И. И. Соколова

КОМПЬЮТЕРНОЕ ТЕСТИРОВАНИЕ КАК НАУКОЕМКАЯ ПЕДАГОГИЧЕСКАЯ ТЕХНОЛОГИЯ

В статье последовательно раскрываются подходы к современному пониманию компьютерной диагностики учебных достижений. Уделено внимание дидактическим основам диагностики результатов обучения. Приведены современные требования к форме и содержанию тестовых заданий, к процедуре компьютерного тестирования, анализируются особенности различных сред для тестирования. Обсуждаются статистические основы обработки результатов диагностики. Показано, что компьютерное тестирование можно рассматривать как современную наукоемкую педагогическую технологию.

Программа модернизации педагогического образования ставит задачи по преодолению отставания материально-технической базы и ресурсно-информационного обеспечения педагогических образовательных учреждений от современных требований, а также создания и внедрения современных средств контроля качества образования на всех уровнях и ступенях системы непрерывного педагогического образования. Одним из элементов системы качества образования является обеспечение мониторинга результатов образовательной деятельности на основе использования современных компьютерных сред и ресурсов.

Тем не менее, вопросы компьютерного тестирования остаются слабо разработанными на уровне их дидактических особенностей, недостаточно осознанны их возможности и ограничения педагогическим сообществом. Задача настоящей статьи — раскрыть и обсудить специфику компьютерного тестирования с целью диагностики результатов обучения студентов в вузе как наукоемкой педагогической технологии.

Наукоемкость современного тестирования обусловлена высоким научным уровнем применяемых в нем трех относительно независимых компонентов — современной базовой статистической теории, передовых компьютерных средств обеспечения тестирования и современной дидактической теории диагностики учебных достижений. Тем не менее, качество тестирования определяется и этими компонентами в совокупности и взаимосвязи. Адекватность оценивания в большой степени обеспечивается использованием современной статистики объектов нечисловой природы. Технологичность обеспечивается использованием современных компьютерных средств создания баз тестовых заданий, проведения тестирования, обработки, хранения результатов. Дидактическая эффективность определяется как первыми двумя компонентами, научный уровень которых обеспечивает оперативность, прогностичность, высокую чувствительность к коррекции, так и собственно педагогическими подходами, лежащими в основе составления тестовых заданий, тестов, банков тестовых заданий. Рассмотрим последовательно современные подходы к выбору эффективного тестологического инструментария.

Прежде всего, при выборе модели оценивания результатов образовательного процесса, диагностики учебных достижений обучающегося нужно иметь в виду, что оценивается сложный объект. Это такой объект, который характеризуется многомерностью, разнотипностью, отсутствием достаточной информации о его внутренних взаимосвязях. Как правило, отсутствует априорная информация о виде функций распределений параметров, характеризующих объект или совокупность объектов. Предположения о нормальности распределения, о линейной регрессионной зависимости, о марковости процесса, об n-мерности гене-

ральной совокупности и т. п. могут стать слишком сильными, и результаты обработки данных не будут отражать адекватно реальные характеристики объекта. Возникает проблема (особенно для малых выборок) статистической устойчивости, неоптимальности полученных решений. Моделируя сложные объекты, приходится оперировать разнотипными переменными. Перед статистикой встает задача вычисления подобия между описаниями практически не сравнимых величин, поиска подходящей метрики для определения типа согласованности признаков, характеризующих сложные объекты.

К сожалению, часто приходится констатировать, что исследования в области тестирования в основном опираются на параметрическую количественную статистику, построенную на гипотезе о том, что большинство значений признаков есть выборки из нормальных генеральных совокупностей. Тем не менее, можно считать доказанным многочисленными исследованиями, что нормальных распределений в исследуемой области не существует, как не существует здесь независимых случайных величин, суммирование которых предусматривается в вероятностной модели нормального распределения, и тем более — нормальных двумерных и многомерных распределений.

Дополнительную сложность в разработке современного статистического аппарата оценивания результатов образовательного процесса представляет тот факт, что в экономике, социологии, психологии, педагогике не все признаки, характеризующие многомерные объекты, могут быть измерены количественно, оценить их можно только качественно. Статистику объектов нечисловой природы называют непараметрической статистикой. В связи с этим направлением можно упомянуть о развитии только еще оформляющегося направления «статистики качеств», а также об интенсивно развивающемся аппарате «кластерного анализа», в котором разрабатываются методики решения задач классификации многомерных объектов по совокупности общих для этих объектов признаков.

Таким образом, само развитие статистики как современной математической науки интенсивно стимулируется необходимостью решать проблемы оценки качества объектов произвольной, в том числе нечисловой природы. Эта проблема отражена на страницах некоторых монографий[1, 2].

Компьютерное тестирование обладает рядом неоспоримых преимуществ перед другими его формами, что обусловлено следующими обстоятельствами:

- возможность использования различных видов наглядности (рисунки, кинофрагменты, компьютерные модели);
- компьютер больше, чем присутствие экспериментатора, побуждает испытуемого к самостоятельности, отсюда — выше диагностическая ценность результатов;
- можно судить об индивидуальных особенностях стратегии деятельности испытуемого, поскольку снижается влияние личностных особенностей экспериментатора и случайных поведенческих факторов на стратегию;
- появляется возможность сравнить экспериментальные данные, полученные разными исследователями;
- автоматизируется сколь угодно сложная обработка данных, появляется возможность быстро получить сравнимые результаты по большим массивам данных, сохранять и сравнивать их;
- достаточно легко совместить функции диагностики и коррекции знаний, например, в обучающих системах;
- появляется возможность применять такие формы, как имитационные упражнения, анализ ситуаций, тесты действия, связанные с тренажерами;
- возможно включать в сферу тестирования элементы психодиагностики, установок, чувств, впечатлений, влияния когнитивных стилей;

• возможно использовать адаптивную стратегию тестирования, когда стратегия исследования меняется в зависимости от полученных ранее результатов.

В компьютерном тестировании, как показывают исследования, обучающихся привлекает, прежде всего, равенство возможностей при установлении собственных достижений любого вида.

Наукоемкость компьютерного тестирования определяется высокими технологиями создания компонентов программного обеспечения всего цикла тестирования — от формирования тестовых баз с помощью комфортных для пользователя тестмейкеров с адекватными задачам тестирования возможностями по созданию различных форм тестовых заданий до обработки результатов тестирования и хранения их в удобной для преподавателя или экспертов форме. Выбирая программное обеспечение для решения задач тестирования, важно также определить, будут ли тестовые задания однородными по сложности, следует ли предусмотреть возможность изменения порядка их представления на соседних компьютерах в аудитории, где проводится тестирование, нужно ли ограничивать время на выполнение тестового задания, имеется ли возможность изменения шкалы оценивания преподавателем и др. С этой точки зрения, выигрышно выглядят среды, предоставляющие возможность адаптивного тестирования, когда компьютер выбирает тестовое задание в зависимости от предыдущего ответа испытуемого, например, — среда АСТ, разработанная в центре тестирования профессионального образования (авторы — В. И. Васильев, Т. Н. Тягунова и др.[3]). Значимым показателем качества программного обеспечения для тестирования является также возможность хранения больших баз данных и оперативного формирования из них тестов. Важными характеристиками компьютерного инструментария для пользователя являются дружественность интерфейса, набор возможностей для представления различных форм тестовых заданий (открытой, закрытой, на соответствие, на последовательность, конструктор, фасет и др.).

Комплекс этих задач на высоком уровне решается далеко не во всех разработках, предоставляемых на рынке образовательных технологий. Так, широко используемая преподавателями HyperTest 1.1 (http://hypertest.virtualave.net) — гибкая, простая в использовании программа — предполагает наличие базы более 1000 вопросов, случайное изменение порядка предъявления тестовых заданий и навигацию по тесту с пропуском заданий, гибкую систему оценки, просмотр и печать протокола тестирования. Однако она требует работы программиста при введении баз данных и только одну форму задания — закрытую, с выбором ответов. В Интернете можно найти различные тестовые системы, например, универсальный тестовый комплекс http://utk.mastak.ru, экзаменатор http://www.pisoft.ru/exam/, контрольно-тестовую систему http://www.omm2002.chat.ru, UniTest System http://www.sight2k.com/rus/unitest и др. При выборе тестовой системы следует обращать внимание также на возможность ее использования в сети, в том числе в Интернете, на возможность набора тестов и работы на различных языках (русский, английский и др.), на возможность использования тестов-тренажеров, на условия защиты и сжатия баз данных, на возможности обработки результатов тестирования, изменения шкалы оценивания, возможность применения и оценки результатов выполнения вопросов разного уровня сложности.

Наконец, третьим важнейшим компонентом, обусловливающим качество системы тестирования, является выбор дидактической концепции и технологии составления заданий для оценивания результатов образовательного процесса.

Согласно современным требованиям, оценке подлежит в первую очередь уровень достижения студентом образовательных стандартов — минимально

необходимых результатов, которые должны быть достигнуты по каждому учебному курсу.

Очевидно, что четырехбалльная шкала не дает представления о полноте и многообразии образовательных результатов студентов. Многие преподаватели расширяют данную систему, вводя к баллам дополнительные знаки «+» и «–», применяют десяти-, двадцати-, стобалльные шкалы, рейтинговую систему. Личностно-ориентированное обучение предполагает расширенные системы качественно-количественного оценивания: диагностические карты студента, дневники личных достижений, портфолио. Однако использование таких систем оценивания является трудоемким для преподавателя и в определенной степени субъективирует результаты обучения.

Оценка успехов в обучении требует анализа вопроса о том, что подлежит измерению, а также вопроса о критериях, показателях, шкалах, единицах и инструментах измерения. Традиционно контролю подлежат знания, умения и навыки обучающихся, которые проверяются с помощью контрольных работ, зачетов, экзаменов. Эти образовательные характеристики можно проверить и в ходе тестирования. Диагностика личностных качеств имеет свою специфику, поскольку результаты обучения имеют две стороны — внешнюю (материализованную образовательную продукцию) и внутреннюю (личностную). Для диагностики способностей кроме различных специальных тестов применяют ситуативный подход, то есть диагностирование с помощью искусственных или естественных образовательных ситуаций. Педагог, опираясь на субъективный опыт, знание особенностей конкретных обучающихся и понимание всегда уникальной ситуации, фиксирует происходящие изменения. Современными методами контроля образовательной деятельности выступают и способы анализа и оценки образовательной продукции обучающихся. Каждый элемент образовательного продукта может быть оценен количественно (баллы) или качественно (рецензии, отзывы). Только с помощью комплексных методик, включающих наблюдение, тестирование, анализ образовательной продукции, преподаватели могут оценивать уровень развития личностных качеств студентов по параметрам, сгруппированным в определенные блоки, например: креативные качества, когнитивные, оргдеятельностные. В число таких личностных качеств могут входить целеполагание, рефлексия, генерация идей, символотворчество, прогнозирование. Сегодня в педагогике принято результаты образовательного процесса выражать через компетентность как способность решать задачи определенного класса. Заметим, что все компетентности требуют различных типов действия. Подходы к диагностике компетентностей в настоящее время только разрабатываются.

Педагогическая диагностика строится на ряде принципов, выступающих в качестве ее исходных оснований и одновременно условий ее органичного включения в целостный педагогический процесс. К числу таких принципов можно отнести:

- системность и конкретность;
- научную обоснованность теории и методики диагностики;
- сочетание констатирующей и корректирующей функций диагностики;
- последовательность и преемственность системы диагностики;
- доступность диагностических процедур педагогам и ученикам;
- оптимизацию форм и методов диагностики;
- комплексность диагностики;
- прогностичность диагностики.

Тестирование как метод современной диагностики опирается на научно обоснованные статистические закономерности при анализе больших объемов проверок образовательных результатов. Оно обеспечивает технологичность диагностических процедур, позволяет хранить и сравнивать результаты диагностики.

Отечественные тестологи справедливо отмечают, что «по результатам тестового контроля знаний учащихся можно сопоставлять качество разных учебников, методических систем обучения. Систематический объективный контроль дает сведения об эволюции качества образования молодежи и при внесении различных изменений в образовательную систему. Появляется возможность сравнивать качество образования в разных странах на основе единых критериев»[4].

Рассмотрим основные понятия, которые используются в дидактической тестологии.

Тест — это инструмент, состоящий из статистически выверенной системы заданий, стандартизированной процедуры проведения и заранее спроектированной технологии обработки и анализа результатов, предназначенных для измерения качеств и свойств личности, изменение которых возможно в результате систематического обучения.

Дидактический тест — система взаимосвязанных заданий для контроля усвоения знаний, сформированности умений, навыков учащихся или других компонентов образованности.

Тестовое задание — одно из составляющих структуры дидактического теста, которое может включать в себя краткую инструкцию для обследуемого, тестовую задачу, эталон ответа (или описание четкого алгоритма выполняемых обследуемым действий). В тестах закрытого типа в структуру тестового задания входят варианты ответа (как правильные, эталонные, так и отвлекающие, «дистракторы»).

Преподаватель может, используя правила тестологии, составить тест для текущей проверки результатов учебы. Однако тестирование выполняет свою основную функцию только при использовании надежных и валидных стандартизированных тестов. Эффективный диагностический тест — это стандартизированная методика, тщательно научно-методически разработанная и проверенная. Тестирование сегодня рассматривается как современная парадигма объективного оценивания учебных достижений.

Стандартизация измерительных инструментов для решения задач объективного оценивания учебных достижений проходит при проведении теста на большой репрезентативной выборке лиц определенного типа, для работы с которыми он предназначен. Эта группа лиц служит для установления нормы, то есть обычного среднего уровня выполнения теста.

Надежность теста означает «согласованность» оценок у обследуемых лиц при их повторном тестировании тем же самым тестом или его эквивалентной формой. Исследование **валидности** должно давать ответ на вопрос, действительно ли данный тест измеряет то, для измерения чего он предназначен с точки зрения разработчиков. Для этого требуются независимые внешние критерии того, что тест должен измерять.

Выделяют тесты входного, текущего и итогового контроля, а также нормативные и критериально-ориентированные тесты. **Нормативные** — позволяют по результатам подсчета набранных баллов (выполненных заданий) сравнивать различные группы обследуемых друг с другом и проводить аналогию внутри каждой группы. **Критериально-ориентированные** тесты определяют, в какой степени каждый испытуемый владеет знаниями и навыками, необходимыми для выполнения каких-либо конкретных учебных или профессиональных задач.

Дидактическое тестирование можно рассматривать как научно-оптимизационную процедуру, позволяющую объективно установить с заданной верностью степень обученности индивида. Прежде всего, характер тестирования определяется его назначением, целью. По этому признаку тесты подразделяют на тесты учебных достижений (учебные проверки) и тесты конкурсных испытаний. Задача

первой группы тестов — определение уровня обученности, компетентности; задача второй группы тестов — отбор и отсев испытуемых. Причем, проверку можно осуществлять на основе нормативной ориентации или личностной ориентации. Во втором случае особо выделяется ориентационная функция проверки, свойственная мониторингу, направленная на непрерывное оценивание уровня обученности тестируемых, на совершенствование по их результатам содержания отдельных тем или разделов учебных дисциплин, а также на коррекцию знаний обучающихся. Нормативно-ориентированные тестовые испытания нацелены на разделение объектов на классы по уровням обученности.

Теория тестирования достаточно хорошо разработана. Методологические правила и рекомендации отражены в международных стандартах (IMS). Обзор современных зарубежных подходов к тестированию имеется в работе М. Б. Челышковой и Г. С. Ковалевой[5].

Согласно рекомендациям тестологов, при составлении тестовых материалов необходимо соблюдать определенную этапность этой работы.

1. Определение цели дидактического тестирования

Цель должна быть поставлена настолько точно и определенно (диагностично), чтобы можно было однозначно сделать заключение о степени ее реализации. Тесты достижений в педагогической диагностике могут осуществлять диагностику знаний (их глубину, осознанность, структурированность), умений, навыков, а также обучаемости, развития навыка как психического новообразования, возникшего под воздействием учебного процесса и самообразования учащегося. Эти тесты могут измерять влияние обучения на развитие логического мышления, критическую оценку выводов, приемы решения задач, на развитие воображения[6]. Широко ориентированные тесты могут диагностировать понимание и применение научных принципов, понимание литературы, восприятие искусства. Специализированные тесты элементарных знаний и умений могут диагностировать: знания законов, их проявления, совокупность фактологических знаний, умения их применять при решении задач на заданные темы или при решении комплексных проблем, навыки чтения и понимание формул, графиков, письма, использования символики, счета, обращения с клавиатурой дисплея и пр.

Прежде всего, дидактическое тестирование направлено на оценку уровня достижения когнитивных целей. Преподавателю традиционных учебных предметов полезно ориентироваться на следующие рекомендации. С помощью теста принципиально можно осуществить проверку *следующих знаний:*

– знания фактов, их причин, различий;

– знания научных и иных проблем по изучаемой теме;

– знания фундаментальных понятий по теме, их определений, представлений об объеме и содержании понятий, знания практических применений понятий;

– знания основных правил, закономерностей и законов, их формулировок, условий и границ проявления, специфики применения;

– знания теорий, опытных фактов, послуживших основой их разработки; основных положений, уравнений, доказательств, выводов, практических приложений, прогностических возможностей.

С помощью теста можно реализовать *проверку умений:*

– владения фактами — установление их причин, взаимосвязей;

– владения проблематикой — формулирование проблем по теме, умение отыскивать возможные пути решения проблемы;

– владения понятиями — узнавание, определение понятий, раскрытие их объема, характеристика количественного состава объектов, их классификация,

взаимосвязи, практическое применение понятий;

– владения правилами, закономерностями и законами — узнавание, формулирование, раскрытие содержания, раскрытие действий, связанных с применением;

– владения теориями — узнавание, отыскание опытных фактов, необходимых для разработки теории, раскрытие содержания теории (характеристика основных положений, уравнений, доказательств, выводов, осуществление на основе теории практических действий).

С помощью теста можно осуществить *проверку навыков:*

– построения и осуществления алгоритма операций выполнения конкретных действий в структуре умения;

– моделирования практического выполнения действий, составляющих данное умение;

– выполнения комплекса действий, составляющих данное умение;

– самоанализа результатов выполнения действий в сопоставлении с целью деятельности;

– по времени выполнения умения.

Тестирование поможет в *проверке усвоенных способов деятельности:*

– узнавания, раскрытия содержания (характеристика действий и операций, составляющих сущность) методов и процедур;

– владения методами и процедурами, связанными с получением знаний и их обработкой;

– применения методов и процедур в различных вариантах последовательности составляющих их действий и в новых условиях;

– характеристики условий и границ методов и процедур.

С помощью теста можно осуществить *проверку развития способностей:*

– выполнение тестов достижений, тестов интеллекта, тестов креативности и пр.;

– выполнение видов деятельности, соответствующих целевым предметным установкам изучаемой темы.

Сложнее оценить проявление творчества при решении задач, понимание материала и точность его усвоения. Частично творческие параметры или параметры глубины усвоения можно проверить и с помощью тестов. Например, применение знаний в нестандартных ситуациях. Возможны также групповые тесты, при работе с которыми учитывается эффект группового взаимодействия испытуемых.

Для углубленной диагностики предлагаются задания, позволяющие выявлять такие особенности решения учащимися учебно-познавательных задач, как[7]:

– реконструкция структуры объекта по заданным преобразованиям (когда изменение условий задачи приводит к изменению свойств взаимодействующих элементов);

– реконструкция структуры объекта при изменении числа ее элементов (включении или исключении их из системы);

– реконструкция преобразования по изменению структуры объекта (свойств исследуемой системы);

– умение замещать схемы операций связями элементов в системе (то есть выявление обобщенности и предметности действия. В этом случае используются в заданиях схемы, фиксирующие состав преобразований некоего объекта, а испытуемые должны воспроизвести в соответствии со схемой определенную группу операций и описать полученные в результате преобразований свойства объектов);

– конструирование схем операций (проверка системности действия), что важно при возникновении необходимости трансформации схемы операций из-за изменения условий анализа объекта.

Наиболее проблематичным является использование тестирования для диагностики способности обучающихся осуществлять оценочную деятельность (оценку адекватности суждений личностного и творческого характера); для диагностики мировоззренческих взглядов и убеждений; для диагностики результатов твор-

ческой деятельности; для диагностики особенностей поиска путей решения комплексных задач и др.

Тесты являются только одним из видов учебных заданий, используемых для контроля и диагностики знаний учащихся. Тест — это система специально составленных заданий, решение которых имеет однозначные правильные ответы либо четкое описание (как минимум — на уровне ранговой шкалы) критериев интерпретации и оценки. Далеко не всякая задача может быть представлена в форме теста. Кроме того, проблематична и фиксация сложных знаний и умений в форме теста, в основном тесты проверяют именно полученные в ходе обучения знания и умения, возможно «натаскивание» на выполнение тестов. Эти недостатки частично можно нивелировать за счет разделения сложных действий на части и их проверки блочными тестами. Большие затраты времени на разработку тестов частично окупаются их долговременным использованием и тем, что они являются надежным инструментарием, в частности, для определения соответствия между условиями обучения и его результатами. Считается, что дидактический потенциал тестирования до сих пор полностью не раскрыт.

2. Составление тестового задания

Тестовое задание — это составная единица теста, отвечающая требованиям технологичности, формы, содержания и статистическим требованиям — известной трудности, достаточной вариации тестовых баллов; положительной корреляции баллов задания с баллами по всему тесту.

При составлении тестового задания должна быть обеспечена:

- содержательная валидность, т. е. соответствие теста содержанию обучения, отображенному в логической структуре и выраженному в форме определения диагностируемым тестом учебных элементов;
- простота (в одном тестовом задании должна быть представлена одна задача данного уровня усвоения);
- определенность (недвусмысленность формулировки),
- однозначность обеспечиваемого конструкцией эталона, в котором должно содержаться полное и правильное решение (или варианты решения задачи).

При составлении банка тестовых заданий важным этапом становится его спецификация. Она может быть представлена в виде таблицы, в которой указываются: в одной колонке — темы курса, в другой — учебные цели или проверяемые дидактические единицы. Далее отмечается, сколько тестов посвящается по каждой теме проверке достижения той или иной цели, указываются соответствующие номера тестовых заданий. Число заданий должно соответствовать важности, объему и целям изучения.

В число требований к тексту тестового задания можно включить установленные эмпирически:

– компактность (наличие одного существенного признака, не более 10 слов, словосочетаний, среднее время на выполнение задания — 1,5–2 мин);

– заданную меру трудности — задание должно быть таким, чтобы только 5% испытуемых с фиксированной степенью обученности смогли выполнить все пункты, мера трудности которых соответствует уровню их обученности.

Чаще всего используются следующие формы тестовых заданий: а) открытого типа, б) закрытого типа, в) на упорядочение, г) на соответствие, д) конструктор.

3. Проведение тестирования, определение качества и корректировка тестовых заданий

Для наиболее эффективной работы во время тестирования рекомендуется стандартизованная процедура его проведения, которая включает ориентацию на следующие типовые правила. Сначала нужно прочитать каждое задание внима-

тельно и до конца, чтобы понять, что требуется сделать. Проработав весь тест, отвечать только на те вопросы, которые не вызывают сомнений, выполнить именно такие задания. Потом подумать над остальными вопросами — не слишком долго, так как время тестирования ограничено, и, по статистике, именно первая реакция оказывается правильной. Попробовать угадать, использовав эрудицию и интуицию, ответы на оставшиеся вопросы, так как этим не ухудшается результат, а приобретаются дополнительные шансы.

В инструкции для тестируемого указываются его права (задавать вопросы по процедуре выполнения теста; в рамках процедуры выбирать темп и порядок выполнения работы) и правила поведения (чаще всего запрещается пользоваться справочными материалами, микрокалькуляторами и другими материалами и инструментами; запрещается разговаривать с кем бы то ни было, консультироваться во время выполнения теста, отвлекать других студентов).

При проведении компьютерного тестирования все инструкции по выполнению тестовых заданий конкретного вида (как работать с клавиатурой, «мышью») даются, как правило, до начала выполнения теста, в так называемом демонстрационном варианте. Затем студенты ориентируются на знакомую форму тестового задания и работают только с текстом задания.

Количественные характеристики каждого тестового задания лучше определять опытным путем. К таким характеристикам относятся: трудность, дифференцирующая сила, точечно-бисериальная корреляция, правдоподобность дистракторов и др.[8] Выбор параметров для оценки теста во многом определяется избранной статистической моделью[9].

Рассмотрим наиболее часто употребляемые способы оценки результатов тестирования в приближении модели нормального распределения. Под параметром **трудности задания** понимают процент испытуемых, верно выполнивших данное задание. Наиболее эффективным считают задание с уровнем трудности 50%. Если все задания такие, то максимальна дифференцирующая способность теста и распределение баллов за выполнение теста близко к нормальному. Следует стремиться к отбору заданий с трудностью 40–60%. Общей рекомендацией для теста на уровне текущего контроля является включение в тест 20% тестовых заданий малой трудности (более 70% правильных ответов), 70% — средней трудности (29–69% правильных ответов) и 10% тестовых заданий большой трудности (менее 25% правильных ответов).

Дифференцирующая способность задания (или дискриминативность) — рассчитывается как разность двух значений: трудности задания для сильной группы тестируемых, составляющих 27% от их общего числа, и трудности задания для слабой группы тестируемых (также 27%). Показатель характеризует способность задания разделять учеников по уровню подготовки (отсюда — валидность и надежность): $D = p_{сильн} - p_{слаб}$.

Желательно, чтобы дифференцирующая способность тестовых заданий была выше 30%, и лишь для легких заданий допускается 20%.

Низкая дифференцирующая способность тестовых заданий может быть обусловлена несколькими причинами, среди которых:

1. Неопределенность понятий в задании (в связи с этим не следует в формулировке заданий употреблять фразы: «главным образом», «преимущественно», «всегда», «никогда», выражения типа «в средней полосе России» и т. п.).

2. Вследствие ошибки составителя в задании отсутствует правильный ответ.

3. Тестовое задание содержит не предусмотренный автором «лишний» правильный ответ или тестовое задание слишком трудно для данного контингента (трудность — менее 25%), например потому, что тема не пройдена. Отсюда —

случайный поиск ответа, угадывание, возможно даже отрицательное значение дискриминативности.

4. Тестовые задания слишком просты для данного контингента (трудность — 80% и более).

Точечно-бисериальная корреляция (ТБК) позволяет ответить на вопрос, как связано выполнение данного тестового задания с выполнением всего теста. Отбираются задания с ТБК больше 0,2; желательно — 0,3–0,5. ТБК для дистракторов должны быть отрицательными и по модулю должны превышать 0,1–0,2.

Затем обязательно анализируются **распределения ответов** по предложенным вариантам ответов. В хорошо составленных заданиях правильный ответ выбирают максимальное число испытуемых, а выборы неправильных ответов распределены равномерно. «Хороший» неверный ответ выбирается большим числом слабых учеников. Выбор неверного ответа большим количеством сильных учеников говорит о том, что дистрактор сформулирован некорректно. Неудачные вопросы и дистракторы определяются по совокупности показателей. Невыполнение задания — также является сигналом к анализу причин. Если его выполнили менее 5% (остальные в принципе не выполняли), — следует внимательно посмотреть, по каким причинам это произошло (слишком трудное, слишком длинный тест и т. д.). Хорошее тестовое задание должно удовлетворять всем вышеизложенным требованиям.

По мнению ряда исследователей, стандартизация педагогических измерений будет происходить по следующим направлениям:

– стандартизация самого метода;

– стандартизация основных (базовых) характеристик метода и средств оценки (измерения);

– стандартизация измерения и интерпретации результатов.

Таким образом, разработка дидактических требований к тестовым заданиям по диагностике достижений студентов является специфической наукоемкой задачей дидактики. Она взаимосвязана с другими компонентами методологической и практической базы компьютерного тестирования: с разработкой статистических моделей, определяющих специфику анализа результатов тестирования и степень их адекватности уровню достижений студента, и с разработкой программных средств для проведения тестирования, обработки и хранения его результатов. Сегодня можно рассматривать компьютерное тестирование как наукоемкую интегративную развивающуюся перспективную образовательную технологию, для освоения которой требуется специальная целенаправленная подготовка специалистов.

ПРИМЕЧАНИЯ

[1] *Васильев В. И., Красильников В. В., Плаксий С. И., Тягунова Т. Н.* Статистический анализ объектов нечисловой природы. М., 2004.

[2] *Дубровский С. А.* Прикладной многомерный статистический анализ. М., 1982.

[3] *Васильев В. И., Тягунова Т. Н.* Основы культуры адаптивного тестирования. М., 2003.

[4] *Михайлычев Е. А.* Дидактическя тестология. М., 2001.

[5] *Челышкова М. Б., Ковалева Г. С.* Основные подходы к оценке качества подготовки обучаемых в России и за рубежом. Обзорный доклад. 8-й симпозиум «Квалиметрия человека и образования. Методология и практика». М., 1999.

[6] *Анастази А.* Психологическое тестирование. Кн. 2. М., 1982.

[7] *Мульдаров В. К., Рубцов В. В.* Метод нормативной диагностики уровней сформированности учебно-познавательных действий у школьников // Вопросы психологии. 1987. № 5. С. 147–153.

[8] *Челышкова М. Б.* Теория и практика конструирования педагогических тестов: Уч. пособ. М., 2001.

[9] *Васильев В. И., Красильников В. В., Плаксий С. И., Тягунова Т. Н.* Указ. соч.

I. Sokolova

COMPUTER TESTING AS A PEDAGOGICAL TECHNOLOGY

The article describes approaches to computer diagnostics of learning achievements. The didactical base of computer diagnostics of learning results is paid attention to. Modern requirements to the text form and content and to the procedure of computer testing are given, features of different testing environments are analyzed. Statistical basis of diagnostics results processing in different testing systems are discussed. It is also shown that computer testing can be viewed as a pedagogical technology.