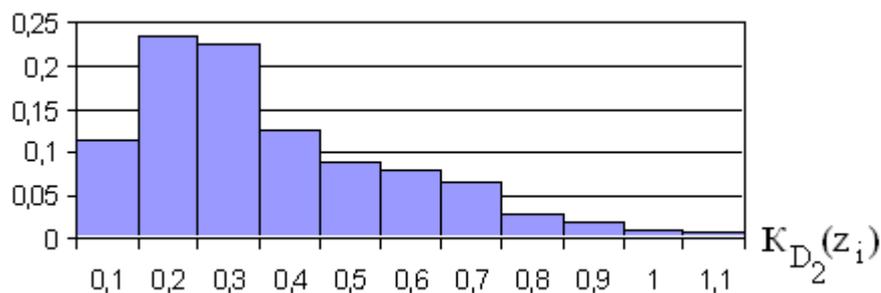


Рис.7. Гистограмма распределения значений ценности диагностического обследования для диагноза D_2

Анализ статистических данных позволил установить взаимосвязи параметров тестовых заданий, представленные корреляционной матрицей табл.10. Здесь тип тестового задания соответствовал следующему: 1 – выбор одного из нескольких, 2 – выбор нескольких из нескольких, 3 – свободная вставка, 4 – упорядочение/соотнесение, 5 – решение задач. Наиболее значимыми являются корреляции диагностической ценности обследования по обоим диагнозам с коэффициентом дискриминантности Кд. Точечные графики для этих параметров приведены на рис. 8 и рис. 9.

Значения коэффициентов парной корреляции параметров тестовых заданий

К-т корр.	Тип ТЗ	К _р	К _т	К _и	К _д	К _{Д₁} (z _i)	К _{Д₂} (z _i)
Тип ТЗ	1,000	-0,416	0,165	-0,276	-0,081	0,050	0,240
К _р	-0,416	1,000	-0,722	-0,020	0,710	0,349	0,209
К _т	0,165	-0,722	1,000	0,007	-0,814	-0,773	-0,375
К _и	-0,276	-0,020	0,007	1,000	-0,486	-0,184	-0,679
К _д	-0,081	0,710	-0,814	-0,486	1,000	0,659	0,617
К _{Д₁} (z _i)	0,050	0,349	-0,773	-0,184	0,659	1,000	0,536
К _{Д₂} (z _i)	0,240	0,209	-0,375	-0,679	0,617	0,536	1,000

Как и следовало предполагать, коэффициенты корреляции между собой ошибок 1-го и 2-го рода не значимы, а обе ошибки значительно снижают показатели ценности диагностического обследования.

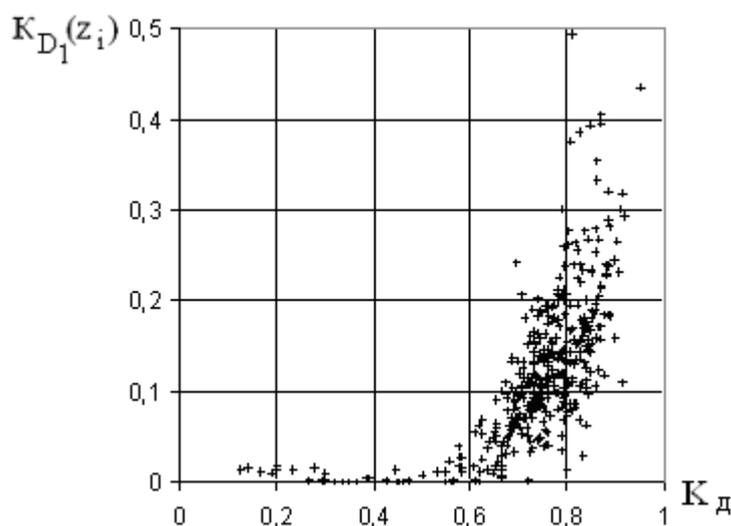


Рис.8. Взаимосвязь коэффициента дискриминантности тестовых заданий с ценностью диагностического обследования для диагноза D₁

Из представленных данных, в частности, следует, что отбор тестовых заданий с достаточно высокими показателями диагностической ценности может производиться с помощью коэффициента дискриминантности, значения которого при селекции тестовых заданий целесообразно принимать не менее 0,6.

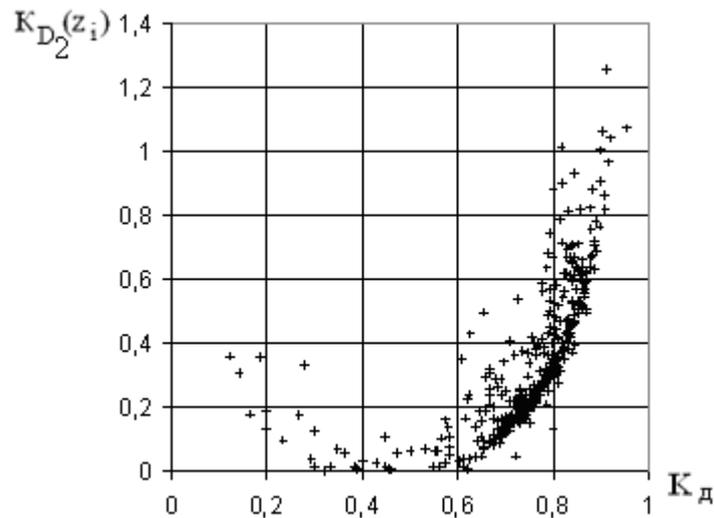


Рис.9. Взаимосвязь значений коэффициента дискриминантности тестовых заданий с ценностью диагностического обследования для диагноза D_2

Высокие значения отрицательной корреляции наблюдаются для показателей ошибки 1-го рода и ценности диагностического обследования для диагноза D_1 , а также для показателей ошибки 2-го рода и ценности диагностического обследования для диагноза D_2 . Соответствующие точечные графики показаны на рис. 10 и рис.11.

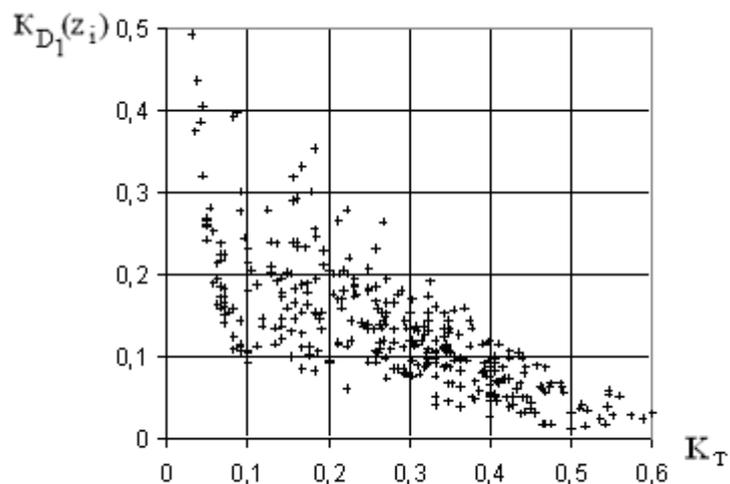


Рис.10. Взаимосвязь значений ошибки 1-го рода тестовых заданий с ценностью диагностического обследования для диагноза D_1

Из приведенных на рис. 10 и рис. 11 данных видно, что наибольшую диагностическую ценность обследования для обоих диагнозов дают тестовые задания со значениями статистических ошибок 1-го и 2-го рода, не превышающими величин 0,2 – 0,4.

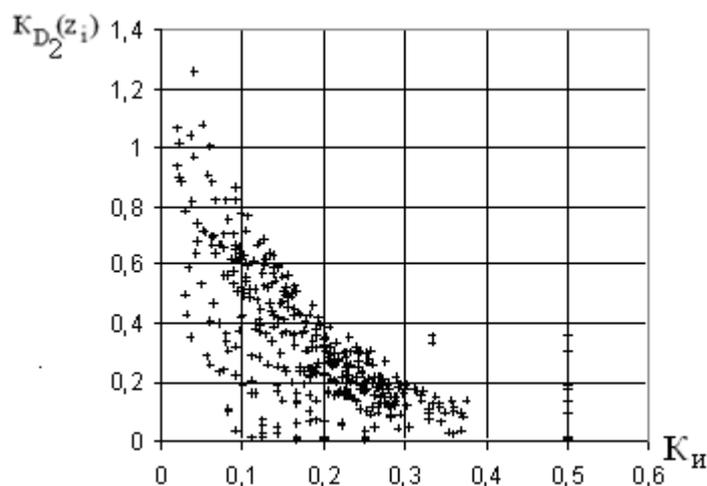


Рис.11. Взаимосвязь значений ошибки 2-го рода тестовых заданий с ценностью диагностического обследования для диагноза D_2

Из анализа данных также следует, что тип тестовых заданий не является критичным для сильных студентов (диагноз D_1), в то же время слабые студенты (диагноз D_2) испытывают наибольшие затруднения при выполнении заданий, требующих проявления более высокого уровня обученности.

При диагностировании чрезвычайно существенным оказывается выбор наиболее информативных признаков для определения состояния объекта. Обычно это связано с трудностью получения диагностической информации (ограниченное время тестирования, большое число таксономических уровней или диагнозов). С теоретической точки зрения процесс диагностического обследования можно представить следующим образом. Имеется объект диагностирования, который может находиться с некоторой вероятностью в одном из

состояний, априорно неизвестном, но по выборке могут быть известны априорные вероятности этих состояний. Энтропия состояний объекта по Шеннону

$$H(D) = - \sum_{j=1}^m P(D_j) \log_2 P(D_j) . \quad (28)$$

В результате полного диагностического обследования по комплексу признаков Z состояние объекта становится известным, например, выясняется, что объект находится в состоянии D_1 , тогда $P(D_1) = 1$, $P(D_j) = 0$ ($j = 2, 3, \dots, m$). После проведения полного диагностического обследования энтропия (неопределённость) состояния объекта

$$H(D|Z) = 0. \quad (29)$$

Внесённая информация, содержащаяся в диагностическом обследовании, или диагностическая ценность обследования

$$K_D(Z) = H(D) - H(D|Z) = H(D) \quad (30)$$

В действительности условие (30) выполняется далеко не всегда. Для подобных состояний остаточная энтропия объекта $H(D|Z) \neq 0$. На практике целесообразно положить

$$K_D(Z) = \xi \cdot H(D) , \quad (31)$$

где ξ – коэффициент полноты обследования, $0 < \xi < 1$.

Коэффициент ξ зависит от требований к надёжности распознавания и для реальных диагностических процессов должен быть достаточно близок к единице. При неизвестных априорных вероятностях диагнозов всегда можно дать верхнюю оценку энтропии объекта с помощью меры Хартли

$$H(D) \leq \log_2 m, \quad (32)$$

где m – число состояний объекта.

Из условия (32) следует, что объём информации, которую необходимо получить при диагностическом обследовании, является заданным и требуется построить оптимальный процесс его накопления. В качестве меры оптимальности диагностического процесса принимают /7/ величину коэффициента оптимальности, определяемого по формуле

$$\lambda = \frac{K_D(Z^{(v)})}{\sum_{i=1}^v C_i}, \quad (33)$$

где $K_D(Z^{(v)})$ - диагностическая ценность обследования по комплексу признаков.

Величина коэффициентов C_i учитывает сложность обследования по диагностическому признаку z_i для диагноза D_j и характеризует трудоёмкость обследования, его достоверность, длительность и другие факторы. Если обследования имеют одинаковую сложность: $C_1 = C_2 = \dots = C_v = C_0$, то

$$\lambda = K_D(Z^{(v)}) / vC_0. \quad (34)$$

Коэффициент оптимальности будет наибольшим, если необходимая величина диагностической ценности получается при наименьшем числе отдельных обследований. В общем случае оптимальный диагностический процесс должен обеспечить получение наибольшего значения коэффициента оптимальности всего обследования. Однако при построении оптимального процесса имеются принципиальные трудности, заключающиеся в том, что реализации диагностических признаков заранее неизвестны и решение на каждом этапе обследования имеет вероятностный характер.

7. Образовательный мониторинг

Понятие «мониторинг» пришло из сферы информатики и означает измерение производительности и контроль за сроками реализации процессов в компьютерных системах. Если оценка состояния образования носит констатирующий характер, то мониторинг качества образования – пролонгированный во времени сопоставительный характер. Под системой постоянного измерения качества (мониторингом качества) понимается регулярное отслеживание состояния объекта и выявление проблем, осуществляемые на основе систематизации и обработки существующих источников информации, а также специально организованных исследований. В качестве эталона при оценке уровня качества выступает норма качества образования, которая определяется как выявленная, общепризнанная и зафиксированная в нормативных документах система требований к качеству образования. Некоторые нормы могут определяться статистически как усредненные значения параметров качества по подразделениям университета, по родственным университетам, по министерству образования в целом.

Анализ традиционных методов проверки и оценки знаний /29, 37/ показывает, что при контроле знаний не устанавливается единых, общепринятых и одинаково понимаемых целей, которых нужно достигнуть в процессе обучения. Цели обучения формулируются очень широко и допускают различную трактовку со стороны преподавателей: дать прочные знания основ наук, сформировать логическое мышление, познавательную активность обучающихся; их мировоззрение и т.д. При такой широкой постановке неясно, какие задания и упражнения, должны выполнить студенты, на какие вопросы ответить, чтобы показать наличие этих характеристик.

Другой фактор, влияющий на обоснованность оценки, - наличие различных объектов контроля. Для одних преподавателей объект оценки - фактический материал, который усвоили студенты, для других - способность применять знания на практике, для третьих - способность переносить знания на решение но-

вых задач и т. д. Соответственно оцениваются различные стороны ответа и разрабатывается своя система проверочных заданий.

Субъективность оценки знаний связана в определенной мере с недостаточной разработкой методов контроля системы знаний. Нередко оценка темы, курса в целом или его частей происходит путем проверки отдельных, часто второстепенных элементов, усвоение которых может не отражать овладение всей системой формируемых знаний, умений, навыков. Качество и последовательность вопросов определяются каждым преподавателем интуитивно, и часто не лучшим образом. Неясно, сколько нужно задать вопросов для проверки всей темы, как сравнить задания по их диагностической ценности.

Каждый из применяемых методов и форм проверки уровня обученности имеет свои преимущества и недостатки, свои ограничения. При устном экзамене качество ответа невозможно предусмотреть, и поэтому затрудняется оценка выявленных знаний. Преподаватель должен быстро реагировать на ответ, задавать, если это надо, дополнительные вопросы, требовать полного, развернутого ответа на основной вопрос, осуществлять непосредственный контакт с учащимся во время проверки. Сделать все это в процессе короткого общения на экзамене очень сложно, что, естественно, сказывается на объективности оценки знаний при устном опросе. Письменная проверка, хотя и более объективна, чем устная, также имеет свои сложности. Преподаватель лишен возможности непосредственно следить за процессом подготовки ответа студента, без чего трудно составить правильное суждение о его знаниях. Письменная форма изложения регулируется особой системой правил, что существенно сказывается на качестве ответов. Сам процесс выражения мысли в письменной форме требует от экзаменуемого не только знания материала, но и других умений, которыми он может не владеть, что также отражается на оценке ответа.

Кроме того, к недостаткам существующей практики проверки и оценки знаний следует отнести стихийность, нерациональное использование методов и форм, отсутствие дидактической целенаправленности, игнорирование препода-

вателем характерных особенностей материала предмета и условий работы в аудитории, отсутствие систематичности в ее проведении.

Справедливой критике подвергается система текущих и вступительных экзаменов. Небольшое количество вопросов не позволяет объективно проверить весь курс; вопросы часто не являются отражением тех знаний, умений, навыков, которые необходимо было сформировать; каждый из экзаменаторов имеет свое суждение о знаниях отвечающего, свои методы и критерии; количество дополнительных вопросов и их сложность зависят от экзаменатора, что также оказывает влияние на общий результат.

Один из наиболее существенных недостатков вступительных экзаменов состоит в том, что они контролируют лишь фактические знания абитуриентов в данный период и не позволяют прогнозировать их будущие достижения. Во многих случаях прогноз более важен, чем настоящий результат, который может измениться за очень короткий промежуток времени. Мониторинг учебных достижений учащихся за достаточно продолжительный период до вступительных экзаменов позволяет оценить возможность прогнозирования успешного обучения при поступлении в вуз /10/.

Другое направление в исследовании этой проблемы связано с изучением воспитательных функций оценки, с анализом психологических условий воздействия оценок преподавателя на студента, с изучением влияния оценки на формирование самооценки обучаемых, на их интерес и отношение к предмету. Существуют объективные условия, при которых оценка преподавателя оказывает свой наибольший и наименьший воспитательный эффект, пути формирования адекватной самооценки студентов с различной успеваемостью. Исследователи установили, что оценка преподавателя приводит к благоприятному воспитательному эффекту только тогда, когда обучаемый внутренне согласен с ней. В действительности студенты ведут оценку своих знаний параллельно и не всегда согласованно с преподавателем. У хорошо успевающих студентов совпадение между собственной оценкой и оценкой, которую поставил им пре-

подаватель, бывает в 46% случаев, а у слабо успевающих - в 11% случаев. По данным других исследований, совпадение между преподавательской и собственной оценкой происходит в 50% случаев. Ясно, что воспитательный эффект оценки будет значительно выше, если студентам станут понятны требования, предъявляемые к ним преподавателями.

Таким образом, проверка и оценка знаний учащихся зависят от многих объективных и субъективных факторов. Применительно к диагностированию обученности можно предложить следующую формулировку: оценка знаний - систематический процесс, который состоит в определении степени соответствия имеющихся знаний, умений, навыков, предварительно планируемыми. Как следует из определения, первое необходимое условие оценки - планирование образовательных целей, без этого нельзя судить о достигнутых результатах. Другое условие - установление фактического уровня знаний и сопоставление его с заданным. Исходя из сказанного, процесс оценки знаний включает в себя следующие компоненты: определение целей обучения; выбор контрольных заданий, проверяющих достижение этих целей; отметку или иной способ выражения результатов проверки.

Все компоненты оценки взаимосвязаны. В зависимости от поставленных целей по-разному строится программа контроля, подбираются различные типы вопросов и заданий.

При использовании объектно-ориентированного подхода любой объект изучения может быть описан определенной совокупностью знаний о нем. Эти знания отражают непосредственные факты, связи между объектами, законы, теории или включают методологические или оценочные знания (свойства, методы и события объектов). В зависимости от целей обучения и начального уровня подготовки одни и те же знания могут изучаться с различной полнотой, глубиной, обобщенностью, осознанностью и т. д. Эти качества знаний могут служить целями обучения. Преподаватель, анализируя задачи занятия, специфику своей дисциплины, конкретную тему, ставит определенную цель, форми-

рует те или иные качества знаний. В одних случаях студенту необходимо получить полные и глубокие знания о предмете, явлении, в другом - планируется лишь знакомство его с определёнными фактами. Часто студенту требуется освоить лишь небольшой объем информации, но уметь применять свои знания на практике. Руководством к выбору целей обучения является раскрытие содержания качеств знания.

Полнота знаний определяется количеством знаний об изучаемом объекте, входящих в учебную программу данной дисциплины; глубина - совокупностью осознанных знаний об объектах изучения. Полнота и глубина знаний - связанные, но не тождественные понятия. Полнота допускает изолированность знаний друг от друга; глубина же, напротив, предполагает наличие осознанных существенных связей, в разной степени опосредованных.

Выделяются и другие качества знаний. Оперативность определяется числом ситуаций или способов, в которых студент может применить то или иное знание; гибкость - быстротой нахождения вариативных способов применения знаний при изменении ситуации. Это качество предполагает умение выбрать нужный в данный момент способ деятельности из ряда известных.

Знания, усваиваемые студентами, могут быть обобщенными и конкретными, свернутыми и развернутыми, системными и бессистемными. Выделенные качества знаний взаимообусловлены, каждое содержит в себе в свернутом виде другие качества. При этом одни знания отражают содержание обучения и не зависят от субъекта, а другие составляют характеристику личности и не могут рассматриваться отдельно. К объективным качествам относятся полнота, глубина, оперативность, конкретность, обобщенность, систематичность, системность, развернутость знаний; к субъективным - гибкость, свернутость, осознанность и прочность.

Важнейшим требованием к знаниям является их осознанность. Это качество выражается в понимании обучаемым связей между учебными элементами, умение выделять между ними существенные и несущественные связи, в позна-

нии способов и принципов получения знаний. Осознанность знаний - наиболее обобщенное качество, отражающее конечный результат усвоения и синтезирующее другие показатели. В зависимости от глубины, полноты и системности одни и те же знания могут быть по-разному осознаны и использованы.

Формулировка целей обучения с помощью системы требований к качеству знаний оказывается полезной преподавателю в его текущей учебной работе.

Мониторинг отличается от обычной оценки знаний тем, что обеспечивает преподавателя оперативной обратной связью об уровне усвоения учащимися обязательного учебного материала.

Образовательный мониторинг - это система организации сбора, хранения, обработки и распространения информации о деятельности педагогической системы, обеспечивающая непрерывное слежение за ее состоянием и прогнозированием ее развития. Одним из предметов изучения является исследование воздействия процесса обучения на личность обучаемого. То есть объектом наблюдения является непосредственно обучаемый и то, какое воздействие на него производит процесс обучения.

Если бы отражение системой самой себя и окружающей среды было бы абсолютно, решения безукоризненны, а исполнение безупречно, то управление исчерпывало бы себя в одном цикле. Однако ввиду эволюции среды и самой системы под действием возмущений, вследствие самосовершенствования и старения, ввиду совершенствования познания и связанного с этим изменения целей управления, наконец, ввиду несовершенства исполнительных органов приходится многократно повторять циклы управления, следовательно, реальная система управления фактически должна функционировать непрерывно.

Информацию о состоянии системы можно разделить на две составляющие:

- собственно показатели качества системы;
- показатели, характеризующие обеспечение качества.

Образовательный мониторинг - категория педагогическая и управленческая, поскольку он не копирует общие положения теории информации, а переводит

их на язык педагогики, психологии и управления. Само понятие «информация» неразрывно связано с динамическими процессами взаимодействия ее источника и приемника, которое может быть реализовано лишь при наличии между ними связи (канала связи). Для выявления информации о состоянии системы, особенно при построении оптимальных диагностических процессов, широко используется теория информации, базирующаяся на работах Н. Винера и К. Шеннона и получившая признание как общая теория связи статистических систем. Одно из центральных мест в теории информации занимает понятие энтропии системы, которая есть мера беспорядка, мера неопределённости наших представлений о внутренней структуре системы.

Так как обучение является передачей информации ученику, то по определению академика В.Н. Глушкова, что информационные технологии - процессы, связанные с переработкой информации, можно сделать вывод о том, что в обучении информационные технологии использовались всегда. Более того, любые методики или педагогические технологии описывают, как переработать и передать информацию, чтобы она была наилучшим образом усвоена учащимися. То есть, любая педагогическая технология - это информационная технология. Когда же компьютеры стали настолько широко использоваться в образовании, что появилась необходимость говорить об информационных технологиях обучения, выяснилось, что они давно фактически реализуются в процессах обучения, и тогда появился термин «новая информационная технология обучения». Таким образом, появление такого понятия «новая информационная технология (НИТ)» связано более широким внедрением компьютеров в образовании.

Состав образовательного мониторинга является педагогической интерпретацией информации о деятельности педагогической системы. Исходя из этого, переведем общие положения в педагогические категории:

Источники информации. Так как объектом наблюдения у нас является студент, то он и есть основной источник информации. На каждом уровне полу-

ченная от студента информация соответствующим образом обобщается и тогда в роли источника выступают группа, поток, курс, факультет, университет.

Обработка информации. Проводится на каждом уровне с учетом его специфики для получения соответствующего обобщения и анализа.

Хранение информации. На каждом уровне сохраняется необходимая информация для сравнительного анализа и прогнозирования.

Распространение информации. Преемниками информации являются практически все уровни управления процессом обучения - от преподавателя до президента университета и выше, при этом для каждого из них выделен свой уровень обобщения и анализа исходных данных об учебных достижениях.

Информационные технологии включают программированное обучение, интеллектуальное обучение, экспертные системы, гипертекст и мультимедиа, микромиры, имитационное обучение, демонстрации. Эти частные методики должны применяться в зависимости от учебных целей и учебных ситуаций, когда в одних случаях необходимо глубже понять потребности учащегося, в других - важен анализ знаний в предметной области, в третьих - основную роль может играть учет психологических принципов обучения.

Как мы видим, главное в НИТ - это компьютер с соответствующим техническим и программным обеспечением, отсюда и определение: новые информационные технологии обучения - процесс подготовки и передачи информации обучаемому, средством осуществления которого является компьютер. Такой подход отражает, как было отмечено выше, первоначальное понимание педагогической технологии как применение технических средств в обучении. Суть современного подхода заключена в идее максимально возможной управляемости деятельностью университета, прежде всего его основного элемента - процесса обучения.

В образовании «педагогическая технология» и «информационная технология» - это в определенном смысле синонимы. Компьютер - это мощный инструмент, позволяющий решать новые, ранее не решенные дидактические зада-

чи, однако абсолютное большинство технологий, основанных на самом широком использовании компьютерной техники опирается на давно известные педагогические идеи. Более того, они вообще не удовлетворяют основным требованиям понятия «технологии». При использовании современных обучающих средств и инструментальных сред, создаются прекрасно оформленные программные продукты, не вносящие ничего нового в развитие теории обучения. Поэтому можно говорить только об автоматизации тех или иных сторон процесса обучения, о переносе информации с бумажных носителей на магнитные и т.д.

Говорить же о *новой информационной технологии обучения* можно только в том случае, если:

- она удовлетворяет основным принципам педагогической технологии (предварительное проектирование, воспроизводимость, целеобразование, целостность);
- она решает задачи, которые ранее в дидактике не были теоретически и/или практически решены;
- средством подготовки и передачи информации обучаемому является компьютер.

Структура управления качеством образования с использованием НИТ может быть представлена в виде следующей последовательности действий:

1. Формирование цели.
2. Конкретизация цели.
3. Создание автоматизированной системы педагогического контроля.
4. Создание педагогической системы психологического контроля.
5. Создание системы педагогического мониторинга.
6. Создание системы психологического мониторинга.
7. Определение начального состояния обучаемого: педагогического и психологического.

8. Выработка прогноза по обучаемому: педагогическая траектория, психологическая траектория.

9. Формулировка цели для учащегося: педагогической, психологической.

10. Анализ полученных результатов процесса обучения.

11. Корректировка целеполагания и деятельности на всех уровнях (внутригрупповой, факультетский, университетский).

Пункты 1- 6 относятся к уровню проекта и являются инвариантными относительно дисциплин, реализующих одинаковое содержание образования. Пункты 7 - 11 являются конкретизацией проекта для реальных студентов выбранного направления подготовки, анализ и корректировка полученных результатов осуществляется на всех уровнях.

8. Психологическое тестирование

"Люди, - писал Платон, - рождаются не слишком похожими друг на друга, их природа различна, да и способности к тому или иному делу также... Поэтому можно сделать все в большем количестве, лучше и легче, если выполнять одну какую-нибудь работу соответственно своим природным задаткам" /56/. Одной из самых распространенных форм диагностики человека в те времена была физиогномика - искусство распознавания характера и способностей человека по его внешнему виду. Гиппократ, который впервые ввёл в обращение этот термин, считал физиогномику наукой. Тогда же появились первые учебники и практические руководства по физиогномике, а также и первые специалисты.

В Афинах обсуждался вопрос о зависимости могущества государства от способностей лиц, им управляющих. На вопрос о том, каких правителей нужно выбирать, Сократ говорил: "Надо отдавать предпочтение самым надежным, мужественным, и по возможности, самым благообразным; кроме того, надо отыскивать людей не только благородных и строгого нрава, но и обладающих

также свойствами, подходящими для такого воспитания. У них... должна быть острая восприимчивость к наукам и быстрая сообразительность. Надо искать человека с хорошей памятью, несокрушимо твердого и во всех отношениях трудолюбивого".

Пифагор обращал также внимание на походку и на смех молодых людей, утверждая, что манера смеяться служит самым хорошим показателем характера человека. Он внимательно относился к рекомендациям родителей и учителей, тщательно вел наблюдение за каждым новичком, особенно после того, как последнего приглашали свободно высказываться и, не стесняясь, смелее оспаривать мнение собеседников /76/.

Основоположником научного подхода к исследованию возможностей человека считают известного английского ученого Френсиса Гальтона. В 1884-1885 годах он проводил серию испытаний для посетителей своей лаборатории, возраст которых был от 5 до 80 лет. За небольшую плату определялись быстрота реакции, вес, жизненная емкость легких, сила кисти, сила удара кулаком, станова́я сила, рост, острота зрения, оценивались способности запоминать буквы и различать цвета, ряд физиологических возможностей организма и некоторые психические свойства. По полной программе было обследовано 9337 человек. Ф. Гальтон отмечал, что методически упорядоченное диагностирование требует специальных условий эксперимента. Это был существенный отход от тысячелетней практики испытаний и проверок, методика и содержание которых основывалось на интуиции. Господствовавший в то время в науке радикальный эмпиризм рассматривался рядом ученых конца XIX века как приемлемая альтернатива идеализму, а эксперимент - как фундамент подлинной науки. Хотя не все, что сделано Ф. Гальтоном, может выдержать критику с позиции современных представлений о диагностировании, он сделал первый шаг на пути создания объективных методов оценки способностей и свойств личности, которые затем легли в основу разработки диагностических методов.

Как и всякая научная дисциплина, теоретическая психофизиология базируется на ряде принципов. Два из них являются главными:

- исследование духовных, психических или физических структур и функций в целом или в отдельности должно быть морально оправданным;
- предпочтительны способы познания соответствующие естественным способам (целостное мировосприятие, системный подход).

Весьма существенным понятием для высшего профессионального технического образования является компетентность специалиста, определяемая на основе психических механизмов реального человеческого поведения. По Дж. Равенну: «Компетентность – это специфическая способность, необходимая для выполнения конкретного действия в конкретной предметной области и включающая узкоспециальные знания, особого рода предметные навыки, способы мышления, а также понимание ответственности за свои действия» /63/. Согласно Дж. Равену, быть компетентным означает иметь определённый набор «специфических или низших» компетентностей (наблюдать, быть глубоко осведомлённым в предмете, самостоятельно ставить вопросы, доказывать собственную правоту, справляться с межличностными конфликтами и т.д.), а также иметь набор «высших» компетентностей (наличие высокого уровня инициативы, способности организовывать других людей для достижения поставленной цели, готовность оценивать и анализировать социальные последствия своих действий и т.д.).

Природа компетентности такова, что она может проявляться только при условии глубокой личной заинтересованности в данном виде предметной практической деятельности. Человеческая психика является и продуктом этой самой предметной деятельности. При этом понятие деятельности играет ключевую, методологически центральную роль. Его место и роль определяются, прежде всего, тем, что они принадлежат к разряду универсальных предельных абстракций. Это означает, что наиболее эффективный путь формирования высших

психических функций заключается в разработке и использовании соответствующих форм предметной деятельности, а не просто в усвоении знаний. Совершенно очевидно, что для эффективного обучения необходимо детальное изучение структуры деятельности, а также поиски операциональных формулировок, выражающих связь предметной деятельности и знания.

Для оценки профессиональной направленности на основе предпочтений используется методика, в основу которой положена схема классификации профессий, предложенная НИИ профтехобразования под руководством Е.А. Климова, в соответствии с которой все профессии делятся на пять групп по предмету труда:

- человек – природа;
- человек – техника;
- человек – человек;
- человек – знаковая система;
- человек – художественный образ.

Понятие политехнического образования традиционно отождествляется с изучением основ общественного производства, которые определяют уровень его технологического развития. Базовыми дисциплинами политехнического образования являются общеинженерные, экономические, экологические и др.

Таким образом, политехническое образование представляет собой изучение законов, происходящих в природе, созданной человеком, и формирование на основе приобретённых знаний взаимоотношений с окружающей средой, обусловленных особенностями общественного производства.

Профессиональное образование (высшее) направлено в первую очередь на изучение методов и особенностей конкретных видов трудовой деятельности, формирование умений и навыков обращения с соответствующими объектами и средствами труда. Профессиональному образованию способствует изучение так называемых специальных дисциплин. В социально-историческом аспекте профессиональное образование является предшественником общего и политехни-

ческого образования. Приобретение профессиональных навыков после прохождения общего и политехнического образования понижает мотивацию изучения последних.

По определению С.А. Тихомирова /59/, инженер - это субъект, занятый преимущественно знаковой деятельностью, направленной на разработку, исследование, усовершенствование и длительную нормальную эксплуатацию технических объектов или организацию их производства, основанную на использовании научно-технических знаний и средств умственного труда.

Объектами инженерной деятельности служат технические изделия (в широком понимании этого слова), существенными признаками которых являются:

- опосредованность их воздействия на материальный субстрат техники;
- научная обоснованность инженерных решений;
- учёт экономических, временных и др. факторов.

Средствами инженерного труда являются:

- научные знания, адекватно отражающие физическую и математическую модели создаваемого технического объекта;
- нормативно-техническая документация (стандарты, технические условия, отраслевые нормы, правила техники безопасности и др.);
- ограничения социально-экономического характера (стоимостные и ценовые показатели, потребности общества в технике, спрос и предложение);
- информационно-вычислительная техника, применяемая для сбора, хранения, обработки и представления технической информации.

Целями и результатами инженерной деятельности могут являться:

- новые технические идеи, отражающие новации в области принципов функционирования технических объектов, областей их возможного использования;
- техническая документация в виде чертежей, схем, программ, графиков, технических условий, пояснительных записок, технологических карт и др.;

- организационные решения - распоряжения, указания, объяснения, обоснования и др.

Техническое мышление - это особая форма ментальной деятельности, направленной на изменение действительности в техносфере. Специфическими особенностями технического мышления и наиболее значимыми в профессиональной деятельности инженера являются:

- способность к абстрагированию, к формализованному восприятию (чертёж - одна из форм формализации технических объектов);
- понимание структуры и принципов функционирования технических объектов;
- способность к структурно-функциональным и системно-элементным переходам;
- способность к преобразованию информации из зрительно-пространственных образов в плоские графические изображения и наоборот;
- способность мыслить по аналогии и контрасту;
- комбинаторное мышление, связанное с комбинированием и взаимозаменяемостью отдельных элементов систем, их свойств, методов и событий;
- способность к анализу и синтезу технических объектов;
- способность к оценке предлагаемых технических решений;
- способность к пониманию структуры и принципов функционирования технических объектов.

Инженерное профессиональное мышление это интегральное многоуровневое отражение и преобразование в обобщённом виде объектных характеристик инженерной деятельности, детерминируемых его целями и задачами, личностными особенностями специалиста.

Структура инженерной деятельности сложна и многообразна, детерминирована внутренним содержанием, потребностью к самовыражению и самореализации, а также общественным разделением труда. При этом могут быть раз-

личными также цели, задачи, содержание, методы и результаты труда. Можно выделить следующие профессиональные группы:

Инженеры-исследователи осуществляют поиск и формирование принципов функционирования технических объектов в основном за счёт развитой технической интуиции, воображения, формально-логического и абстрактно-логического мышления, критичности суждений, способности прогнозировать, анализировать и обобщать получаемые результаты. Профессиональная деятельность инженеров-исследователей носит явно выраженный творческий характер. Склонность к такому виду инженерной деятельности является скорее следствием аффективных качеств, нежели результатом, достигнутым в результате обучения. Наиболее характерные черты инженера-исследователя - это достаточно высокий уровень теоретической подготовки, способность к осмыслению и интерпретации результатов исследований, общая направленность на когнитивные виды деятельности, внутренняя неудовлетворённость, целеустремлённость.

Инженеры-проектировщики, занимающиеся разработкой нормативно-технической документации и решением конструкторских задач. Их в меньшей степени интересуют вопросы изготовления или эксплуатации конструкторских разработок и в большей - проблемы их схемного решения или конструктивного исполнения. Работа носит прикладной характер, направлена на практический результат в виде знаковой модели (эскиз, чертеж, пояснительная записка, технические требования, спецификация и др.) Из разрабатываемой документации должно быть понятны состав, назначение, принцип действия предлагаемого конструктивного решения.

Инженеры-технологи, обеспечивающие процесс изготовления технического объекта с помощью имеющихся или разрабатываемых средств производства. Содержанием их деятельности является проектирование технологического процесса, выбор технологического оборудования, назначение режимов металлообработки, контроль соответствия параметров технологического процесса

установленным требованиям. Типичными профессиональными качествами инженера-технолога являются систематичность мышления, склонность к анализу, способность к реконструктивной деятельности, т.е. к переходу от абстрактного к конкретному мышлению. Результатом их деятельности является построение маршрутных и технологических карт.

Инженеры-эксплуатационники ответственны за соблюдение норм и правил эксплуатации технических объектов, выполнение технических условий, организацию и проведение профилактики технических объектов, их технического обслуживания, ремонта, обеспечение запасными частями. Анализируя причины отказов, они должны видеть возможные причины их возникновения, т.е. должны обладать способностью в рассуждениях переходить от конкретного к абстрактному.

Инженеры-организаторы осуществляют мобилизацию коллектив или рабочую группу на выполнение исследовательских, конструкторских, технологических и производственных задач; выполняют административно-хозяйственные функции по планированию и организации производства, подбору и расстановке кадров, техническому обеспечению производства, охране труда и управлению людьми.

Известно, что качественное описание всегда является недостаточно точным, поскольку вербальная модель не всегда в состоянии передать дифференцированность изучаемых явлений и своеобразие их динамики. Одно только качественное описание не позволяет определить также тенденции развития и величину погрешности экспериментального исследования. Количественный анализ всегда предшествует и, что особенно важно, следует за качественным анализом. Таким образом преодолевается субъективизм, формулируемые суждения становятся более независимыми от личности исследователя, обеспечивается возможность из объективной проверки. Поэтому Б.Г. Ананьев предложил классификацию методов психологической диагностики, исходя из порядка операций в диагностическом обследовании. Все методы им распределены по четырём

группам. В первую группу или в группу организационных методов относится лично-ориентированный лонгитюдный метод, который предполагает сравнительное исследование одних и тех же объектов в течение достаточно продолжительного времени. Вторая группа методов включает в себя известные методы эмпирической психодиагностики: наблюдательные методы (наблюдение и самонаблюдение), экспериментальные методы (тесты, опросники, анкеты, интервью, беседы), праксиметрические методы (приёмы анализа процессов и продуктов деятельности: хронометрия, профессиографическое описание, экспертная оценка выполненных работ), моделирование (математическое, кибернетическое) и биографические методы (исследование жизненного пути). Третью группу методов составляют методы качественного и количественного анализа на основе стандартных методов статистической обработки данных, включая классификационные методы, разработку типологии, составление психологической казуистики (описание наблюдаемых случаев). Четвёртую группу методов – интерпретационные методы – образуют методы генетического и структурного анализа. Генетический анализ рассматривает весь экспериментальный материал в характеристиках его развития, а структурный анализ – в типах связей между отдельными компонентами структуры изучаемой личности или структуры социальной группы.

Для психологического тестирования, так же как и для педагогического тестирования, чрезвычайно важное и актуальное значение имеет доказательство валидности и надёжности предлагаемой методики исследования. Применяемая методика может считаться валидной лишь в том случае, если между результатами психологического тестирования и оценкой результативности обучения будет обнаружена устойчивая положительная связь. Надёжность тестирования, как и отмечалось выше, определяет пригодность данной методики в качестве измерительного инструмента, а также точность проводимых с её помощью измерений. Приступая к выбору измерительных методик, необходимо иметь чёткое представление о цели исследования и удовлетворят ли результаты измере-

ния требованиям адекватного решения конкретной исследовательской или практической задачи.

Одной из важнейших задач, которую позволяет решить психологическое обследование индивидуализация обучения, под которой понимается такая организация учебного процесса, при которой выбор способов, приёмов, темпа обучения учитывает индивидуальные различия учащихся, уровень развития их способностей к учению, т.е. позволяет осуществлять дифференцированный подход в обучении. Между понятиями дифференцированного и индивидуального подходов существуют методологические различия. Под индивидуальным подходом понимается такая обучающе-корректирующая деятельность преподавателей по управлению учебной познавательной деятельностью студентов, которая основана на учёте их индивидуально-психологических особенностей, что, в конечном итоге, обеспечивает более высокую успешность этой деятельности. Индивидуализированное обучение ставит и решает задачи, связанные с индивидуализацией процесса обучения, но не затрагивает содержание и структуру учебного материала. Методика индивидуализированного обучения предусматривает выработку концепции организации процесса обучения, включающего следующее:

- обоснование и конкретизация понятийно-категориального аппарата индивидуализированного обучения;
- систему принципов индивидуализированного обучения;
- систему диагностики индивидуальных особенностей студентов;
- систему дидактических приёмов, средств и форм организации индивидуальной учебной работы со студентами.

В направлениях зарубежной педагогики существуют различия в трактовке этих понятий. В США, например, понятие индивидуализации обучения охватывает любые формы и методы учёта индивидуальных особенностей учащихся, и иногда определяются как стратегии обучения. Во Франции, согласно исследованиям М.В. Кларина, сложилось понимание индивидуализации как совер-

шенствование самостоятельной работы, планируемой с учётом индивидуальных особенностей студентов. Если студенты работают самостоятельно с одним и тем же учебным материалом, то это – индивидуальная работа, если же все задания разные и подобраны с учётом индивидуальных особенностей – это индивидуализированное обучение.

И.Э. Унт под дифференциацией понимает учёт индивидуальных особенностей учащихся в той форме, когда они группируются на основании каких-либо особенностей для отдельного учебного процесса. При этом индивидуализация обучения должна исчерпывающе отражать учёт индивидуальных особенностей обучаемых, охватывать все формы и методы учёта этих особенностей.

В сравнительной экспериментальной психологии изучение профессионально значимых характеристик личности предполагает как достаточно сложный процесс измерения, так и специфическую обработку данных эксперимента. Сложность организации выражена в первую очередь в подборе групп испытуемых, различия между которыми классифицируются по какому-либо признаку (направление подготовки, курс, пол, возраст и др.) Другой особенностью организации эксперимента является то, что при выделении нескольких групп испытуемых необходимо соотнесение результатов экспериментального исследования между собой. С целью получения таких сопоставимых результатов диагностические исследования следует проводить поэтапно. На первом этапе методика изучения функциональных характеристик не отличается от требований, предъявляемым к методам и процедуре исследований в области общей и прикладной психологии. Следующий этап исследования, имеющий целью определение уровня развития интеллектуальных функций, предполагает специальную статистическую обработку данных для получения интегрального показателя, наиболее адекватно отражающего исследуемую психическую функцию и характеризующего её уровень продуктивности. На основе этого показателя возможно соотнесение полученных данных по разным группам диагностируемых.

Для получения этого показателя производят первичную обработку данных, зафиксированных по каждой методике отдельно. При этом абсолютные значения разных показателей сводятся в условные оценки – баллы, полученные путём обычного шкалирования. На заключительном этапе исследования, связанного с изучением индивидуальных особенностей психических функций диагностируемых, проводится специальная математико-статистическая обработка экспериментальных данных, предполагающая получение сопоставимых между собой результатов исследования разномодальных характеристик путём перевода интегрального показателя развития отдельных свойств функции в шкальные оценки. После такой обработки экспериментальных данных мы имеем возможность сопоставить результаты исследования с точки зрения профессионально значимых индивидуально-типических особенностей.

Поскольку воспитание и обучение нормативны по своей природе, то психологические результаты, которые первоначально фиксируются в целях обучения должны, быть заданы в конкретном перечне требуемых и взаимосвязанных психологических качеств с определением желательной степени их выраженности. Соответствие конечных или промежуточных психологических результатов обучения с произведенными затратами разного рода (временными, материальными, физическими, психическими и др.) даёт производный от результативности показатель – эффективность обучения, который свидетельствует не столько об уровне об уровне достижения целей, сколько о трудоёмкости и экономичности выбранных способов обучения.

Для высшей технической школы, являющейся школой профессиональной, психологический результат выступает, прежде всего, в форме профессионально и социально значимых психологических качеств личности, всё то, что определяет профессиональную компетентность и профессиональное мастерство.

Профессиональная направленность выступает как системообразующее качество личности оптанта (оптанта - выбирающий профессию от лат. optatio – же-

ление, избирание), определяющее её психологический склад. В этой направленности выражаются цели, во имя которых действует личность, её мотивы и субъективное отношение к различным сторонам действительности. По этому поводу С.Л. Рубинштейн отмечал: «Проблема направленности – это, прежде всего, вопрос о динамических тенденциях, которые в качестве мотивов определяют деятельность, сами, в свою очередь, определяясь её целями и задачами». Между развитием мотива и овладением деятельностью существуют сложные взаимоотношения: формирование мотива часто опережает формирование деятельности, а иногда, наоборот, отстаёт, что сказывается и на результате степени овладения деятельностью. Каждый период жизни человека, изменение обстоятельств приводят к изменению мотивационной сферы. Адекватная имеющейся мотивации проблема выбора вида и типа профессиональной деятельности оказывает существенное влияние на успешность профессионального обучения и на продуктивность учебной деятельности. Данные анализа мотивов выбора профессии показывают, что значительную роль в этом играют советы окружающих: 25% выбирают профессию под влиянием друга, который более самостоятелен; 17% - по совету родителей; 9% - под влиянием средств массовой информации, 9% руководствуются в выборе малозначительными факторами, например, близость к дому; и только 40% опрошенных выбирают профессию, ориентируясь на содержание будущей профессиональной деятельности, хотя многочисленные исследования показывают, что профессиональные намерения являются более устойчивыми, обучение навыкам профессиональной деятельности проходит быстрее и эффективнее, если главной причиной выбора является ориентация на содержание будущей профессиональной деятельности. Исследования позволили выявить и сгруппировать качества, определяющие успешность учебной деятельности, показанные в табл. 11:

Свойства личности, необходимые для успешной учебной деятельности

Св-ва	Краткая характеристика свойства	Набор качеств, характеризующих свойство
Память	Память – накопление человеком прошлого опыта, сохранение и последующее воспроизведение полученной информации в целях преобразования природы и самого человека	Способность воспроизводить усвоенное Скорость деятельности Сохранение информации Точность воспроизведения Быстрота реакции
Воля	Воля – способность к совершению сознательных целенаправленных действий и поступков, осуществление которых связано с преодолением значительных трудностей.	Настойчивость Работоспособность Самообладание Требовательность к себе Самостоятельность
Мышление	Мышление – производство идей, представлений, сознания, отражение действительности в таких её объективных свойствах, в которые включаются и недоступные естественному восприятию объекты	Сообразительность Осознанность Умение обобщать Интуиция Осмысленность действий Систематизация ассоциаций
Внимание	Внимание – избирательная, сосредоточенная направленность сознания на определённый объект	Активность Устойчивость Переключаемость внимания Наблюдательность
Восприятие	Восприятие – процесс отражение человеком предметов и явлений объективной действительности в ходе их воздействия на органы чувств	Способность анализировать Адаптивность
Представление	Представление – чувственно-наглядный образ предметов и явлений объективного мира, которые воспринимались нами ранее, но в данный момент не воздействуют на наши органы чувств.	Рефлексия Чёткость действий

Таким образом, модель обучаемого – это целостная, сложно-организованная система свойств и качеств, включающая в себя отдельные подсистемы, обуславливающая комплексный системный подход к осуществлению индивидуализации учебной деятельности. Структурно-функциональная схема такой модели показана рис.12.

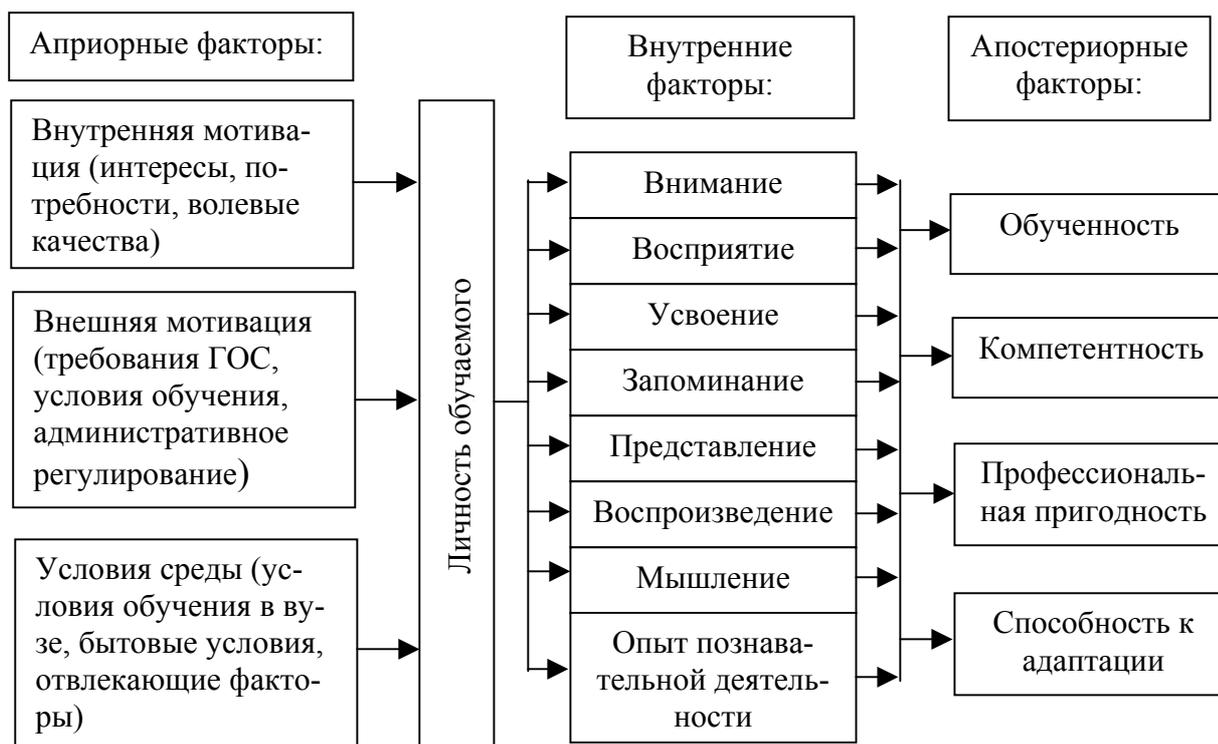


Рис. 12. Структурно-функциональная модель обучаемого

Под мотивацией традиционно понимаются побуждения, вызывающие активность и определяющие её направленность. Мотивация оказывает влияние на профессиональное самоопределение, на удовлетворённость человека своим трудом. Профессиональная мотивация – это действие конкретных побуждений. Которые обуславливают выбор профессии и продолжительное выполнение обязанностей, связанных с выбранной профессией. Профессиональная ориентация динамична и представляет собой непрерывный процесс, протекающий под постоянным воздействием объективных и субъективных факторов.

При оценивании внутренней мотивации следует иметь в виду, что интересы являются важной составной частью направленности личности, её мотивационной сферы и представляют собой форму проявления потребностей. Своеобразие интересов заключается в том, что они выражают познавательное отношение к предмету, тогда как потребности выражают тенденцию к реальному овладению им. Под профессиональным интересом понимают избирательную на-

правленность личности к овладению навыками той или иной профессии. При этом необходимо различать заинтересованность и устойчивый профессиональный интерес, различие между которыми заключается в объёме и содержании информации о выбираемой профессии. Профессиональный интерес, кроме того, характеризуется большей направленностью на сущностные стороны профессии.

Современное образовательное пространство характеризуется новой парадигмой образования, основанной на индивидуализации и дифференциации, вариативности и альтернативности образовательных систем, её прогностичности и адаптивности к изменяющимся социально-экономическим условиям, интересам и способностям обучаемых. При этом объективно существует ряд противоречий:

- между требованиями современной социокультурной ситуации к сформированности методологического мышления специалистов и сложившейся системой их подготовки;
- между требованиями к образованию как основе становления специалиста нового типа и стереотипным построением образовательного процесса;

Для преодоления обозначенных противоречий необходима разработка модели специалиста, обеспечивающей устойчивые связи между целями, содержанием, условиями, средствами и результатами образовательного процесса.

Все содержательные критерии эффективности обучения в вузе относительно независимы, но вместе с тем взаимосвязаны многозначными соотношениями. Это единство разнообразных содержательных критериев оценки психологических результатов обучения находит своё выражение в таком интегральном, формализованном и едином для всех учебных заведений критерии, как академическая успеваемость. Она определяется с помощью педагогической оценки, выраженной в числовой форме – отметке. Отметка, как показатель успеваемости, в большей мере фиксирует сам результат. Когда же хотят выделить ещё и способы обучения, позволившие достигнуть тот или иной результат – высокую или низкую успеваемость, применяют понятие учебной успешности,

впервые введённое Б.Г. Ананьевым. Профессиональная компетентность, профессиональная направленность, учебная активность, умственная самостоятельность, академическая успеваемость и учебная успешность – все эти разнообразные по содержанию и форме критерии можно отнести к внутренним критериям эффективности обучения. Внутренние критерии эффективности работы учебных заведений должны дополняться внешними. Для технических вузов такими внешними критериями могут быть: адаптация выпускников на местах их профессиональной деятельности, профессиональная устойчивость и миграция, темпы роста профессионального мастерства и связанное с ним должностное продвижение (служебная карьера) и др. Эти общие критерии эффективности при реализации каждой конкретной педагогической технологии на разных уровнях её функционирования могут дополняться и конкретизироваться более частными и разнообразными по содержанию критериями, служащими для оценки локальных, промежуточных и конечных результатов обучения.

Анализ существующих психологических тестов показал, что в рамках имеющихся возможностей (отсутствие централизованной вузовской системы сбора и обработки статистической информации; отсутствие нормативных документов, определяющих статус психологического тестирования в вузе и т.д.), целесообразно на первом этапе использовать четыре теста: Кейрси, Мехрабина, Айзенка и Спилбергера /58, 59/.

При проведении психологического тестирования и обработке его результатов применяются те же методы диагностирования, но в качестве диагнозов выступают психологические типы личностей, а в качестве диагностических признаков - поступки людей с тех или иных ситуациях. Если нельзя наблюдать и фиксировать сами поступки, то о многом можно судить по намерению их совершать или не совершать. Для этого в тесте предлагается ситуационная модель, к которой люди с различным психологическим типом будут действовать по-разному. Строится диагностическая матрица, и определяются апостериорные вероятности психологических диагнозов. Как выполняются подобные дей-

ствия, мы рассмотрели выше. Заслуга психологов-исследователей, что они произвели дискретизацию и классификацию свойств личности человека, определили систему балльных оценок, собрали и обработали огромный статистический материал: частоты наблюдения диагностических признаков в том или ином диагнозе. Мы можем воспользоваться большей частью этих достижений. С точки зрения теории измерения всё множество различных измерительных процедур, применяемых в психологии, является процедурами построения шкалы психологической переменной, иначе говоря, процедурами психологического шкалирования.

Тест Кейрси позволяет диагностировать наличие (отсутствие) и получить количественные оценки восьми качеств личности:

E (Extraversion) - экстраверсия,	I (Introversion) - интроверсия
S (Sensation) - здравомыслие,	N (iNtuition) - интуиция,
T (Thinking) - логичность,	F (Feeling) - чувство,
J (Judging) - рассудительность,	P (Perceiving) - импульсивность.

В результате обработки диагностической информации получаем один из 16 возможных психологических портретов: ISNP, ESTP, ESFP, ISFP, ISFJ, ESFJ, ESTJ, ISTJ, INFP, ENFP, INFJ, ENFJ, INTP, ENTP, INTJ, ENTJ с числовыми значениями, характеризующими яркость проявления тех или иных качеств. Значения 0,1-0,3 яркость характеризуют как низкую, 0,4-0,7 - как среднюю и 0,8-1 - как высокую. Суммарное значение всех четырёх показателей позволяет разделить все личности на «яркие» (сумма равна 2-4) и «неяркие» (сумма - 0-2). Уровень яркости личности характеризует устойчивость сохранения психологического портрета во времени. Набор буквенных символов может быть охарактеризован следующим образом:

ESFP - оптимизм и теплота, избегают одиночества, идут по жизни смеясь, жизнь - сплошные приключения, игнорируют всё мрачное, щедры, легко поддаются соблазнам, умеют работать с людьми, хорошо подвешен язык, хорошо проявят себя в бизнесе, торговле, наука - не для них;

ISFP - свойства личности проявляются в искусстве, эпикурейский образ жизни, острота ощущений текущей минуты, высокая чувствительность к оттенкам и полутонам, свобода, оптимистичность, уход от всякого рода ограничений, области приложения способностей - музыка, танцы и рисование, проблемы - с устной и письменной речью;

ESTP - энергия, игра, неистощимый на выдумки, остроумный, искусственный в обращении с людьми, успешная работа в условиях риска и на грани катастрофы, поиск острых ощущений;

ISTP - не признают субординации, жажда действия, бесстрашие, умение обращаться с инструментом, хорошо водят автомобиль, боевые искусства, братские взаимоотношения, часто бросают учёбу и не стремятся к высшему образованию;

ESFJ - открытый и практичный, житейская мудрость, компанейский, гостеприимный, общительный, ответственный;

ISFJ - спокойный и ответственный, интересы организации - превыше всего, традиции, связь времён, хозяин в доме, выполнять поручения лучше, чем руководить;

ESTJ - лидер, ответственный, долг, иерархия, порядок, практичный, открытый, всё по плану и без лишних выдумок, бесхитростный, исполнительный, цельная натура;

ISTJ - человек чести и слова, долг, ответственный, спокойный, твёрдый, надёжный, логичный, малоэмоциональный, основательность и детальность;

ENFJ - лидер, общительный, внимательный к чувствам других людей, нетерпеливый по отношению к рутине и монотонной деятельности, умение распределять роли в группе;

INFJ - проницательность и прозорливость, успешное самообразование, ранимость, не любит споров и конфликтов, радость друзей - радость и для него, богатое воображение, поэтичность, склонность к метафорам, гармония человеческих взаимоотношений;

ENFP - умение влиять на окружающих, видит людей насквозь, отрыв от реальности в поиске гармонии, чувствительность, умение подмечать необычное, чувствительность, отрицание формальной логики, оптимизм, творчество, энтузиазм, богатая фантазия, экстравагантность, щедрость (иногда до безрассудства);

INFP - спокойный, идеалист, чувство собственного достоинства, борьба со злом за идеалы добра и справедливости, лирический символизм, крайне уживчивый и покладистый, способности в изучении языков, противопоказано занятие бизнесом;

ENTJ - руководитель, ориентация на достижение целей, логичный, эффективность в работе и карьера - превыше всего, интеллигент, требовательный, неутомимый;

INTJ - самоуверенный, интересы в будущем, авторитеты не имеют значения, теоретик, логика, высокие руководящие должности, отсутствие эмоциональности, высокая способность к обучению, независимость, интуиция, проблемы с сочувствием;

ENTP - изобретатель, применяет интуицию на практике, энтузиаст, важна воплощённая идея, а не идея сама по себе, хороший собеседник, инициативный в общении, нетерпелив к рутине, юмор, стремление понимать людей;

INTP - ценитель мыслей и речи, мгновенная оценка ситуации, логичность, познание законов природы, интеллектуал, несколько высокомерный, философ, генератор новых идей, сложный внутренний мир.

Д. Кейрси определил совместимые и несовместимые типы личностей и предложил систему группировки их в 4 базовых типа темперамента:

SP - ESFP, ISFP, ESTP, ISTP;

SJ - ESFJ, ISFJ, ESTJ, ISTJ;

NF - ENFP, INFP, ENFJ, INFJ;

NT - ENTP, INTP, ENTJ, INTJ.

Для базовых темпераментов Д. Кейрси установил особенности поведения, которые с успехом могут быть учтены в педагогической практике. Он же установил противоположные психологические портреты, которые формируются в пары следующим образом:

INTP \leftrightarrow ESFJ,	ENTP \leftrightarrow ISFJ,
INTJ \leftrightarrow ESNP,	ENTJ \leftrightarrow ISFP,
INFP \leftrightarrow ESTJ,	ENFP \leftrightarrow ISTJ,
INFJ \leftrightarrow ESTP,	ENFJ \leftrightarrow ISTP.

Если у преподавателя почему-либо не налаживается контакт с каким-либо студентом, необходимо обратить внимание на их взаимную психологическую совместимость. Для этого, правда, потребуется знать также и психологический портрет преподавателя. Вероятно, скоро наступят времена, когда психологическое обследование профессорско-преподавательского состава будет проводиться не реже их флюорографического обследования. Ещё одна рекомендация: при выявленной несовместимости не пытаться «переломить» студента, заставить его быть похожим на преподавателя. Т. Элиот утверждал: «Нет человека, достойного присвоить себе право переделывать другого по своему образу и подобию». Лучше попытаться использовать в своих целях лучшие стороны личности учащегося, найдя к нему особый, индивидуальный подход.

Тест Мехрабина служит для диагностирования уровня поведенческой мотивации. Различают два вида мотиваций: МУ - мотивация на достижение успеха, МН - мотивация на избежание неудач. Тест имеет две формы: (М) - для мужчин и (Ж) - для женщин.

После обработки результатов тестирования получим показатель уровня мотивации. Его высокие значения ($> 0,67$) говорят о мотивации на успех, низкие ($< 0,33$) - о мотивации на избежание неудачи. Исследования психологов доказывают, что более ценным для развития личности является стремление избегать неудачи. Ожидания успехов в большинстве случаев оказываются несостоятельными, оканчиваются разочарованием, утратой стимулов. Умение избегать не-

удачи, напротив, стимулирует деятельность. В военной науке также декларируется, что тот, кто сумел избежать поражений просто «обречён» на успех, на победу.

Тест Айзенка позволяет диагностировать четыре типа темперамента: холерик, сангвиник, флегматик и меланхолик.

Таблица 12

Психологические типы, выявляемые с помощью теста Айзенка

		Нестабильный тип AN = 1			
Интровертированный AE = 0	Легко расстраивающийся	<i>Меланхолик</i>	Чувствительный	<i>Холерик</i>	Экстравертированный AE = 1
	Тревожный		Беспокойный		
	Ригидный		Агрессивный		
	Склонный к рассуждениям		Возбудимый		
	Пессимистичный		Изменчивый		
	Сдержанный		Импульсивный		
	Необщительный		Оптимистичный		
	Тихий		Активный		
		<i>Флегматик</i>	<i>Сангвиник</i>		
Интровертированный AE = 0	Пассивный		Общительный		Экстравертированный AE = 1
	Осмотрительный		Контактный		
	Рассудительный		Разговорчивый		
	Доброжелательный		Отзывчивый		
	Миролюбивый		Непринуждённый		
	Управляемый		Жизнерадостный		
	Ровный		Не склонный к беспокойству		
	Спокойный		Склонный к лидерству		
		Стабильный тип AN = 0			

Выявление экстраверсии/интроверсии, а также типов темперамента личности может осуществляться на основе использования адаптированного варианта опросника Г. Айзенка. После обработки результатов получим два значения параметров AE и AN (показатели экстраверсии и нейротизма по Айзенку), значения которых и позволят не только идентифицировать тип темперамента, но и оценить выраженность типичных свойств выявленного типа. Попробуем представить себе реакцию на отрицательный результат тестирования для типичных представителей всех четырёх типов.

Сангвиник постарается узнать возможность повторного тестирования, поинтересуется правильными ответами на задания, с которыми ему не удалось справиться в ходе выполнения теста.

Холерик начнёт ворчать, что времени для ответа недостаточно, вопросы содержат ошибки, предложенную задачу он решил правильно, но почему-то зачтена она не была.

Флегматик подумает, что большой трагедии не случилось, тест не сдало ещё очень много тестируемых, будет передача теста и тогда его результат наверняка может стать лучше.

Меланхолик расстроится, что очередной раз в жизни не повезло, так что учи или не учи, всё равно толку от этого будет не много.

Тест Спилбергера позволяет оценить уровни личной (сиюминутной) и ситуативной (долговременной) тревожности. Этот показатель характеризует степень «психологической комфортности». Высокая тревожность обычно затрудняет эффективную деятельность по выполнению заданий предметного теста, низкая тревожность требует вмешательства для пробуждения чувства ответственности и заинтересованности в положительном результате тестирования.

Распределения значений коэффициентов психологического портрета, полученные по данным результатов психологического обследования студентов СПбГПУ показаны на рис. 13 – 14.

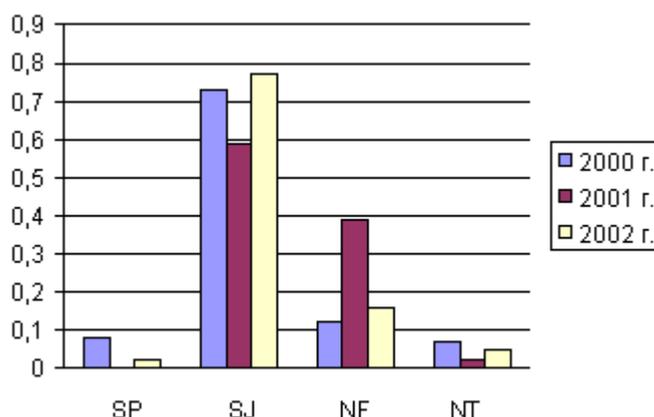


Рис. 13. Результаты прохождения психологического теста Кейрси

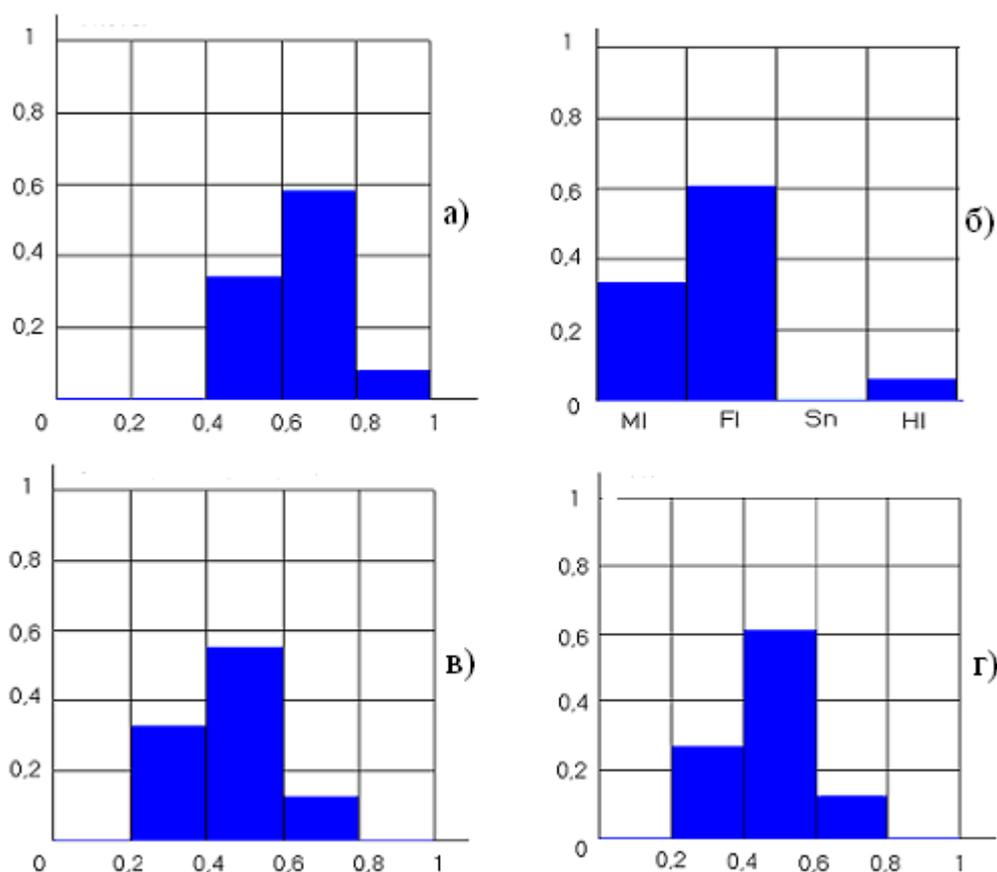


Рис. 14. Распределения значений коэффициентов мотивации (а), психологических типов Меланхолик-Флегматик-Сангвиник-Холерик (б), ситуативной (в) и личностной (г) тревожности

9. Анкетирование

Тестирование не только позволяет объективно оценивать результаты учебной деятельности, но принципиально изменяет характер взаимоотношений и взаимодействия администрации образовательных учреждений (деканат, ректорат), субъектов (преподавателей) и объектов (студентов) учебного процесса. Преподаватель из строгого судьи превращается в доброго наставника, заинтересованного в высоких результатах прохождения тестов своими подопечными. С переходом на коммерческое обучение заказчиками и потребителями образовательных услуг становятся студенты. Их мнение о качестве учебного процесса

должно учитываться и приниматься во внимание. Выявление мнения по тем или иным аспектам деятельности образовательных учреждений целесообразно проводить путём анкетирования.

Несколько общих правил для составления опросных листов анкет, выработанных в ходе организации и проведения анкетирования:

- должна быть сформулирована цель анкетирования, и должно быть ясное представление о том, какая информация нужна;
- вопросы должны быть краткими и простыми, не содержать двусмысленностей;
- ответы на вопросы также должны быть просты, отвечающий должен быть в состоянии на них ответить; для неопределённых вопросов следует предусматривать эталон ответа «не знаю»;
- вопросы личного характера могут быть заданы анкетизируемому, после того как с помощью нескольких простых вопросов более общего характера завоёвано некоторое доверие опрашиваемого;
- логика следования вопросов должна помогать опрашиваемому управлять собственными мыслями, сделает его ответы более разумными и последовательными;
- следует иметь в виду, что анкетизируемые не всегда говорят правду полностью или хотя бы частично, что иногда может быть отнесено к разряду мысленной оговорки, иногда – к особенностям характера (хвастовство, чрезмерная стыдливость и др.);
- часто ответы являются стремлением действовать в рамках принятых или предполагаемых норм поведения, и поэтому желательно, чтобы опрашиваемый чётко понимал каждый вопрос, но он не должен предугадывать точную цель обследования;
- более надёжный ответ может вызвать косвенный вопрос, например, вместо вопроса, предпринимали ли опрашиваемые такое-то действие, можно спросить, когда они это делали;

- в ответах на вопросы следует стремиться в меньшей степени полагаться на память опрашиваемого: процедура анкетирования обычно не предполагает активного напряжения памяти.

Анкета может иметь «жесткую» форму, требующую однозначного выбора (один из нескольких), и «мягкую» форму, предполагающую выбор нескольких из числа предложенных эталонов ответов (несколько из нескольких) и допускающую пропуск или указание всех эталонов. В силу последних обстоятельств при «мягкой» форме нарушаются прямые пропорциональные сравнения, распределения ответов затруднительно или практически невозможно выразить в долях или процентах от целого.

Возможные темы анкетирования зависят от целей его проведения, например, профориентация, оценка уровня и качества образования, обеспеченность учебными пособиями и др. Часто используется методика SWOT- анализа как средства диагностирования со следующих четырех позиций:

- *сильные стороны (Strengthes)* - внутренние положительные качества объекта или предмета анкетирования;
- *слабые стороны (Weaknesses)* - внутренние отрицательные качества;
- *возможности (Opportunities)* - внешние факторы, улучшающие перспективы развития;
- *угрозы (Threatneses)* - внешние факторы, которые потенциально могут подорвать ожидаемый будущий успех.

Например, результаты опроса, проведенного в СПбГПУ, показали, что 51% опрошенных устраивает работа деканатов, 27% не обращались в деканат, поэтому не имеют представления о его работе. Не посещают лекции 3% студентов, а 30% посещают их выборочно. Перспективной считают выбранную специальность 68% опрошенных, причём 20% из них считают, что полученных знаний будет в будущем недостаточно.

10. Система компьютерного тестирования

Полномасштабная реализация изложенных выше подходов к диагностированию возможна только на основе использования информационных технологий с широким применением средств вычислительной техники. Анализ функциональных возможностей зарубежных и отечественных программ, используемых для подготовки банков тестовых заданий, для компоновки компьютерных тестов, для проведения сеансов компьютерного тестирования и обработки его результатов, а также отечественный опыт разработки «Комплексной технологии компьютерного тестирования» и использования ее в централизованном компьютерном тестировании абитуриентов в России в 2001-2002 гг. более чем в девяноста вузах, позволили сформулировать основные требования, которым должны удовлетворять современные системы компьютерного тестирования /25, 46/.

Среди требований к системе создания банков тестовых заданий и разработки компьютерных тестов - обеспечить использование в тестовых заданиях текста и графики, работу с заданиями, в которых тестируемый в качестве правильного ответа выбирает подходящую альтернативу (альтернативный выбор), либо отмечает все подходящие элементы множества (множественный выбор), либо вводит набор символов (свободный ответ). Должны применяться разные алгоритмы выбора следующего задания - линейный, блочный, случайный, адаптивный.

Сеансы тестирования должны проводиться на автономных или объединенных в локальную сеть компьютерах. При этом должны использоваться парольный метод запуска тестов и подсистема информационной безопасности. Обязательными являются требования регистрации данных о тестируемом, возможность его анкетирования, итоговый показ количества представленных заданий, количества использованных минут, количества правильных ответов, количества набранных баллов по 100-балльной шкале.

При передаче данных через Интернет необходимо иметь специальные программные средства, позволяющие учитывать «специфику» каналов передачи данных и обеспечивающие необходимый уровень защиты передаваемых данных. Система оперативного статистического анализа результатов компьютерного тестирования должна обеспечивать выбор объекта статистической обработки (группа, поток, курс, факультет, направление подготовки, вуз). Получение для этого объекта следующих данных: количество тестируемых, их распределение по оценкам и по количеству набранных баллов, процент правильных ответов на вопрос теста. Сравнительный анализ результатов тестирования различных объектов статистической обработки и т.д. Система статистического анализа качества тестовых заданий по их статистическим параметрам должна обеспечивать анализ «траекторий» сеансов компьютерного тестирования и формировать комплекты статистических форм. Система обработки апелляций на результаты компьютерного тестирования должна обеспечивать полный анализ выполнения тестируемым каждого задания компьютерного теста.

Важными требованиями для современной системы компьютерного тестирования являются комплексность ее технологии, апробация ее программного обеспечения в массовых процедурах компьютерного тестирования, обоснованность научно-методических решений и их патентная защищенность, мобильность, наличие готовых комплектов репетиционных и экзаменационных компьютерных тестов, наличие документации и т.д.

На основании рассмотрения подходов к принципам организации диагностирования могут быть сформулированы основные требования к инструментальным средствам системы тестирования:

- автоматизация создания, сбора, хранения и использования информации реализуется на основе компьютерных методов представления и обработки информации;
- для организации данных используется технология баз данных;

- доступ к данным обеспечивается посредством интерфейса ADO (ActiveX Data Objects) и средств UDA (Universal Data Access) с помощью OLE DB;
- для наполнения и обработки данных используется база данных Access, входящая в состав Microsoft Office;
- обеспечивается сетевой доступ к данным;
- защита информации производится на уровне парольной защиты баз данных Access;
- тестовые задания должны иметь текстовую, графическую и звуковую (аудио) формы представления;
- порядок предъявления заданий тестируемому рандомизирован или устанавливается с помощью SQL-запросов;
- система компьютерного тестирования должна представлять собой пакет сервисных и утилитных программ, являющихся приложениями, которые работают под управлением операционной системы Windows;
- функции проведения тестирования, формирования базы данных и обработки результатов тестирования относятся к компетенции различных программ;
- программа компьютерного тестирования должна позволять производить предметное и психологическое тестирование, а также анкетирование;
- режим работы программы компьютерного тестирования должен задаваться с помощью объектов базы данных.

Заключение

Отметим здесь несколько общих правил, следование которым помогут успешнее решать проблемы разработки тестов и использования тестовых технологий в учебном процессе.

1. Не следует путать методологии процесса обучения и процедуры контроля его результатов. В первом случае допускается и даже приветствуется многообра-

зие форм, средств и методов осуществления педагогического процесса: в ходе обучения студент обязан проявлять любознательность, уметь хорошо говорить и аргументировать свою точку зрения, владеть риторикой, логикой, демонстрировать способность рассуждать и т.п. Тестирование же, как одна из форм контроля - сугубо формальная диагностическая процедура: эмоциональность, многословие (или, напротив, молчание), субъективные суждения могут только затруднить получение достоверной информации. Во многих случаях при тестировании бывает достаточным проставить галочки или крестики в опросных листах, то есть изложить своё мнение по каким-то вопросам в форме ответов типа «да/нет».

2. Тест должен быть объективен, т.е. независим от субъективных оценок, а также должен быть прозрачен для контроля со стороны независимых экспертов. На практике это может означать своего рода принцип «разделения власти»: преподаватель учит и должен быть морально готов к тому, что результаты обучения будет оценивать не он сам. Результаты же тестирования покажут, насколько хорошим было обучение.
3. Тест должен давать достаточно большую и достоверную статистическую информацию, анализ которой может служить основой для совершенствования учебного процесса. Тестирование не проводится ради «галочки» в отчёте или морального удовлетворения. Основная задача – выявить проблемы в подготовке и помочь с ними справиться в ходе обучения. Наибольший интерес для того, кто анализирует результаты тестирования, представляют допущенные ошибки и причины их появления. Правильные ответы менее интересны и информативны.
4. Тестовые задания принципиально отличаются от заданий письменного и, тем более, устного экзамена. Необходимо научиться (и это не так просто) формулировать тестовые задания в корректной форме. Традиционные для устного экзамена вопросы, начинающиеся со слов «Как?», «Зачем?», «Почему?», для теста непригодны, так как требуют при ответе его аргументиро-

вать, т.е. высказать одну из возможных точек зрения по предлагаемому вопросу. При тестировании многозначность и многословность в ответе должны быть исключены.

5. Задания в тесте должны быть сформулированы в простой, корректной и лаконичной форме, не допускающей различного толкования. По статистике более половины тестовых заданий обычно, к сожалению, таковы, что студенту не понятно, что именно преподаватель хочет от него, каких действий ожидает. В качестве примера, не допустимого для использования в тесте, можно привести вопрос из Логического словаря справочника Н.И. Кондакова /30/ «Когда на Марс вступит первый житель Земли?». Полагаю, что ответы «Никогда, так как первый житель Земли давно уже скончался», «Не скоро», «15 августа 2032 года», «Через 8 лет» или «Утром, в 9 часов по местному марсианскому времени» и многие другие могут рассматриваться в качестве потенциально допустимых.
6. Не следует Вам пытаться с помощью Вашего теста оценивать общий уровень эрудиции или компетенции тестируемого, включая в тест задания смежных или ранее изученных дисциплин. Общий принцип таков: старайтесь контролировать только то, чему Вы научили сами. Справедливо и обратное: всё чему Вы научили должно быть обязательно проконтролировано. При этом повышается мотивация к обучению со стороны студентов, из курса сами собой исключаются «лишние» знания.
7. Все тестовые задания должны быть профессионально значимы (с точки зрения будущей профессии тестируемого), по объёму и содержанию удовлетворять требованиям необходимости и достаточности. Тест - это визитная карточка Вашей дисциплины. По тестовым заданиям, их объёму и содержанию, можно судить о том, сколько нового, полезного и интересного студенты узнали от Вас. По результатам тестирования отдельных студентов Вы можете сравнивать уровень их знаний, определять, в чём состоят их проблемы и ин-

дивидуализировать учебный процесс. Тест контролирует и измеряет то, чему и насколько хорошо мы учим.

8. Не существует учебных дисциплин, для которых тестовый опрос не возможен или не эффективен. Он должен наряду с лекциями, упражнениями, семинарами, лабораторными работами, зачётами и экзаменами стать одним из неперменных атрибутов учебного процесса.

ЛИТЕРАТУРА

1. Аванесов В.С. Вопросы объективизации оценки результатов обучения. Обзорная информация. - М.: НИИ проблем высшей школы, 1976. - 66 с.
2. Аванесов В.С. К вопросам истории и теории тестов. /www.oim.ru
3. Аванесов В.С. Централизованное тестирование лучше единого государственного экзамена. Тезисы докладов Всероссийской научно-методической конференции «Развитие тестовых технологий в России» /Под ред. Л.С. Гребнева. – Центр тестирования Министерства образования РФ, 2002. с.71-73.
4. Беспалько В.П. Не пора ли менять стратегию образования? – Педагогика № 9, 2001.
5. Беспалько В.П. Основы теории педагогических систем. - Воронеж: изд. ВГУ, 1977. - 304 с.
6. Беспалько В.П. Слагаемые педагогической технологии. - М., 1989. – 153 с.
7. Биргер И.А. Техническая диагностика. – М.: Машиностроение, 1978 .- 240 с.
8. Битинас Б.П. Педагогическая диагностика: сущность, функции, перспективы. //Сов. педагогика. - 1993, №12, с. 3 - 13.
9. Бычков А.И. Компьютерный учёт знаний учащихся школы как средство повышения качества образования. / Проблемы качества образования. Книга 2. Оценка качества подготовки выпускников образовательных учрежде-

- ний. // Материалы XI Всероссийской научно-методической конференции. - М.; Уфа: Исследовательский центр проблем качества подготовки специалистов, 2001. - 30 с.
10. Гершунский Б.С. Педагогическая прогностика: методология, теория и практика. - Киев. 1986. – 197 с.
 11. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976.- 495 с.
 12. Голицин Н.Н. Исторические этюды Древней Греции. Пифагор // Москвитянин. 1855. №19-20. Кн.1-2
 13. Гомоюнов К.К. Совершенствование преподавания технических дисциплин. Методологические аспекты анализа учебных текстов. – Л. Изд-во Ленингр. Ун-та, 1983. – 206 с.
 14. Горский Д.П. Вопросы абстракции и образования понятий. М., 1961. – 351 с.
 15. Гришанова Н.А. Проблема целеполагания в государственных образовательных стандартах высшего профессионального образования. / Проблемы качества образования. Книга 1. Качество высшего образования как объект системного исследования. Образовательные стандарты и качество образования. // Материалы XI Всероссийской научно-методической конференции. - М.; Уфа: Исследовательский центр проблем качества подготовки специалистов, 2001. - 74 с.
 16. Гутник И.Ю. Педагогическая диагностика образованности школьника (Теория. История. Практика): Учебное пособие. – СПб.: Издательство РГПУ им. А.И. Герцена, 2000. – 158 с.
 17. Дандамаев М.А. Вавилонские писцы. М.: Наука, 1983. -245 с
 18. Диагностика как фактор развития образовательной системы. Учебное пособие. /Авторский коллектив: Г.А. Деминова, Т.И. Дормидонова, А.В. Кулев, В.Н. Максимова, А.Д. Храмцова. - СПб., 1995.

19. Диагностика познавательных и профессиональных способностей: Сб. научных трудов. // Отв. ред. В.Д. Шадриков. - М., 1988.
20. Диагностика способностей и личностных черт учащихся в учебной деятельности / Под ред. В.Д. Шадрикова. - Саратов, 1989.
21. Жуков Р.Ф. Пути развития активных методов обучения в университете. Сб. науч. тр. Технология акмеологических методов обучения / Отв. ред. Р.Ф. Жуков. - СПб.: СПбГИЭУ, 2000. - 160 с.
22. Загвязинский В.И. Методология и методика педагогических исследований. - М., 1982. - с. 160.
23. Закон Российской Федерации "Об образовании". Постановление Верховного Совета РФ от 10.07.1992, № 3267-1 - М., 1992. - 57 с.
24. Иванов Б.С. Педагогические измерительные материалы. Материалы научно-методологической конференции "Эдукология: природа, проблемы, пути и методы её решения" (6 сессия) 25-28 апр. 2000.
25. Иванов Б.С. Педагогические основы и программное обеспечение компьютерной тест-системы. Материалы 19-й международной конференции "Школьная информатика и проблемы устойчивого развития" 19.04.2000.
26. Иванов Б.С., Тихомиров С.А. Контроль результатов учебной деятельности и управление качеством образовательного процесса. Материалы международной научно-практической конф. "Формирование профессиональной культуры специалистов XXI в. в техническом университете, СПбГТУ, 26-28 марта 2001 г.
27. Иванов Б.С., Жуков В.А. Статистические показатели тестирования. Труды научно-методической конференции ММФ СПбГТУ 29 янв.2002, СПбГТУ.
28. Ингенкамп К. Педагогическая диагностика. - М.: Педагогика, 1991.
29. Кальней В.А., Шишов С.Е. Технология мониторинга качества обучения в системе «учитель - ученик». - М.: Педагогическое общество России, 1999.
30. Кондаков Н.И. Логический словарь-справочник. 2-е изд. испр. и доп. М., «Наука», 1975. – 717 с.

31. Концепция, структура и содержание многоуровневой системы образования. /Под ред. Васильева Ю.С., Козлова В.Н. - СПб.: СПбГТУ, 1993.- 202 с.
32. Концепция оценки достижения учащимися требований общеобразовательного стандарта. Временный государственный образовательный стандарт. - М., 1993.
33. Кузнецов А.А. Проблемы оценки достижения требований образовательных стандартов // Образовательные стандарты и контроль качества образования. - Вологда, 1966. - с. 60-66.
34. Куклин В.Ж., Наводнов В.Г., Петропавловский М.В. КАМЕРТОН - технология проведения тестирования и анализа результатов. - Йошкар-Ола, 1995.
35. Куклин В.Ж., Масленников А.С., Наводнов В.Г. Технология разработки педагогических испытательных материалов // Прикладные исследования в электронике и новые технологии в обучении студентов. Йошкар-Ола, 1996.- с. 44 - 46.
36. Куклин В.Ж., Мешалкин В.И., Наводнов В.Г., Савельев Б.А. О компьютерной технологии оценки качества знаний. // Высшее образование в России. 1993.- №3.-с.146-153.
37. Майоров А.Н. Мониторинг в образовании. - СПб.: Образование и культура, 1998.
38. Майоров А.Н. Тесты школьных достижений: конструирование, проведение, использование. - СПб.: Образование и культура, 1996.
39. Майоров А.Н., Сахарчук Л.Б., Сотов А.В. Элементы педагогического мониторинга и региональных стандартов в управлении. - С.-П., 1992.
40. Макарова Т.Д. Обеспечение итогового контроля. // Образовательные стандарты и контроль качества образования. - Вологда, 1996. - с. 67-77.
41. Матрос Д.Ш., Полев Д.М., Мельникова Н.Н. Управление качеством образования на основе новых информационных технологий и образовательного мониторинга. - М.: Педагогическое общество России, 1999.

42. Минияров В.М., Бердников И.Г. Педагогическая диагностика личности школьника. СГПИ, Самара. 1993. - 98с.
43. Монтэ П. Египет Рамсесов. \ Пер. с франц. Ф.Л. Мендельсона .- Смоленск: Русич, 2000. – 416 с.
44. Наводнов В.Г., Петропавловский М.В., Ельцын А.В. Автоматизированное проектирование педагогических измерительных материалов: Препринт № 2/97. - Йошкар-Ола, 1997.
45. Налимов В.В. Вероятностная модель языка. – 2-е изд. перераб. и доп. – М., Наука, 1979, 303 с.
46. Нардюжев В.И., Нардюжев И.В. Требования к системам компьютерного тестирования. Тезисы докладов Всероссийской научно-методической конференции «Развитие тестовых технологий в России» /Под ред. Л.С. Гребнева. – Центр тестирования Министерства образования РФ, 2002. с.284-285.
47. О сертификации педагогических испытательных материалов / Куклин В.Ж., Масленников А.С. Наводнов В.Г., Савельев Б.А. // Квалиметрия человека и образование: методология и практика: Тез. докл. - М., 1996.- с.134-135.
48. Образовательное тестирование в Российской Федерации // Проект реформирования образования с участием Мирового Банка (Авторский коллектив: Болотов В.А., Полле А.Б., Королёв М.Ф., Панов В.Б., Решетников Н.Н., Шадриков В.Д., Шишов С.Е., Родионов Б.У., при участии Беспалько В.П. и др.) - М., 1995. - 36 с.
49. Образовательные стандарты // Материалы международной конференции. - Вологда, Вологодский ИПК и ППК, 1995. - 131 с.
50. Образовательные стандарты и контроль качества образования. - Вологда, 1996. - 100 с.

51. Образовательные стандарты в Российской Федерации. Материалы по разработке национально-региональных компонентов государственного образовательного стандарта. - Якутск, 1996. - 138 с.
52. Отчет Учебно-методического объединения и Научно-методического совета за 2000 год. СПб.: Изд-во СПбГТУ, 2001.- 192 с.
53. Оценка качества подготовки выпускников средней (полной) школы по информатике. / Сост. А.А. Кузнецов, Л.Е. Самовольнова, Н.Д. Угринович. - М.: Дрофа, 2001. - 64 с.
54. Оценка качества подготовки выпускников средней (полной) школы по физике. / Сост. В.А. Коровин, В.А. Орлов. - М.: Дрофа, 2001. - 192 с.
55. Педагогическая энциклопедия. М., 1976. В 2-х т.
56. Платон. Сочинения. В 3-х т. М.: Мысль, 1971.
57. Подготовка и проведение учебных курсов в заочно-дистанционной форме обучения: Метод. рекомендации преподавателям. Т.П. Зайченко, В.В. Зверев, Б.С. Иванов и др. / Под ред. И.А. Цикина. СПб.: Изд-во СПбГТУ, 2000. -126 с.
58. Практикум по общей, экспериментальной и прикладной психологии: Учеб. пособие / В.Д. Балин, В.К. Гайда и др. // Под общ. ред. А.А. Крылова, С.А. Маничева. - СПб.: Изд-во "Питер", 2000.- 560 с.
59. Практическая психология: Учебник /М.К. Тутушкина, С.А. Волков и др. Под ред. д-ра психол. наук, проф., акад. БПА М.К. Тутушкиной - 3-е изд., перераб., доп. - СПб.: Изд-во "Дидактика Плюс", 2000.- 336 с.
60. Примерная методика обобщения и анализа информации о качестве образования в вузе и системе его обеспечения // Серия: Нормативно-методическое обеспечение мониторинга качества образования в России. / Под научной редакцией д-ра техн. наук, проф., Н.А. Селезнёвой, д-ра экон. наук, д-ра филос. наук А.И. Субетто. - М.: Исследовательский центр по проблемам качества подготовки специалистов, 2001. - 21 с.

61. Примерная методика организации структурно-функционального анализа деятельности вуза // Серия: Нормативно-методическое обеспечение мониторинга качества образования в России. / Под научной редакцией д-ра техн. наук, проф., Н.А. Селезнёвой, д-ра экон. наук, д-ра филос. наук А.И. Субетто. - М.: Исследовательский центр по проблемам качества подготовки специалистов, 2001. - 33 с.
62. Психологические механизмы целеобразования. // Под ред. О.К. Тихомирова. - М.: Наука, 1977. - 260 с.
63. Равен Дж. Педагогическое тестирование: Проблемы, заблуждения, перспективы /Пер с англ. Изд. 2-е испр. – М.: «Когито-центр», 2001. – 142с.
64. Радионов Б.У., Татур А.О. Стандарты и тесты в образовании. - М., 1995.
65. Ришар Ж.Ф. Ментальная активность. Понимание, рассуждение, нахождение решений. Сокр. пер. с франц. Т.А. Ребенко. М.: Изд-во «Институт психологии РАН», 1998. - 232 с.
66. Сборник нормативно-методических документов по высшему профессиональному образованию. СПб.: Изд-во СПбГТУ, 2001.- 182 с.
67. Симонов В.П. Педагогический менеджмент. - М.: Педагогическое общество России, 1999. – 427 с.
68. Сухомлинский В.А. О воспитании. М., Политиздат, 1973. – 272 с.
69. Татур Ю.Г. Основные положения, определяющие создание фондов оценочных средств для итоговой государственной аттестации выпускников вузов на соответствие требованиям ГОС ВПО. / Проблемы качества образования. Книга 2. Оценка качества подготовки выпускников образовательных учреждений. // Материалы XI Всероссийской научно-методической конференции. - М.; Уфа: Исследовательский центр проблем качества подготовки специалистов, 2001. - 30 с.
70. Уланов В.Я. Опыт методики истории в начальной школе. М. 1914. - с. 114-115.

71. Фонды комплексных квалификационных заданий по специальностям высшего образования. Методические рекомендации по разработке /Под ред. В.П. Беспалько и Н.А. Селезневой. - М.: Исследовательский центр, 1989.
72. Хинтиikka И. Вопрос о вопросах. В сб. Философия в современном мире. М. Изд-во «Наука», 1974. – с.303-362.
73. Челышкова М.Б. Разработка педагогических тестов на основе современных математических моделей. - М.: Исследовательский Центр, 1995.
74. Челышкова М.Б., Савельев Б.А. Методические рекомендации по разработке педагогических тестов для комплексной оценки подготовленности студентов в вузе. - М., 1995.
75. Шишов С.Е., Кальней В.А. Школа: мониторинг качества образования. – М.: Педагогическое общество России, 2000. – 320 с.
76. Шюре Э. Великие посвященные. Пер. с фр. Калуга: Лотос, 1914.- 419 с
77. Энгельс Ф. Диалектика природы. М., 1987. – 576 с.
78. Якунин В.А. Педагогическая психология: Учебн. пособие / Европ. ин-т экспертов. СПб.: Изд-во Михайлова В.А.: Изд-во "Полиус", 1998.- 639 с.

(с) Санкт-Петербургский государственный политехнический университет, 2003.

Подписано к печати 15.02.2003

Печ. л. 5

Уч.-изд.л. 5

Зак.

Типография СПбГПУ