

## ОСНОВНЫЕ ЭМПИРИЧЕСКИЕ ПАРАМЕТРЫ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ

**Атоев Эhtiёр Худоёрович**

*Бухарский инженерно-технологический институт*

**Рузиев Бехрузжон**

*студент группы 201-22 ТНГ БухИТИ*

В последние годы проблеме тестирования уделяется большое внимание в области теории и методики обучения химии. Имеются многие учебно-методические разработки, в которых эта проблема анализируется: стандартизированные тесты, классифицируемые по форме их строения, метод, повышающий надежность результатов тестирования, примеры статистической обработки результатов контроля; типология тестов (дополнения, выборки, сличения, напоминания, группировки, ранжирования, альтернативный, профнаправленный, методика составления и использования тестов разного вида в формировании мотивации учения; стандартизированные тесты, сконструированные на конкретном материале учебного предмета химия, предназначенные для оценки уровня знаний, умений и навыков; требования к тестам обученности, характеристика эталонов к заданиям тестового типа, методика оценки и др.

В зависимости от диагностических целей и задач, специфики обследуемого контингента, формы представления и структуры композиций, программно-методического обеспечения и мастерства разработчика, дидактические тесты могут весьма различаться, как по типу, так и по качеству их составления. Но в любом случае эффективный дидактический тест (традиционный или компьютерный) – это набор тестовых заданий (субтестов) имеющие четко определенные цели диагностирования и проверенные на репрезентативной выборке с достаточно высокой надежностью и валидностью.

Эмпирическое знание - это совокупность высказываний о реальных, эмпирических объектах. Эмпирическое знание основывается на чувственном познании. Рациональный момент и его формы (суждения, понятия и др.) здесь присутствуют, но имеют подчиненное значение. Поэтому исследуемый объект отражается преимущественно со стороны своих внешних связей и проявлений, доступных созерцанию и выражающих внутренние отношения. Эмпирическое, опытное исследование направлено без промежуточных



звеньев на свой объект. Оно осваивает его с помощью таких приемов и средств, как описание, сравнение, измерение, наблюдение, эксперимент, анализ, индукция (от частного к общему), а его важнейшим элементом является факт (от лат. factum — сделанное, свершившееся). Наблюдение — это преднамеренное и направленное восприятие объекта познания с целью получить информацию о его форме, свойствах и отношениях. Процесс наблюдения не является пассивным созерцанием. Это активная, направленная форма гносеологического отношения субъекта по отношению к объекту, усиленная дополнительными средствами наблюдения, фиксации информации и ее трансляции. К наблюдению предъявляются требования: цель наблюдения; выбор методики; план наблюдения; контроль за корректностью и надежностью полученных результатов; обработка, осмысление и интерпретация полученной информации. Измерение - это прием в познании, с помощью которого осуществляется количественное сравнение величин одного и того же качества. Качественные характеристики объекта, как правило, фиксируются приборами, количественная специфика объекта устанавливается с помощью измерений.

Для эмпирического анализа тестовых заданий, прежде всего, необходимы сведения о степени сложности каждого задания, его селективности и о выверке дистракторов (отвлекающих вариантов ответа – подсказок в закрытом тесте с выбором вариантов). Рассмотрим кратко методы определения или оценки основных параметров, определяющие дидактическую эффективность и диагностичность тестовых заданий:

1) Значение **коэффициента усвоения** ( $K_{уcв.}$ ), одного из наглядных показателей диагностической эффективности теста зависит от структурной сложности, "процентного веса" и значимости тестового задания в общем комплексе диагностируемых знаний и умений. В первом приближении эту величину можно оценить в интервале от 0 до 1 по формуле:

$$K_{уcв.} = \frac{n(\text{количество правильных ответов})}{N(\text{общее количество тестируемых})}$$

Беспалько В.П., предлагает вместо  $K_{уcв.}$  вводить понятие "коэффициента качества усвоения" ( $K_{уcв.}$ ), определяемый по формуле:

$$K_{уcв.} = \frac{a(\text{количество выполненных операций})}{p(\text{общее количество операций в задании})}$$

$K_{уcв.}$  - уровень усвоения, определяемый уровнем деятельности.



Более точных и объективных методов определения коэффициента усвоения и его взаимосвязи с типом и уровнем сложности тестовых заданий пока еще не разработаны, хотя в ряде работ В.И. Михеева предложены методы оценки обобщенного показателя успешности обучения (5).

**2) Степень сложности** ( $P_{\text{сложн.}}$ ) заданий определяют в процентах от общего количество респондентов:

$$P_{\%} = \frac{\text{Кол} - \text{вореспондентов, правильно решивших задание}}{\text{Общее количество респондентов}} \%$$

Так, если задание правильно решили 80% испытуемых, степень его сложности равна 80, т.е. чем легче задание, тем больше ( $P_{\%}$ ) в числовом выражении. Оптимальным считается  $P = 50\%$ , приемлемым являются величины в интервале от 20 до 80 %, при  $P \leq 5\%$  или  $P \geq 95\%$  данное задание исключается из теста.

Такая интерпретация степени сложности заданий не отвечает современным требованиям педагогической диагностики. Как известно, из классических работ Талызиной Н.Ф, Беспалько В.П, Ерецкого М.И. и других авторов, степень сложности задания является объективной характеристикой тестовых заданий не зависящей от уровня подготовленности тестируемых, месторасположения задания в структуре всего теста (батареи тестов) и от последовательности его представления.


Для оценки степени сложности тестовых заданий, необходимо, на наш взгляд, хотя бы говорить требуемый уровень усвоения дидактического теста. Например, для тестов I уровня можно допустить, что количество существенных операций (шагов), необходимых для решения задания ( $P = 1$ ) на первом уровне деятельности, и сложность задания также равна 1. Во всех остальных случаях, определение степени сложности по Ингенкампу К. нельзя считать корректной (4).

**3) Коэффициент селективности** ( $K_{\text{sel.}}$ ) отражает взаимосвязь данного задания с решением всех заданий теста и является одним из основных показателей его диагностической ценности.

Согласно К. Ингенкампа (4) при первом приближении можно принять, что:

1. а)  $K_{\text{sel.}} \rightarrow 1$ , если обучаемые не выполнившие определенное задание не могут решить и все остальные задания ;
2. б)  $K_{\text{sel.}} \rightarrow 0$ , если хорошо успевающие студенты не выполняют данное задание также часто, как и неуспевающие их коллеги;
3. в)  $K_{\text{sel.}} \rightarrow -1$ , если слабые студенты решают данное задание лучше, чем сильные.





Случай (в) часто наблюдается при тестировании с помощью заданий " закрытого типа" с готовыми, но неудачно сформулированными ответами.

Если разность между количеством правильных решений в лучшей (А) группе и в худшей группе (В) составляет 10 %, то селективность считают "удовлетворительной", ~ 15 % – "хорошей". На наш взгляд, задания с величиной  $K_{sel.} \leq 0,3$  должны исключаться из теста, как не селективные. Однако, в случае использования тестов различного уровня сложности, весьма проблематично выявление логической взаимосвязи между заданиями с помощью данной величины.

Особенностью этой величины является то, что она связана с индексом различения (ID) тестовых заданий. В частности, как указывает ведущий тестолог США А. Анастази (1), величина ID прямо пропорциональна разности между числом **правильных** и **ошибочных** различений, выявляемых заданием. Это позволяет оценить диагностическую ценность тестового задания с помощью двух взаимосвязанных параметров.


**4) Сбалансированность дистракторов тестовых заданий** является главным показателем качества их составления и характеризуется равномерностью распределения внимания респондентов между отвлекающими вариантами ответа – эталона.

В тестах с набором вариантов ответа, ни один дистрактор не должен быть столь невероятным, чтобы его выбрали менее 5% респондентов. Если такое произошло, дистрактор **необходимо заменять** или перестраивать данное тестовое задание и проверять его снова. Таким образом, сбалансированность тестового задания однозначно характеризуют научно-методический уровень и квалификацию разработчика теста.

**5) Скорость выполнения** задания является одним из основных факторов при организации и проведения дидактического тестирования. В компьютерных дидактических тестах необходимо также учитывать скорость и качество выполнения заданий. Время в педагогических процессах относится к числу основных организующих факторов, хотя до сих пор, "в педагогике отсутствует понятие **времени** как педагогической категории" (2).

На наш взгляд, во всех контролирующие-тестирующих программах должны быть предусмотрены возможности контроля и фиксации рационального использования контрольного времени выполнения заданий, скорость и качество выполнения должны поощряться начислением "премиальных" или "штрафных" баллов. Это позволило бы оценить косвенно степень трудности и





уровень сложности заданий, качества подготовленности обследуемых, варьировать количество заданий в тестах и субтестах.

**6) Степень трудности** ( $Q_{\text{трудн.}}$ ) – тестовые задания являются тем параметром, от точности определения которого зависит степень завершенности и полнота понятийного аппарата дидактической тестологии.

Понятие "трудность" в дидактике неоднозначно и имеет как объективные, так и субъективные компоненты. Объективный компонент трудности задания обусловлен степенью (уровнем) сложности, теоретической глубины и структурированности задание, его терминологической, математической и графической (символьной) насыщенностью, четкостью логической структуры. Субъективной компонент трудности (для тестируемого) зависит, в первую очередь, от уровня базовой подготовки обследуемого к тестовому контролю знаний, а во вторых от условий тестирования (шум, освещенность, присутствия представителей кафедры, деканата, регламент проведения теста и др.).

На основе изложенного можно констатировать незавершенность математического аппарата анализа качества тестовых заданий и необходимость разработки более совершенных методик на основе рационального применения компьютерных средств.

#### ЛИТЕРАТУРА:

1. Атоев Э. Х., КУРБАНОВ М. Т. ЭКСПЕРТНАЯ ОЦЕНКА КАЧЕСТВА ПРЕДМЕТНЫХ ТЕСТОВЫХ ЗАДАНИЙ-ВАЖНЫЙ АСПЕКТ ПРИ ОБУЧЕНИИ ГУМАНИТАРНЫХ НАУК //Поколение будущего: Взгляд молодых ученых-2014. – 2014. – С. 258-259.

2. Атоев Э. Х. СТАНДАРТИЗАЦИИ ПРОЦЕДУР ДИДАКТИЧЕСКОГО ТЕСТИРОВАНИЯ //Аллея науки. – 2019. – Т. 5. – №. 1. – С. 168-172.

3. Атоев Э. Х., Гайбуллаев Х. С. АНАЛИЗ РЕЗУЛЬТАТОВ ДИДАКТИЧЕСКОГО ТЕСТИРОВАНИЯ //ИННОВАЦИИ, КАЧЕСТВО И СЕРВИС В ТЕХНИКЕ И ТЕХНОЛОГИЯХ. – 2014. – С. 22-25.

