**Тестирование в учебном процессе: его история и возможности**

Иван Павлович Подласый, профессор, доктор педагогических наук, Челябинский государственный педагогический университет.

Качество обучения напрямую зависит от количества, глубины, своевременности и объективности оценки получаемых знаний. Грамотно составленные тесты позволяют определять уровень усвоения знаний и степень формирования навыков в процессе обучения.

Применение диагностических тестов в зарубежных школах имеет давнюю историю. Признанный авторитет в области педагогического тестирования Э. Торндайк (1874—1949) выделяет три этапа внедрения тестирования в практику американской школы:

1. Период поисков (1900—1915). На этом этапе происходило осознание и первоначальное внедрение тестов памяти, внимания, восприятия и других, предложенных французским психологом А. Бинэ. Разрабатываются и проверяются тесты интеллекта, позволяющие определять коэффициент умственного развития.

2. Последующие 15 лет — годы «шума» в развитии школьного тестирования, приведшие к окончательному осмыслению его роли и места, возможностей и ограничений. Были разработаны и внедрены тесты О. Стоуна по арифметике, Б. Зекингема для проверки правописания, Э. Торндайка по диагностике большинства школьных предметов. Т. Келли разработал способ измерения интересов и наклонностей обучаемых (при изучении алгебры), а Ч. Спирмен предложил общие основы использования корреляционного анализа для стандартизации тестов.

3. С 1931 г. начинается современный этап развития школьного тестирования. Поиски специалистов направляются на повышение объективности тестов, создание непрерывной (сквозной) системы школьной тестовой диагностики, подчиненной единой идее и общим принципах, созданию новых, более совершенных средств предъявления и обработки тестов, накопления и использования диагностической информации.

Напомним в связи с этим, что педагогика, развившаяся в России в начале века, безоговорочно приняла тестовую основу объективного школьного контроля.

После известного постановления ЦК ВКП(б) «О педологических извращениях в системе Наркомпроса» (1936) были ликвидированы не только интеллектуальные, но и безобидные тесты успеваемости. Попытки возродить их в 70-х годах ни к чему не привели. В этой области наша наука и практика значительно отстали от зарубежной.

В школах развитых стран внедрение и совершенствование тестов шло быстрыми темпами. Широкое распространение получили диагностические тесты успеваемости, использующие форму альтернативного выбора правильного ответа из нескольких правдоподобных, написания очень краткого ответа (заполнения пропусков), дописывания букв, цифр, слов, частей формул и т. п. С помощью этих несложных заданий удается накапливать значительный статистический материал, подвергать его математической обработке, получать объективные выводы в пределах тех задач, которые предъявляются к тестовой проверке. Тесты печатаются в виде сборников, прилагаются к учебникам, распространяются на цифровых носителях.

Проблема соотношения устных и письменных форм контроля разрешается в большинстве случаев в пользу последних. Считается, что хотя устный контроль больше способствует выработке быстрой реакции на вопросы, развивает связную речь, он не обеспечивает надлежащей объективности. Письменная проверка, обеспечивая более высокую объективность, кроме того, способствует развитию логического мышления, целенаправленности: обучаемый при письменном контроле более сосредоточен, он глубже вникает в сущность вопроса, обдумывает варианты решения и построения ответа. Письменный контроль приучает к точности, лаконичности, связности изложения мыслей.

В области определения системы оценок (отметок) наблюдается большое разнообразие как в принципах, так и в конкретных подходах, выборе способов оценивания и выставлений оценок. В зарубежных образовательных заведениях практикуются различные системы оценивания знаний, умений, навыков, приняты различные шкалы оценок, включая сто-, двенадцати-, десяти-, двухбалльную и др. Во французском лицее, например, при сдаче выпускных экзаменов результаты определяются по 20-балльной шкале. При этом для каждого предмета устанавливается определенный весовой коэффициент, определяющий значимость данного предмета для избранной обучаемым специальности, профиля или отделения лицея. Благодаря этому оценки по профилирующим предметам приобретают большую весомость.

При значительной децентрализации большинства зарубежных образовательных систем, при условии свободного выбора школой учебных планов и программ, а обучаемыми — учебных предметов редко проводятся централизованные проверки и оценивание знаний, умений. В Великобритании, например, функцию практического контроля за качеством академической подготовки берут на себя университетские комиссии, к которым на период сдачи выпускных экзаменов прикрепляются все грамматические и привилегированные школы, ориентирующиеся в подготовке своих выпускников на дальнейшее университетское образование.

В целом же нужно признать, контроль и учет успеваемости, диагностирование обученности в зарубежных школах носят характер объективной констатации результатов. Он не предполагает заботу о всеобуче. Общепринятый принцип индивидуализации обучения диктует один подход — каждый идет своим путем и темпом, обучается в меру своих возможностей, потребностей, реальных оценок будущего.

До революции в России существовала шестибалльная система оценки знаний с баллами от нуля до пяти. В 1918 г. оценка «0» была упразднена. Но постепенно и оценка «1» стала использоваться все реже, а начиная с 50-х годов все меньше стала использоваться и оценка «2». Пятибалльная система оценок фактически превратилась в трехбалльную, а для большинства обучаемых, которые не могут учиться на «4» и «5», эта шкала стала двухбалльной. Такая оценочная система очень слабо стимулирует учебный труд, «ступенька» между тройкой и четверкой непреодолима для большинства обучаемых.

Однако многие педагоги используют «дополнения» к обычной пятибалльной шкале в виде знаков «плюс», «минус». Реально получается три градации пятерки («пять с плюсом», «пять», «пять с минусом»), три градации четверки (аналогично), три градации тройки и двойки. Получается типичная десятибалльная шкала оценок.

Научно-исследовательские учреждения РАО предложили новые оценочные шкалы, которые проходят экспериментальную проверку в различных регионах страны. Некоторые регионы склонны принять двенадцатибалльную систему оценок, в которой кроме названных десяти баллов существует два экстремальных: балл «1» — «спасайте» — свидетельствует о том, что учащийся требует немедленной индивидуальной помощи или особенного внимания, вплоть до помещения в специальное учебное учреждение; высший балл «12» — экстремальный максимум («ура»), свидетельствующий о появлении способного и чрезвычайно одаренного учащегося, которого следует обучать индивидуально по специальной программе или же в учебном заведении с углубленным изучением предметов.

**Тестирование достижений и развития**

Слово «тест» английского происхождения и на языке оригинала означает «испытание», «проверка». Тест обученности — это совокупность заданий, сориентированных на определение (измерение) уровня (степени) усвоения определенных аспектов (частей) содержания обучения.

Правильно составленные тесты обученности должны удовлетворять ряду требований. Они должны быть:

относительно краткосрочными, т. е. не требовать больших затрат времени;

однозначными, т. е. не допускать произвольного толкования тестового задания;

правильными, т. е. исключать возможность формулирования многозначных ответов;

относительно краткими, требующими сжатых ответов;

информационными, т. е. такими, которые обеспечивают возможность соотнесения количественной оценки за выполнение теста с порядковой или даже интервальной шкалой измерений;

удобными, т. е. пригодными для быстрой математической обработки результатов;

стандартными, т. е. пригодными для широкого практического использования — измерения уровня обученности возможно более широких контингентов обучаемых, овладевающих одинаковым объемом знаний на одном и том же уровне обучения.

Если в основу классификации тестов положить различные аспекты (компоненты) развития и формирования человеческих качеств, то она будет выглядеть следующим образом:

1. Тесты общих умственных способностей, умственного развития.

2. Тесты специальных способностей в различных областях деятельности.

3. Тесты обученности, успеваемости, академических достижений.

4. Тесты для определения отдельных качеств (черт) личности (памяти, мышления, характера и др.).

5. Тесты для определения уровня воспитанности (сформированности общечеловеческих, нравственных, социальных и других качеств).

Применение тех или иных тестов будет наиболее успешным и обеспечит надежные выводы лишь при условии правильного их сочетания со всеми другими группами тестов. Поэтому тестовые испытания всегда имеют комплексный характер. Делать общие выводы, например, об уровне развития обучаемых на основе применения лишь тестов обученности было бы непростительной ошибкой. Когда ставится задача диагностирования обученности в связи с достижениями и развитием личности, нужно применять соответствующие виды тестовых заданий и предписанные им методики измерения, не забывая о локальном характере диагностирования.

При разработке тестов важно, насколько они соответствуют запроектированным целям обучения, образования, развития обучаемых. Важнейшими критериями диагностических тестов обученности являются действенность (валидность, показательность), надежность (вероятность, правильность), дифференцированность (различимость).

Действенность теста по своему содержанию близка к требованию полноты, всесторонности проверки, пропорционального представления всех элементов изучаемых знаний, умений. Термин «действенность» имеет, как минимум, два синонима — валидность (от англ. valid — имеющий значение, ценный) и показательность, представительность, трактуемые так же, как репрезентативность. Всегда имеется в виду, что составитель теста обязан тщательно изучить все разделы учебной программы, учебные книги, хорошо знать цель и конкретные задачи обучения. Лишь тогда он сможет составить тесты, которые будут действенными для определенной категории обучаемых.

Четкая и ясная постановка вопроса в пределах освоенных знаний — неотъемлемое условие действенности теста. Если тест выходит за пределы освоенного содержания или же не достигает этих пределов, превышает запроектированный уровень обучения, то он не будет действенным для тех обучаемых, которым он адресован. Действенность теста определяется статистическими методами. Величина 0,7—0,9 свидетельствует о высокой действенности тестов обученности. Если же коэффициент корреляции достигает 0,45—0,55, то действенность теста считается удовлетворительной, при более низких значениях она считается неудовлетворительной.

Степень надежности характеризуется стабильностью, устойчивостью показателей при повторных измерениях с помощью того же теста или его равноценного заменителя. Количественно этот показатель характеризуется вероятностью достижения запроектированных результатов (правильностью значений). Грамотно составленные и апробированные тесты обученности позволяют достичь коэффициента надежности 0,9. Установлено, что надежность теста повышается при увеличении количества тестовых заданий.

Установлено также, что чем выше тематическое, содержательное разнообразие тестовых заданий, тем ниже надежность теста. Это следует понимать так: тест, нацеленный на проверку усвоения конкретной темы, всегда будет более надежным, чем тест, направленный на проверку всего раздела (курса), охватывающий значительное количество материала — закономерностей, концепций, фактов. Происходит это именно потому, что содержательное разнообразие последнего выше.

Надежность тестов обученности значительно зависит от трудности их выполнения. Трудность определяется по соотношению правильных и неправильных ответов на тестовые вопросы. Включение в состав тестов таких заданий, на которые все обучаемые отвечают правильно или же, наоборот, неправильно, резко снижает надежность теста в целом. Наибольшую практическую ценность имеют задания, на которые правильно отвечают 45—80% обучаемых.

Характеристика дифференцированности (различимости) связана с использованием таких тестов, где нужно выбирать правильный ответ из нескольких возможных альтернатив. Если, скажем, все учащиеся безошибочно находят правильный ответ на один вопрос и также дружно не могут ответить на другой, то это сигнал для совершенствования теста в целом. Его необходимо дифференцировать, сделать различимым. Иначе подобные задания не помогут отделить тех, кто усвоил материал на необходимом уровне, от тех, кто заданного уровня не достиг. Практически дифференцируют тесты по результатам статистического анализа, сравнивая результаты выполнения теста в целом с результатами выполнения отдельных заданий. Если коэффициент корреляции между ответами на конкретные задания и на тест в целом больше 0,5, то это свидетельствует о достаточной дифференцированности теста.

При подготовке материалов для тестового контроля необходимо придерживаться таких основных правил:

1. Нельзя включать ответы, неправильность которых на момент тестирования не может быть обоснована учащимися.

2. Неправильные ответы должны конструироваться на основе типичных ошибок и должны быть правдоподобными.

3. Правильные ответы среди всех предлагаемых ответов должны размещаться в случайном порядке.

4. Вопросы не должны повторять формулировок учебника.

5. Ответы на одни вопросы не должны быть подсказками для ответов на другие.

6. Вопросы не должны содержать «ловушек».

Тесты обученности применяются на всех этапах дидактического процесса. С их помощью обеспечивается предварительный, текущий, тематический и итоговый контроль знаний, умений, учет успеваемости, академических достижений.

Предварительный контроль. Успех изучения любой темы (раздела или курса) зависит от степени усвоения тех понятий, терминов, положений и т.д., которые изучались на предшествующих этапах обучения. Если информации об этом у преподавателя нет, то он лишен возможности проектирования и управления в учебном процессе, выбора оптимального его варианта. Необходимую информацию преподаватель получает, применяя предварительный контроль (учет) знаний. Последний необходим еще и для того, чтобы зафиксировать (сделать срез) исходный уровень обученности. Сравнение исходного начального уровня обученности с конечным (достигнутым) позволяет измерять «прирост» знаний, степень сформированности умений и навыков. Если известны входные и выходные характеристики системы, проблемы ее оптимизации считаются во многом решенными.

Собрать максимальный объем информации о входных характеристиках обученности и оценить их в количественных показателях удается с помощью тестирования, которое осуществляется с помощью специально разработанных для этой цели заданий.

Текущий контроль. Он необходим для диагностирования хода дидактического процесса, выявления динамики последнего, сопоставления реально достигнутых на отдельных этапах результатов с запроектированными. Кроме собственно прогностической функции текущий контроль и учет знаний, умений стимулирует учебный труд учащихся, способствует своевременному определению пробелов в усвоении материала, повышению общей продуктивности учебного труда.

Тематический контроль. Составление тематического тестового задания требует кропотливого и тщательного труда. Ведь речь идет не просто о проверке усвоения отдельных элементов, а о понимании системы, объединяющей эти элементы. Значительную роль при этом играют синтетические, комплексные задания, объединяющие вопросы об отдельных понятиях темы, направленные на выявление информационных связей между ними.

Итоговый контроль. Осуществляется во время заключительного повторения, а также в процессе экзаменов (зачетов). Именно на этом этапе дидактического процесса систематизируется и обобщается учебный материал. С высокой успешностью могут быть применены соответствующим образом составленные тесты обученности.

Естественно, не все необходимые характеристики усвоения можно получить средствами тестирования. Такие, например, показатели, как умение конкретизировать свой ответ примерами, знание фактов, умение связно, логически и доказательно выражать свои мысли, некоторые другие характеристики знаний, умений, навыков диагностировать тестированием невозможно. Это значит, что тестирование должно обязательно сочетаться с другими традиционными формами и методами проверки.