

## Методология и технология диагностики обученности

*Идиатулин Валентин Сергеевич, заведующий кафедрой физики Ижевской государственной сельскохозяйственной академии, кандидат физико-математических наук, доцент*

РАССМОТРЕНЫ МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ ДИАГНОСТИКИ ОБУЧЕННОСТИ. ОТМЕЧЕНО, ЧТО ДЛЯ ОБУЧЕННОСТИ НЕ ХАРАКТЕРНЫ НОРМАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ И ПРЕДПОЛОЖЕНИЯ О СЛУЧАЙНОМ КОМПОНЕНТЕ ТЕСТОВЫХ БАЛЛОВ, ЕЁ ПАРАМЕТРЫ НАПРАВЛЕННО МЕНЯЮТСЯ В ПРОЦЕССЕ ОБУЧЕНИЯ. ПОКАЗАНО, ЧТО ОЦЕНКА ОБУЧЕННОСТИ ТРЕБУЕТ АНАЛИЗА, КАК ЕЁ СТРУКТУРЫ, ТАК И ДИАГНОСТИРУЕМЫХ ВИДОВ УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ.

*Methodological grounds for diagnostics of learnedness are considered. It is noted that the accidental component of learnedness and the normal distribution are not in point. Thorough training changes them directionally. It is shown that evaluation of learnedness demands both their structure and the kinds of diagnostic educational work analysis.*

Развитие любой науки начиналось с качественных понятий и их субъективных оценок, затем устанавливались сравнительные степени этих качеств, причём далеко не всегда степень качества могла быть выражена числом. До сих пор в ходу ранговая шкала Мооса твёрдости минералов, шкала Бофорта для оценки силы ветра, сейсмическая шкала Рихтера и др. В естествознании измеряемые свойства носят название величин — таковы, например, размеры, массы или заряды объектов; неизмеряемые величинами не являются, это может быть цвет объекта, его материал, форма и т.п. Неметризуемы престиж, репутация, надёжность, ответственность, добросовестность, настроение. Они не имеют выражения в каких-либо естественных единицах, но часто именно этические, психологические и социальные факторы используются для мотивации цели и оценки результата в педагогике. Пока область знания не метризована, она не может иметь статус точной науки. С однозначностью интерпретации связан вопрос об определении типа шкалы. Считается, что в педагогике наиболее значимой является шкала порядка, а полнее всего реализует функции измерения экспертная оценка. В научно организованном виде она использует стандартизованные процедуры отбора экспертов и представления данных 20. Задача верификации экспертной оценки оконча-

тельно не решена, но её вероятный разброс в группе экспертов позволяет рассчитать степень достоверности усреднённых характеристик. Экспертная оценка во многих случаях является единственно возможной, она особенно сложна на творческом и акмеологическом уровнях, а также в инновационных областях.

Проблема диагностики обученности как результата обучения входит в систему оценки его качества федеральной программы развития образования в России. Диагностика стала частью методологии педагогики, обобщающей принципы и способы добывания точных знаний о педагогической действительности, методы их преобразования и интерпретации. Качество образования в целом трудно определяемо, однако частные и внешние проявления могут быть измерены. Для дидактики характерны косвенные измерения, выводимые из ответов на задания, но не сводимые к ним. К настоящему времени в педагогике сложились три уровня квалиметрии: общая, специальная и предметная. Конечная цель каждой заключается в разработке и совершенствовании методов и методик, с помощью которых качество диагностируемого объекта может быть выражено числом, характеризующим степень удовлетворения им общественной или личной потребности. Субъекты обучения не тождественны друг другу по миропониманию, мотивации, ценностям, структурам знаний и способам усвоения, а такая гетерогенность обучаемых делает неразрешимой практически любую дидактическую задачу 1. Эта опасность требует обеспечения, заложенного в технологию обучения, диагностики каждого обучаемого, его готовности к обучению, обучаемости и обученности.

Особенностью педагогической диагностики является то, что она оказывает активное воздействие на объект, как-то меняет его состояние; это наводит на предположения о вероятностном характере объекта диагностики. Стохастические модели обученности получили весьма широкое распространение в образовании. Допустимые статистические операции ограничивают выбор типа шкалы. Полностью они выполнимы только в пропорциональной шкале отношений, имеющей естественный нуль. Метрической является и шкала интервалов, позволяющая вычислять среднее значение (математическое ожидание) признака и отклонения от него. Для ранговой шкалы адекватна лишь медиана — срединное значение упорядоченного признака, для которого вероятности больших и меньших значений одинаковы и равны половине. Для номинальной шкалы достоверна только мода — наиболее частое значение признака. Равноудалённость рангов практически не реализуема и только симулирует точность шкалы, причём над числовыми значениями рангов нельзя производить обычные арифметические операции. Принципиальное значение имеет то, что точность измерений в ранговой шкале даже падает с ростом их числа, тогда как в интервальной и пропорциональной растёт. Точность наиболее высока в отсутствие случайных ошибок, отсутствие же систематических ошибок определяет правильность измерения. Постоянство статистических результатов обеспечивает их воспроизводимость и устойчивость при повторных измерениях. Надёжность характеризует степень точности измерения признака, последняя не может быть выше его стабильности, что и выявляет корреляция повторных измерений. Измерение надёжно, если его повторение даёт тот же результат в пределах погрешности измерения. Коэффициент надёжности мало информативен, во многих отношениях предпочтительнее стандартная ошибка с известной доверительной вероятностью или другая оценка погрешности.

С методологической точки зрения фундаментальной характеристикой диагностики является валидность. Она отражает степень соответствия диагностируемого показателя объекту диагностики. Понятие валидности не абсолютно, она зависит от вида критерия. В оценке обученности различают валидность содержания, определяемую по степени охвата учебного материала, функциональную валидность по степени владения этим материалом; выделяют валидность соответствия результатов, полученных разными методами, прогностическую валидность согласования с последующими измерени-

ями, конструктивную валидность соответствия модели объекту и др.

Обученность представляет собой результат обучения, выраженный в действиях обучаемого: это совокупность знаний, умений и навыков по данной учебной дисциплине, и её структурные составляющие — градации объёмов знаний и уровней их усвоения. Диагностика обученности состоит в выявлении определяемого качества и измерении его интенсивности. Методология оценивания обученности разработана авторами работы [16], она предполагает, что любое качество и свойство обучаемого может быть выражено в единицах той или иной меры. Существует объективная уверенность, что интенсивность проявления качества является упорядоченной в рамках определённой шкалы. Проблема в том, что шкалы разных индивидуумов оказываются неидентичными и не могут быть наложены одна на другую. Возникает необходимость в некоторой эталонной шкале, не зависящей от субъектов процедуры диагностики.

Только с введением единицы измерения можно сводить отношения между свойствами объектов к отношениям между числами, которые устанавливаются в результате абстрагирования от их конкретной природы. Определение измерения как процедуры сравнения величины с некоторой однородной её единицей применимо лишь к экстенсивным величинам, которые отвечают принципу аддитивности, во всех других случаях используется существующая связь между измеряемой интенсивной и какой-либо экстенсивной величиной. Для обученности объективна пропорциональная шкала, однако, достижение пропорциональности бывает затруднено качественной разнородностью её структуры.

С энциклопедической точки зрения мерой знания является количество его элементов. Элементами научного знания могут быть идеи и формы их представления или их достаточно автономные фрагменты. Наименьшая значимая единица знания, выделяемая в отдельное высказывание, элемент семантической памяти, связывающий идеи и понятия, в когнитивной психологии названа пропозицией [17]. Современная психолингвистика подтверждает пропозициональный характер доречевых когнитивных референтов, близость к ним по структуре фраз и предложений облегчает оперирование последними, причём семантические связи преобладают над формально грамматическими [12]. Любая осмысленная информация, в конце концов, перекодируется в вербальную форму, которая является основной в обучении. Понимание

также основано на пропозициях, упорядоченных в когнитивных структурах — понятно уже выраженное знакомым образом.

В дидактике часто используется понятие смысловых учебных элементов, как перенесённых из науки в учебные программы объектов изучения. Их число и количество логических связей между ними заведомо ограничено объёмом учебного материала дисциплины и порогом насыщения способности к ассоциациям человеческого ума. Учебные элементы характеризуются степенью научной абстракции, уровнями и рядом качеств усвоения. Задачей контент-анализа ставится выделение некоторого набора достаточно постоянных и однородных элементов знания. Их перечень можно было бы составить, приписывая каждому элементу индекс значимости, определяемый, например, частотой употребления в учебной и научной литературе. Такое упорядочение элементов знаний делает их, хотя бы в принципе, доступными для наблюдения.

Знания как совокупность усвоенных элементов опыта человека и общества, проверенные общественной практикой, воспроизводятся обучаемым в речи и деятельности. Восприятие информации органами чувств интерпретирует и фильтрует её с участием высших когнитивных механизмов — памяти, мышления, воображения, знаний о мире, подключая нервные связи мозга. Скорость забывания определяется многими факторами: объём материала, трудность усвоения, интерференция, ретроактивное торможение, мнемоническая задача, интеллектуально насыщенные повторения, реминисценция, вербальное осмысление... Элементы знания существуют в их конкретной связи с другими. Их количество экстенсивно характеризует эрудицию, а число и сила связей — интенсивность и продуктивность ума. Когнитивная сфера многомерна во многих отношениях: по классам знаний, по языкам их выражения, по факторам общих способностей, по способам мышления, по семантическому дифференциалу, по формам представления информации, по её характеристикам, по структуре ценностей.

Недостаточно обоснованна продолжающаяся тенденция одномерного шкалирования знаний, несмотря на выявляемую многомерность их структуры 5. Различные свойства, характеристики, параметры знания несводимы друг к другу. Многомерность проявляется в анализе общей и индивидуальной культуры как совокупность параметров значимости, абстрактности, специфичности и связности элементов знания [14]. Уровни усвоения учебных элементов также

ориентированы на весьма различные виды деятельности, каждый из которых объективно оценивается независимо от других. Введение весовых характеристик всегда произвольно и неоднозначно, оно часто вызвано необходимостью ранжирования обучаемых, что характерно для всех рейтинговых систем. В структуре обученности традиционно выделялась триада: знания умения, навыки, эпистемология также всегда различала декларативное и процедурное знание, в том числе и реализуемое без контроля сознания. Обученность как результат обучения проявляется в деятельности обучаемого, которая может быть репродуктивной и продуктивной, каждый вид разделяется на уровни, разные авторы детализируют их по-разному, но общие черты совпадают: репродуктивными уровнями являются фактуальный (декларативный) и операционный (процедурный), продуктивными — эвристический и творческий. Эта таксономия когнитивных целей достаточно полна для диагностики структуры обученности с оговоркой, что творческий уровень в меньшей степени является результатом обучения, он имеет не деятельностную, а личностную природу, практически плохо измерим из-за отсутствия эталона объективно нового, его элементы могут включаться в эвристический уровень. Таким образом, в структуре обученности достаточно выделять следующие уровни усвоения учебных элементов: фактуальный, содержащий то, что можно усвоить и выучить; операционный, включающий то, чему можно научиться; эвристический, требующий трансформации усвоенного, интуиции и логики, развития интеллектуальных способностей личности. Подобная трёхмерная метрологическая характеристика цвета, например, является основным принципом колориметрии, которая базируется на разложении цвета на три независимо измеряемые составляющие.

Ни валидность, ни надёжность диагностики не достижимы, если не обеспечивается её объективность. В основе измерителей, широко применяемых в педагогике, лежат разного рода контрольные задания, поскольку наблюдаемы не латентные процессы, а внешние формы их проявления. Измерению доступны не всякие, а только конвергентные результаты, которые могут быть представлены в однозначно лучшем или строго правильном виде. Дивергентные достижения могут приводить в принципе к разным, но одинаково ценным результатам, диверсификация которых не имеет объективной меры. Креативная сфера неизмерима даже методами экспертной квалиметрии, когнитивная простирается от знания фактов до аргументированных оце-

нок. Обученность может быть соотнесена с результатами других, которые определяют социальную норму, с прежними результатами обучаемого, что характеризует индивидуальную норму, либо с поставленными целями, которые представляют критериальную норму. Последняя наиболее отвечает целям и задачам педагогики и признаётся потребителями образовательных услуг. Наиболее технологичным средством диагностики служат тесты уровней обученности, позволяющие конструировать системную технологию организации учебного процесса.

Принципиальная неметрируемость системы образования на творческом уровне владения изученным материалом допускает введение шкал обученности на уровнях знаний фактов и понятий, терминов и определений, принципов и законов, формул и формулировок, умений выполнять стандартные операции по освоенному образцу, правилу, алгоритму, их неоднозначные процедуры и эвристический анализ. Тесты уровней обученности, включающие задания на выполнение деятельности требуемого уровня и систему оценки с выработанным эталоном решения, должны соответствовать смыслу и содержанию выявляемого свойства и уровню его усвоения [13]. Содержательная и функциональная валидность тестов определяются на основе групповых экспертных оценок и подвергаются эмпирической проверке на репрезентативной выборке обучаемых.

Критериально ориентированные тесты уровней обученности, область содержания которых охватывает знания, умения и навыки для данной учебной дисциплины, в состоянии выполнять все основные функции педагогического контроля. Обучающая функция обеспечивается деятельностью обучаемых, направленной на усвоение и приобретение знаний, совершенствование навыков и умений при выполнении тестовых заданий, а также последующем анализе ошибочных и правильных решений. Развивающая функция воплощается при выполнении заданий, ориентированных на зону ближайшего развития обучаемых, а также заданий эвристического уровня, в т.ч. имеющих учебный проблемный характер. Активизация обучения проявляется в соревновательном характере процедуры тестирования, в её действенной мотивации, возникающей в условиях педагогического контроля. Она может быть усилена элементами дидактической игры. Закрепляющая и обобщающая функции реализуются при предъявлении заданий в логической последовательности, в оригинальной форме представления учебного материала. Воспитание систематической ответст-

венности обеспечивается регулярностью тестирования, объективностью его результатов, их гласностью. Системность и индивидуальность контроля не оставляют безразличным ни одного обучаемого, становятся психологическим фактором его совершенствования.

При тематическом тестировании выявляются не только достижения, но и конкретные пробелы в знаниях и умениях обучаемых. По данным тестирования может быть проведена индивидуальная диагностика степени обученности каждого обучаемого и построена дифференциальная карта уровней его обученности. Суммирование индивидуальных тестовых баллов даёт возможность изучения структуры обученности как академической группы, так и всего потока обучаемых, что позволяет оперативно производить коррекцию учебного процесса, организовывать эффективное управление им в реальном времени [4].

Богатый опыт тестирования, накопленный в психологии и социологии и перенесённый в педагогику в качестве средства диагностики обученности, равно как и экспертные методы квалиметрии, не всегда в нужной мере учитывают особенности дидактического объекта. Тестовый балл данного обучаемого не является случайной величиной, на чём основываются многие теории нормативно ориентированного тестирования, он не может быть охарактеризован ни устойчивой частотой, ни аксиоматической вероятностью. Это величина неопределённая, такого рода величины всегда сосуществуют с детерминированными и случайными. Неопределённой величиной является результат голосования эксперта, поскольку его вероятность априорно неизвестна, иначе эксперта можно было бы заменить, например, бросанием игральной кости — именно к такому выводу пришли в своё время Лаплас и Пуассон [19]. Теория вероятностей применима только к статистически устойчивым событиям, частоты любых подпоследовательностей которых совпадают, либо к событиям, которые могут быть охарактеризованы аксиоматической вероятностью. Ни тестовый балл, ни оценка эксперта такими свойствами не обладают, их вероятности априорно не известны, они статистически неустойчивы, не наблюдаемы многократно в однородных условиях, поэтому являются величинами не случайными, а вообще говоря, неопределёнными. Задача математической статистики, заключающаяся в нахождении вероятностей каких-либо событий по эмпирическим данным, тоже разрешима в условиях статистической устойчивости. Аппарат теории измерений основан на непосредственных независимых, как правило, равнозначных измерениях одной величины.

Критериально ориентированные тесты не предполагают высокой различающей способности, они не имеют целью ранжирование обучаемых по какому-либо нормативному признаку, а являются тестами учебных достижений каждого испытуемого [8, 10]. Статистически определяемые характеристики — средний тестовый балл, дисперсия и среднееквадратичное отклонение — не дают оценок надёжности и точности отдельного измерения, критерий выполнения теста может быть весьма жёстким. Критериальный тестовый балл должен быть достаточно далёк от математического ожидания числа правильных ответов при их случайном выборе и находиться за пределами стандартного отклонения. Чтобы полностью избежать элементов угадывания ответов, в необходимых случаях можно требовать краткое обоснование выбора наиболее правильного из них, т.е. формулировку, расчёт, объяснение и т.п. При этом тестовое задание превращается в истинное или ложное суждение с вариантом доказательства этого. Такая форма тестового контроля сохраняет все преимущества теста закрытого типа, включая известные оперативность и объективность проверки, а также предоставляет дидактически мотивированную возможность дополнительного коммуникативного взаимодействия преподавателя с обучаемым.

В дидактических целях допустимы и другие отклонения от строгой предметной чистоты и единства формы тестовых заданий. Они могут возникать при исключении неработающих дистракторов, достижении необходимой полноты альтернатив, различий в степени градуирования объектов и явлений, последовательном проведении логических принципов и т.д. Концентрации внимания способствует отражение в инструкции к тесту возможности более чем одного правильного ответа на какое-либо задание теста. Ограничение числа слов допустимо только в той мере, которая не искажает смысл задания, его понятность для обучаемого. Стоит иметь в виду, что читается, как правило, задание целиком, при этом охватывается весь набор дистракторов, а не подстановка их поочерёдно в высказывание. По этой причине вовсе не обязательно стремиться к такому грамматическому совершенству, которое по форме правильно, но логике восприятия противоречит; грамматика нужна для понимания, но от неё стоит отходить тогда, когда это способствует ещё большему и лучшему пониманию.

Конструирование тестовых заданий должно наиболее полно соответствовать учебной цели, определение чего возможно экспертным путём, хотя оценка сложности заданий экспертами до-

стоверна только в небольшом числе случаев. Для критериально-ориентированных тестов формулы классической теории тестов не работают, распределение ответов может не быть нормальным, за исключением тех случаев, когда предпринимаются специальные меры по их нормализации. Определение его вида и параметров становится самостоятельной задачей эмпирического исследования. Исследования показывают, что эмпирические среднее, дисперсия, асимметрия и эксцесс распределения степени обученности направленно меняются в процессе обучения [2-4].

Основная дидактическая цель использования критериально-ориентированных тестов заключается в оптимизации учебного процесса по выбранным параметрам, при этом достаточно бывает критериев, установленных преподавателем. Преимущества дидактических целей в том, что они могут быть сформулированы практически весьма точно, для контроля их достижимости как раз и нужна частая диагностика, которая даёт информацию не только о конечных результатах, но и о приводящих к ним конкретных дидактических процессах, в том числе и ситуативно обусловленных, что позволит сразу выделять педагогические находки для обеспечения условий эффективного обучения, использовать полученную информацию в педагогическом поиске.

К преимуществам тестовых измерителей справедливо относят высокую объективность контроля, его независимость от взаимоотношений субъектов учебного процесса [8,15]. Точность измерений обеспечивает обоснованность и дифференцированность оценки, а скорость обработки данных позволяет работать с большим контингентом обучаемых. Тестовой диагностике более доступна проверка степени усвоения просто организованного материала, чем глубинного понимания, поэтому многие считают, что тесты оценивают не развитие способностей, а лишь умение оперировать фактами, тем не менее, руководители образования не раз говорили, что тестирование даёт возможность выявить талантливых и способных, делая это особым справедливым и объективным способом.

По форме тест представляет собой систему большого числа достаточно кратких заданий, объединённых общей логикой построения. Область содержания теста охватывает по возможности полно тот объём знаний, умений и навыков, которыми должен овладевать обучаемый. При преобразовании в тестовые задания учебные цели необходимо конкретизируются и опе-

рационализируются. Учебная дисциплина состоит из разделов, в которых выделены темы, состоящие из достаточно ограниченного числа учебных элементов, которые сохраняют самостоятельный смысл, а их дальнейшее расчленение ведёт к утрате включённой в них учебной и научной информации [8,9]. Смысловые учебные элементы чаще всего уже представлены в виде некоторой системы пропозиций, отражающих положительное содержание учебной дисциплины. Логическая структура теста как раз и использует эту форму их представления, которая в зависимости от выбора испытуемого превращается в истинное или ложное высказывание, кодируемое дихотомным тестовым баллом. Интеррогативная форма задания исключает эту возможность, так как вопросы не могут быть истинными или ложными, а ответы на них, как правило, многовариантны, неопределённые, многословны и нетехнологичны [8]. В вопросе важную роль играют интонация, логические ударения, акценты которых непросто выразить графически достаточно однозначно.

Тесты учебных достижений должны измерять качества усвоения учебного материала данной учебной дисциплины: объёмы знаний и уровни владения ими, т.е. то, что охватывается понятием обученность. Они валидны, если дадут низкий балл недоучившемуся и высокий обученному, они надёжны, если дадут один и тот же балл одинаково знающим. Способствует этому унификация тестов по форме, избегание всяких ухищрений, которые больше выявляют интеллект и способности, нежели знания и умения. Тесты обученности не предназначены для измерения свойств интеллекта, трудоспособности, профпригодности, прилежания или личностных характеристик. Диагностика последних в массовом порядке во многих странах запрещено, это недопустимо и опасно без учёта психологического состояния испытуемых. В своём докладе [21] С.Е. Шишов предостерегал также от опасности ранжирования обучаемых даже по объёмам полученных знаний. Думается, что эту проблему можно разрешать в духе медицинской этики — каждый вправе знать об уровне своих знаний, но это не обязательно должно становиться достоянием других. Этика и право не препятствуют диагностике обученности и когнитивных возможностей обучаемых.

Стремление к измерению вполне естественно, требуется, однако, известная осторожность в верификации и интерпретации его результатов. Ставка на положительные эмоции намного эффективнее, чем фиксация недостатков. Объективная мера обученности не всегда восприни-

мается психологически комфортно, поскольку не оставляет места на скидки из-за возможности необъективного оценивания преподавателем. Вопрос о том, измеряют ли тесты, интеллектуальные способности давно стоял перед психологами, причём довольно надёжно установлено, что интеллектуальные тесты не диагностируют каких-либо процессов, а только сравнивают их результаты. Задания, репрезентативные для одних социальных или этнических групп, не всегда пригодны для других. Даже сама степень знакомства с процедурой тестирования может порождать различия в тестовых баллах, не связанные ни со способностями, ни с уровнем обученности. Решение задач, выполнение тестовых заданий всегда подключает весь арсенал когнитивных процессов, таких как восприятие, память, распознавание, мышление, причём практически невозможно таким путём выделить отдельную способность и приходится анализировать их в совокупности. Определений интеллекта так много, что это свойство индивида едва ли является измеримым. Даже такие психологические способности, как память, невозможно рассматривать как отдельные свойства, которые можно измерять, поскольку в познавательной деятельности имеем дело не со свойствами, а с процессами, зависящими от конкретной ситуации, мотивации, установки и других факторов.

Трактовка обученности как результата обучения, выраженного в действиях обучаемого, снимает многие психологические проблемы, хотя и не все и не всегда. Поэтому столь стандартизованные требования предъявляются не только к форме задания, но и к самой процедуре тестирования. Спецификация теста устанавливает соответствие области содержания теста и тестовых заданий предъявляемым требованиям и предметной области учебной дисциплины. Они содержат общее описание того, что измеряет тест, включает в себя инструкцию по проведению тестирования, характерные формы и содержание заданий, правила представления ответов на задания, критерии их оценки и способ оценивания. Таким образом, надо решить какие знания, умения, представления, навыки должны быть и могут быть измерены. В структуре обученности по каждой дисциплине возможно выделение минимального объёма мировоззренческих знаний, которые должны оставаться в памяти любого обучаемого, к ним примыкают базовые знания, которые необходимы для усвоения учебного материала самой изучаемой дисциплины. Следующая категория включает остальные программные знания, которые нужны

в первую очередь для обеспечения готовности к усвоению дисциплин учебного плана и, в определённой мере, к профессиональной деятельности. Эти три градации могут быть дополнены блоком специальных сверхпрограммных знаний, которые рассчитаны на индивидуальные потребности обучаемых [2,16], либо отражают требования специальности, вуза, региона.

Достаточно надёжно измерим фактуальный уровень владения знаниями, т.е. умение узнавать, идентифицировать, воспроизводить основные факты, сведения, термины, определения, формулировки, формулы, принципы, законы, положения изучаемой дисциплины, всё то, что можно узнать, усвоить, выучить и потом воспроизвести по памяти или выделить среди подобного. Закрытая форма тестовых заданий на этом уровне наиболее уместна, т.к. оценка правильности выбора может быть сделана вполне однозначно, а открытый ответ бывает трудно верифицируем. Воспроизведение и узнавание как мнемонические процессы не имеют существенных различий, и нет особых оснований полагать, что задания с выбором ответа измеряют нечто иное, чем задания со свободным ответом [10]. Операционный уровень владения изученным материалом измерим столь же уверенно и надёжно. Он предполагает выполнение действий и операций по усвоенному образцу, правилу, рецепту, алгоритму, т.е. умение делать всё то, чему можно было научиться. Подразумевается, что они выполнимы настолько однозначно, что для оценки их правильности достаточно сравнения полученного результата с эталоном ответа, причём валидность задания обеспечивается исключением операций, которые не являются результатом обучения по данной дисциплине. На этом уровне одинаково применимы задания как открытой, так и закрытой формы, однако последняя предпочтительнее в дидактическом и методологическом отношении. Первое обеспечивается разработкой дистракторов, мотивированных типичными ошибками обучаемых, для последующего их анализа и исправления; открытая форма в этом отношении не вносит никакой информации. Методологически существенно, что закрытая форма даёт естественную меру вероятности случайного выбора, с которой возможно сравнение эмпирических частот.

Умение проводить эвристический анализ действий и строить процедуры освоенных операций без известного однозначного алгоритма, но с чётко установленным эталоном решения, также доступно тестовому измерению. Объективно те-

сты могут измерять те же признаки, что и традиционные формы проверки, и не могут то, что не могут они [10]. Эвристический анализ опирается во многом на догадку, развитую интуицию, способность к трансформации усвоенного на ещё неизведанные обучаемым области. Этот уровень усвоения включает всё то, что нельзя просто выучить, чему нельзя и научиться, не обладая определёнными задатками, которые обуславливают развитие способностей, позволяющих успешно овладеть такой деятельностью; от творческого уровня он отличается, пожалуй, лишь тем, что результат деятельности только субъективно неизвестен, хотя для обучаемого он является личным открытием пусть даже небольшого масштаба. Закрытая форма заданий эвристического уровня чётко выделяет эталон правильного ответа, но налагает особые требования на искусство составления дистракторов.

Значимость элементов учебной дисциплины требует включения в тест тех из них, без которых её усвоение будет неполным. Научная достоверность связывается с объективной истинностью содержания и его рациональной аргументированностью и обоснованностью. Полное соответствие современному состоянию науки едва ли достижимо, поскольку на её переднем крае ещё многое недостаточно достоверно, переходящее и недоступно обучаемым. Репрезентативность теста означает полноту отражения в его заданиях необходимых связей учебного материала. Принцип возрастающей трудности заданий лучше применять по стадиям и с учётом логических связей. Вариативность содержания отслеживает изменения в учебной дисциплине и требованиях к уровню её усвоения. Системность отражает связь заданий с общей логической структурой учебного материала. Комплексность теста необходима при рубежном и итоговом контроле, она достигается включением в тест заданий, связывающих элементы разных тем. Связь содержания и формы, предполагает в первую очередь отбор материала диагностируемого в тестовой форме и разных её видах. Результат тестирования должен давать возможность объективной диагностики качества усвоения и оказывать стимулирующее влияние на субъекты обучения, что достижимо при опоре на индивидуальные диагностические данные.

Дидактический тест ещё не завершил стадию полного признания педагогической общественностью, иногда причина этому в тех случаях, когда его использование ни методологически, ни психологически, ни методически несостоятельно. Страстное неприятие вызывает как раз

распространение тестов на области недоступные им, либо их крайне неудачное составление и применение. Тестирование является довольно сложной технологической процедурой, разработка, проверка и применение тестов требуют высокой квалификации и опыта работы. Их возможности заведомо ограничены степенью формализуемости учебного материала. Как и любой другой метод контроля, тест проверяет в большей степени конечный результат, а не способы его получения, которые имплицитно включают в себя способности к трансформации, классификации, систематизации, логическим операциям, ассоциативному мышлению [4,8]. Ограничение времени тестирования и возможность свободно распределять его между отдельными заданиями аналогичны подчинению правилам шахматных соревнований, когда глубокое погружение в размышления над каким-то одним заданием (ходом) проводит к цейтноту и только накопленный опыт позволяет вырабатывать правильную стратегию в распределении ресурса времени. Но такие же проблемы встают при выполнении любой контрольной работы. Тест не оценивает навыки связного письменного изложения — для этого нужны другие средства, зато его логическая форма способствует их формированию. Тест не оценивает коммуникативные способности, однако его результат оказывается всегда дидактически мотивированным поводом для проведения последующего собеседования с обучаемым. Критика как тестов, так и любых других измерителей лишь указывает на правильный путь — совместить и адекватно использовать достоинства разных способов контроля, и со знанием дела минимизировать их недостатки. К подобному выводу приходит и М. Панин [15], однако он видит выход в накоплении условных баллов за каждый вид учебной деятельности, включая посещение лекций, практических и лабораторных занятий, чтение текстов, аннотирование и конспектирование, контрольные работы, т.е. в соединении меры количества с оценкой качества. Гораздо продуктивнее разделить эти два понятия и оценивать только качество выполненных работ, которые предусмотрены учебным планом, причём оценивать привычной педагогической оценкой. Их совокупность становится intersubъективной, когда разные виды работ проводят и оценивают разные преподаватели. Она становится объективизированной, когда сопровождается тестовым мониторингом тех уровней усвоения, которые доступны измерению.

После представления диагностируемой части учебного материала в виде утверждений [6] и

перевода их в тестовую форму должно быть разработано достаточное количество параллельных тестов, содержащих одинаковое число тестовых заданий. При этом содержательная основа (стем) задания на фактуальном уровне, как правило, повторяет сформулированное утверждение, на операционном уже предлагает на его основе сделать расчёт или вывод, а на эвристическом уровне включает в себя нестандартную либо проблемную ситуацию, устанавливающую логические связи утверждений, их причины и следствия. Правильный ответ уже включён в структуру утверждения, для композиции задания в целом бывает необходима его переформулировка для формирования вариативных частей стема (фасет) и составления достаточного числа эффективных дистракторов, т.е. выражений или отдельных слов, включение которых в утверждение, превращает его в ложное высказывание, которое обучаемый ошибочно принимает за истинное из-за неполноты, непрочности, неуверенности в своих знаниях, несформированности умений и навыков, либо неразвитости способностей к трансформации знаний. По результатам эмпирической проверки и групповой экспертизы тестов проводится их сертификация, устанавливающая параметры надёжности и валидности.

В соответствии с изложенными принципами были составлены тематические тесты уровней обученности в двух частях [7], включающие более тысячи заданий закрытого типа разных уровней. Подготовительный этап завершается составлением инструкции по проведению тематического тестирования, содержащей указания по представлению и оформлению его результатов, порядок начисления тестовых баллов. Инструкция устанавливает время тестирования, содержит требования к точности вычислений и другие необходимые сведения. Содержание инструкции доводится до сведения испытуемых перед тестированием.

Тестовые баллы каждого обучаемого суммируются по всем тематическим модулям и служат индивидуальной характеристикой степени обученности. По результатам тестирования преподаватель даёт рекомендации: повторить тему, либо её определённые фрагменты; обратить внимание на знание фактического материала, либо на умения выполнения типовых заданий, или же на эвристический анализ нетиповых задач.

Текущая диагностика не отражает полностью результаты обучения, оставляет незавершёнными его дидактические циклы, умения и навыки

продолжают совершенствоваться и на другом учебном материале [4]. Она даёт верный прогноз для значимой части обучаемых, однако не решает проблемы оценки оставшихся. Карты рубежного тестового экзамена составляются на основе банка тестовых заданий, апробированных в процессе модульно-тематического тестирования. Доля заданий каждого модуля определяется экспертным путём с учётом его веса в общем объёме учебного материала. Для тестового экзамена должно быть подготовлено достаточное число параллельных тестов, охватывающих все изучаемые в семестре темы.

Автор не находит обоснованным оценивание всех обучаемых только по их тестовым баллам. В выборке наиболее успевающих иерархия уровней оказывается инвертирована: более двух третей из них на фактуальном уровне усваивали учебный материал хуже, чем на операционном или эвристическом. Учёт их индивидуальные особенности как раз и позволяют тесты уровней обученности, раскрывающую её многомерную структуру. Решить проблему оценки поможет мониторинг всех видов учебной деятельности, включающий отчёты по лабораторным работам, контрольные на практических занятиях, индивидуальные задания и всё то, что не охвачено диагностикой в тестовой форме. Несовпадение оценок станет поводом для выявления причин расхождений и подлинно педагогического оценивания, которое недоступно пока никакому компьютеру. Совпадение же послужит гарантией объективности оценки и продемонстрирует валидность соответствия тестовой квалиметрии другим методам диагностики обученности.

Априорное установление соответствия тестового балла какой-либо академической оценке практически невозможно без изучения структуры обученности, исследования её динамики, соотношения иерархии её уровней с реальным состоянием контингента обучаемых. Для этого необходима технология обучения, основанная на когнитивной предметной квалиметрии обученности, исследовательский характер которой позволяет ставить и разрешать вопросы такого рода [4,9].

Установление требуемых критериев обученности обеспечивает перевод тестовых баллов в принятую шкалу академических оценок, либо в рейтинговую шкалу. Существуют и объективизированные способы перевода шкал [18]. В граничных областях необходимо руководствоваться теоретической оценкой погрешности тестирования (до 5%) 10 или эмпирической оценкой

погрешности по данным статистического анализа. Дополнительное собеседование необходимо проводить также во всех спорных и вызывающих сомнение случаях резкого расхождения тестовых баллов рубежного экзамена с результатами тематического тестирования и мониторинга учебной деятельности для выявления его причин.

Проблема угадывания всегда служила основой критики заданий закрытого типа. Из наших результатов следует, что оно почти всегда отсутствует: в области вероятного угадывания (для пяти вариантов она лежит около 20% степени усвоения) эмпирические распределения имеют минимум частоты. Если бы угадывание имело место для всех, то здесь был бы максимум, если бы для некоторой части, то локальный максимум, который при большом размахе и СКО распределения мог бы перекрываться последним, а такое отмечалось при входном контроле слушателей курсов довузовской подготовки. Без анализа эмпирических распределений обученности не следует априорно вводить коррекцию на угадывание, особенно там, где оно не имеет места. Упорный труд над заданиями, поиск правильного ответа — всё это требует напряжения ума, работы мысли и следует отделять их от бездумного гадания. Смешивание этих двух процессов как раз и служит почвой для сторонних критиков, которые не давали себе труда вникать в существо тестирования как умственной деятельности, происходящей на высоком для тестируемого уровне.

Психология и педагогика обычно не связывают повышение качества обучения с регулярным контролем; более того, считается, что быстрая оценка не способствует обучению, так как в условиях неопределённости обучаемый продолжает работать над ситуацией. Это, может быть, верно в отношении контроля, сопровождающегося выставлением оценки. Диагностика же обученности отделена от её оценивания, она указывает на конкретные пробелы, над устранением которых надо работать. Неопределённость ситуации и поиск правильного решения могут продолжаться достаточно долго, а указание на то, что уже найденный ответ неверен, лишь подталкивает к продолжению работы, а не завершает её, как это делает оценка. Диагностика не будет отвлекать от учебной деятельности, если диагностирующие задания прямо включены в её содержание. Измерение действующих на учебный процесс факторов превращает его в деятельность, увлекающую педагога-исследователя. Учебный процесс подчиняется объективным законам, и теория даёт решение многих дидакти-

ческих задач, однако многофакторный характер обучения, зависимость его от конкретной ситуации всякий раз заставляют искать пути решения задач, в том числе и не охваченных общей теорией.

## Литература

1. **Беспалько В.П.** Педагогика и прогрессивные технологии обучения. М.: Институт проф. образования РАО, 1995.
2. **Идиатулин В.С.** Естественно-научная подготовка инженера: квалиметрический подход // Образование и наука. № 1. 2001. С. 41–53.
3. **Идиатулин В.С.** Квалиметрия уровней обученности // Стандарты и качество. № 1. 1999. С. 80–82.
4. **Идиатулин В.С.** Когнитивная технология обучения студентов. Ижевск: Шеп, 2002.
5. **Идиатулин В.С.** Многомерная квалиметрия обученности // Материалы V симпозиума «Квалиметрия человека и образования. Методология и практика». М.: Исследовательский центр проблем качества подготовки специалистов, 1996. С. 51–52.
6. **Идиатулин В.С.** Основные понятия общей физики. Опорный конспект. Ижевск: ИжГСХА, 1998; Idem, 2004.
7. **Идиатулин В.С.** Тематические тесты уровней обученности по курсу физики. Часть I. Ижевск: ИжГСХА, 1998; Часть II. Ижевск: Изд-во УдГУ, 2000.
8. **Идиатулин В.С.** Тесты уровней обученности как средство диагностики учебных достижений // Вопросы тестирования в образовании. № 4. 2002. С. 18–30.
9. **Идиатулин В.С.** Технология обучения на основе когнитивной предметной квалиметрии // Проблемы теории и методики обучения. №4. 1999. С. 8–11.
10. **Ингенкамп К.** Педагогическая диагностика: Пер. с нем. М.: Педагогика, 1991.
11. Концепция национальной политики России в области качества продукции и услуг // Стандарты и качество. № 4. 2001. С. 15–20.
12. **Леонтьев А.А.** Основы психолингвистики. М.: Смысл, 1997.
13. **Люсин Д.В.** Основы разработки и применения критериально-ориентированных педагогических тестов. М.: Исследовательский центр проблем качества подготовки специалистов, 1993.
14. **Моль А.** Социодинамика культуры. М.: Прогресс, 1973.
15. **Панин М.** Морфология рейтинга // Высшее образование в России. № 1. 1998. С. 90–94.
16. **Родионов Б.У., Татур А.О.** Стандарты и тесты в образовании. М.: МИФИ (ТУ), 1995.
17. **Солсо Р.** Когнитивная психология. М.: Тривола, 1996.
18. **Хубаев Г.** О построении шкалы оценок в системах тестирования // Высшее образование в России. №1. 1996. С. 122–124.
19. **Хургин Я.И.** Как объять необъятное. М.: Знание, 1985.
20. **Черепанов В.С.** Экспертные оценки в педагогических исследованиях. М.: Педагогика, 1989.
21. **Шишов С.Е.** О федеральном компоненте государственных образовательных стандартов для образовательных учреждений // Материалы I симпозиума. «Квалиметрия человека и образования. Методология и практика». М.: Исследовательский центр проблем качества подготовки специалистов, 1992. С. 18–24.