

Инструменты диагностики качества знаний учащихся

В.П. Беспалько, академик Российской академии образования, доктор педагогических наук, профессор педагогики

Лекция № 3. Содержание лекции

Краткая ретроспектива лекций № 1 и № 2
Требование валидности измерения качества обучения
Процесс измерения и оценки качества знаний учащихся

1. Краткая ретроспектива лекций № 1 и № 2

Люди издревле применяют в своей трудовой деятельности всевозможные инструменты для проверки качества своей работы. Не буду здесь приводить примеры всем известных инструментов, используемых и медиком, и инженером, и агрономом. Важно, однако, особо подчеркнуть, что потрясающий прогресс во многих отраслях промышленности, сельского хозяйства, медицины, пищевой промышленности и других видах деятельности человека был бы невозможен, если бы этот прогресс не предварялся адекватным прогрессом соответствующих средств измерения. Отсюда логически следует вывод о том, что без соответствующих средств измерения качества продуктов человеческой деятельности никакой научно-технический прогресс был бы невозможен. На глазомерных оценках велосипеда не построишь и тем более в космос не полетишь.

Если это так, то вполне понятно, почему столетиями и уже тысячелетиями образование топчется на месте, а люди не становятся всё более образованными и воспитанными, а человеческое общество всё более человеческим. Причина этого явления предельно проста: не имея в своём распоряжении точных измерительных инструментов для диагностики качества знаний учащихся и их воспитанности, довольствуясь глазомерными измерениями, мы пребываем под гипнозом иллюзий и привычек, вместо осознанного и точного знания. И деградируем вместе с этими иллюзиями. Одна из таких катастрофических иллюзий состоит в том, что нам кажется, что мы знаем, что такое «хорошее образование» и на глазок определяем, насколько оно хорошее. Ниже мы рассматриваем проблему инструментальной диагностики качества знаний учащихся с опорой на приведённые в первой и второй лек-

Что измерено, то понято.

Тесты как инструменты объективного оценивания знаний учащихся
Структура теста
Создание тестов различного уровня
Тест лестница
Тесты, экзамены и демократия

циях исходные положения.

Для лучшего восприятия материала этой лекции полезно напомнить следующие сведения из **первой** лекции: *понятие о структуре личности и возможности диагностического целеобразования по отдельным её свойствам.*

Из **второй** лекции полезными будут сведения об объективных *параметрах диагностической цели.* Именно благодаря открытию этих параметров нам удалось построить непротиворечивую *теорию и практику* объективного контроля и оценки знаний учащихся. Последние открывают безграничные возможности осознанного и реального совершенствования образования в целом.

2. Требование валидности измерения и оценки качества обучения

Слово «валидность» означает соответствие правилу или истине.

Можно говорить о валидном выводе или доказательстве, поступке или действии. Говоря о валидном измерении и оценке знаний учащихся, подразумевают, что они осуществляются по вполне определённым правилам и результаты этих действий *объективно, точно и однозначно* характеризуют подготовленность учащегося в данном предмете или к данной деятельности. К примеру, ученик получил оценку «пять» по предмету при выпуске из школы, а при поступлении в вуз его знания были оценены «четвёркой» или, того хуже, «тройкой». Что это означает? Это означает, что одна из его оценок или обе вместе не валидны («инвалидны»), т.е. они ничего определённого о действительной подготовке учащегося не говорят. В школе оценивание велось по одним признакам и своими методами, а в вузе и признаки, на которые ориентировались экзаменаторы, и методы оценивания могли быть совершенно другими. Что часто и бывает.

Как появляется инвалидная оценка? Инвалидность оценки имеет два источника.

Первый. Качество знаний проверялось с отступлением от обозначенного программой содержания обучения. В школе это могло быть из-за сокращения учителем числа учебных элементов (УЭ), на базе которых проводился экзамен. В вузе, наоборот, экзаменаторы часто выходят за пределы школьных программ, предлагая учащимся «догадываться» о существовании ещё и других УЭ, не вошедших в школьную программу, но известных экзаменаторам. В обоих случаях это отступление от *содержательной валидности* измерения и оценки знаний учащихся, которая требует, чтобы *проверка и оценка знаний осуществлялась на тех же учебных элементах, на которых проводилось обучение*.

Приведу один яркий пример, как отступление от содержательной валидности теста подрывает нормальную работу важнейшей отрасли народного хозяйства. Американская медицина испытывает острую нужду в медицинских сёстрах. Для получения права работать медсестрой надо сдать экзамен. Но как можно сдать экзамен, который предваряется почти Райкинской фразой: «Чтобы сдать экзамен, вам недостаточно тех знаний, которые вы получили в колледже: вам придётся самостоятельно додумываться до правильного ответа.» И далее, вместо содержательно валидных тестов, испытуемым предлагаются ситуации с множественными правильными решениями, из которых надо выбрать «наиболее правильный», ни в каком пособии не описанный и известный только изобретателю теста. В результате даже врачи-иностранцы не могут сдать этот иезуитский экзамен, и страна продолжает остро нуждаться в медсёстрах, а больные лишены медицинской помощи. На таких же принципах построены и многие тесты SAT, TOEFL и им подобные волонтаристские изобретения (см. Лекцию № 4).

Вторым источником отступления от валидности оценки является:

а) предложение выполнить действия, отличающиеся от заданной программой диагностической цели по α , β , τ , ψ (если цель задана диагностично). Например, цель α_1 , а тест α_2 ;

б) предложение выполнить действия, которые заведомо не могли быть сформированы предшествующим учебным процессом.

Этими отступлениями нарушается так называемая *функциональная валидность* измерения и оценки качества знаний учащихся (типичное явление при традиционном обучении и субъективных методах контроля).

Итак, измерение и оценка качества знаний учащихся должны удовлетворять требованию валидности этих процедур. В противном случае школьники либо незаслуженно поощряются либеральным экзаменатором, либо столь же незаслуженно наказываются слишком строгим испытателем. И то и другое вредно для учащегося и вредно для образования.

3. Процесс измерения и оценки знаний учащихся

Для более осознанного понимания процесса измерения и оценки знаний учащихся надо ясно представлять себе процесс их формирования. На рисунке 1 схематически изображён этот процесс. Важно усмотреть в этой схеме, что это переходный процесс восхождения к мастерству: *формирование каждого следующего уровня усвоения начинается только после некоторой сформированности предшествующего уровня и эта взаимосвязь уровней усвоения сохраняется в ходе всего процесса обучения. Тот же механизм взаимодействия уровней усвоения сохраняется и в процессе забывания, но в обратном направлении.*

Обратите особое внимание на соотношение времён, необходимых для достижения разных уровней усвоения деятельности на одних и тех же УЭ. Это соотношение выражается пропорцией квадратов номеров уровней усвоения:

$$T_1 : T_2 : T_3 : T_4 = (\alpha_1)^2 : (\alpha_2)^2 : (\alpha_3)^2 : (\alpha_4)^2 = 1 : 4 : 9 : 16.$$

$$T\alpha_1 : T\alpha_2 : T\alpha_3 : T\alpha_4 = 1 : 4 : 9 : 16$$

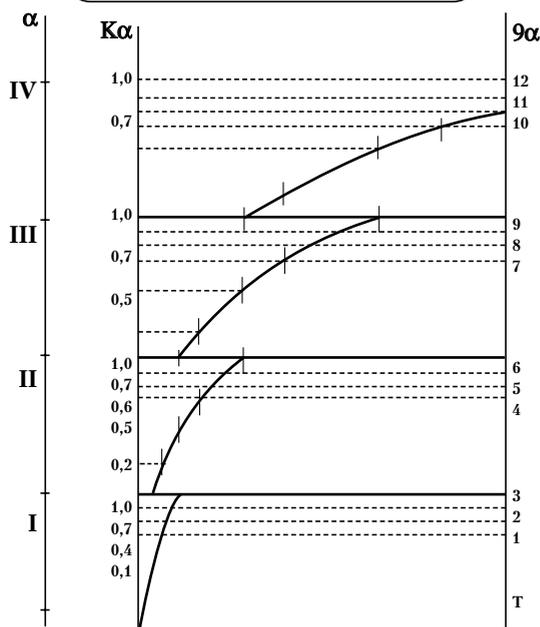


Рис. 1

Это означает, что для подъёма на второй уровень мастерства учащемуся потребуется в **четыре** раза больше времени, чем для достижения первого уровня. Ему потребуется в **девять** раз больше времени, чтобы достичь третьего уровня усвоения, и в **шестнадцать** раз больше времени, чтобы достичь четвёртого уровня (при условии соответствующей одарённости). Отсюда понятна абсурдность одинаковой оценки за выполнение тестов разных уровней, практикуемое в отечественной и зарубежной практике оценивания знаний.

Примечателен участок кривой восхождения между точками 0 и 0.7. Он отражает процесс начальной организации (усвоения) опыта учащимся. Этот процесс завершается по достижении учащимся $K\alpha = 0.7$, когда можно считать, что процесс управляемого обучения завершился и начался процесс самообучения. Т.е. учащийся настолько усвоил предмет на данном уровне усвоения, что в ходе его последующей учебной и практической деятельности он сумеет самостоятельно видеть свои ошибки и исправлять их, улучшая качество усвоения предмета. **При $K\alpha < 0.7$, процесс обучения на данном уровне не закончен**, и из-за неспособности учащегося видеть и исправлять свои ошибки, при прекращении обучения, совершенствования деятельности не происходит, и учащийся работает с систематическими и неистребимыми ошибками. Стремление дать учащимся возможно больший объём знаний заканчивается недостаточным его усвоением ($K\alpha < 0.7$) со всеми вытекающими из этого последствиями для учащегося и общества.

Вот уж действительно: «лучше меньше, да лучше».

Вернемся, однако, ещё раз к рис.1 и обратим внимание на процессы формирования каждого уровня усвоения. Из рисунка хорошо видно перекрытие кривых, означающее, что формирование каждого следующего уровня усвоения не ждёт полного завершения формирования предыдущего уровня. Так, уже при $K_1 = 0.4$ начинается формирование второго уровня усвоения и, когда K_1 достигает 0.7, $K_2 = 0.2$, а при $K_1 = 1.0$, $K_2 = 0.5$ и начальная организация опыта учащегося на втором уровне приближается к своему завершению. То же самое происходит и на других уровнях усвоения: перекрытие кривых свидетельствует о постепенности процесса обучения. Попытка сразу «прыгнуть» на второй – третий уровень усвоения, минуя первый уровень, является иллюзорной, т.к., пока не сформируется устойчивый первый уровень ($K_1 > 0.7$), все по-

пытки формирования второго уровня будут тщетными, а неуправляемое, попутное с формированием второго уровня, восхождение учащегося на первый уровень потребует больших усилий и времени.

Измеряться и оцениваться знания могут различными способами. Это может быть *глазомерная* процедура, при которой экзаменатор в уме в свёрнутом и сокращённом виде выполняет операции оценивания. Это *субъективный* подход к оцениванию со всеми присущими ему иллюзиями, неизбежными нарушениями валидности контроля и влиянием отношения экзаменатора к ученику и его ответу. При субъективном подходе к измерению и оценке знаний учащихся результат процесса **не воспроизводим** и при повторных пробах может варьироваться в широких пределах.

Примерами субъективного подхода к оцениванию знаний могут служить все виды устного опроса, диктанты, изложения, сочинения, решения задач, если при их выполнении не соблюдается последовательность операций процесса оценивания.

Другой подход, который мы назовём *объективным*, состоит в том, что для *выявления* наличия знаний всегда используется *адекватный инструмент*, позволяющий соблюсти содержательную и функциональную валидность пробы. К примеру, для проверки умения читать используется задание на *чтение*, а не на понимание, интерпретацию или реконструкцию текста. В печати часто можно прочесть, что американские школьники настолько безграмотны, что даже в старших классах не умеют читать. Это следствие содержательно инвалидного тестирования, при котором вместо проверки умения читать проверяется понимание учащимся многозначной семантики и логики текста со специально завуалированным межстрочным смыслом. В одной из американских публикаций, рассматривающей причины (мнимые) плохой подготовки абитуриентов вузов, утверждается, что этой причиной является их неумение читать *сложные* тексты. В публикации расшифровывается, что такое *сложный* текст. Это текст в котором:

- а) *смысл сообщения выражен неявно*, а скрыт внутри текста;
- б) *отношения между персоналиями или идеями не очевидны* и о них можно только догадываться;
- в) *словарь сообщения нетипичен и состоит из редких слов*.

Возникает естественный вопрос: кто виноват в том, что абитуриент проваливается на вступи-

тельном экзамене? Во многих случаях вина за провалы испытуемых падает на «шифровальщика» текстов, грубо нарушающего содержательную валидность теста. В школе учащиеся читают тексты, которые тщательно отредактированы для того, чтобы в них был *явно выражен смысл, чтобы были явно выражены отношения между персоналиями или идеями, и использует словарь общепотребительной лексики*. Для чего же на приёмных экзаменах в вуз испытывать абитуриентов инвалидными «сложными» текстами?

При объективном подходе к оценке знаний учащихся после выявления *ранее усвоенных знаний* всегда измеряется абсолютный показатель его качества (сколько знаний почерпнул ученик в учебном процессе). В том же примере оценки качества чтения это может быть объём прочитанного, число допущенных ошибок, время на прочтение текста.

Измерение относительного показателя качества знаний состоит в определении различного рода коэффициентов, позволяющих сопоставлять и сравнивать успехи различных учеников и одного и того же ученика в разные периоды обучения. Обращаясь к тому же примеру с чтением, назовём такие его относительные показатели, как отношение объёма прочитанного ко времени чтения или число допущенных ошибок определённого рода к объёму прочитанного или времени чтения. Во второй лекции мы ввели два относительных показателя качества обучения: коэффициент усвоения $K\alpha$ и коэффициент автоматизации $K\tau$. В той же второй лекции было показано, как использование относительного показателя качества усвоения позволяет *соотнести* его с определённой шкалой. Очень удобной шкалой для суждения о качестве знаний оказалась 12-балльная шкала отношений относительного $K\alpha$.

По известному коэффициенту усвоения $K\beta$ не составляет труда выполнить последнюю операцию *объективного* процесса измерения и оценки знаний ученика: присвоить измеренному знанию соответствующую оценку, которая однозначно показывает уровень усвоения и его качество.

Главное достоинство объективного измерения и оценки знаний учащихся состоит в **воспроизводимости** результата оценивания при любом числе повторных проб (при надлежащей, конечно, надёжности измерения).

Субъективные методы контроля качества знаний учащихся я бы рекомендовал применять

только для текущей классной работы и получения самого грубого представления об успешности проведённого занятия.

Для итогового оценивания качества знаний учащихся и косвенной оценки качества занятия надо использовать только *объективные инструментальные методы*.

С учётом приведённого анализа процесса восхождения по уровням усвоения и способов оценки знаний учащихся, рассмотрим процесс создания и применения **тестов успешности усвоения** как инструментов объективного оценивания. Этот процесс состоит из следующих операций:

- выявление наличия знаний;
- измерение абсолютного показателя качества знаний;
- определение относительного показателя качества знаний;
- соотнесение относительного показателя качества с принятой шкалой оценки;
- присвоение качеству соответствующей оценки.

4. Тесты как инструменты объективного оценивания знаний учащихся

История тестирования и тестов восходит к началу XX века, когда два известных французских психолога А. Бине и Т. Симон впервые предложили тесты для оценки качества интеллекта личности. До инициативы Бине-Симона тесты использовались для изучения свойств других психических процессов и не касались высших психических функций человека: мышления и в целом его интеллектуальных операций. Предложение французских психологов осуществлять *опосредованное* измерение и оценку качества человеческого интеллекта было встречено мировой психологической общественностью с энтузиазмом, т.к. открывало новые возможности исследования высших психических функций человека и давало им в руки вполне осязаемый инструмент их диагностики. Чего психологи не хотели замечать с самого начала и на что закрывают глаза до сих пор, так это отсутствие хоть какой то бы ни было теоретической базы под широко развернувшимся тесто-творчеством. Даже само понятие интеллекта, лежащее в основе исходного Бине-Симона теста, до сих пор не имеет однозначного определения в психологии.

С тех пор как Бине-Симон предложили свой первый тест уткло много воды и времени, опубликовано много разных тестов, на разные

лады трансформирующие исходный Бине-Симона тест. В Европе и Америке развилась целая индустрия тестирования и образовался широкий рынок тестов. Производство и применение тестов поставлено на поток. Было изобретено множество разновидностей тестов, среди которых наибольшую популярность получили *интеллектуальные, личностные, способностей, достижений, профессиональной пригодности, психического здоровья и пр. разновидности тестов*. Примечательно, что за все сто лет составления и применения тестов не сложилась сколько-нибудь различимая психолого-педагогическая теория создания тестов и тестирования. Тесты создаются различными авторами и организациями по своему собственному усмотрению и копируются по внешним признакам многочисленными предпринимателями. Неудивительно, что такое массовое коммерческое развитие деятельности тестирования привело к тому, что предлагаемые тесты не выдерживают требований валидности и не воспроизводят объективного процесса измерения и оценки тестируемого качества. Ниже мы покажем эти дефекты массового тестирования на примерах американского и российского опыта (см. Лекцию № 4).

Но, прежде всего, необходимо познакомиться с теорией тестирования как она сложилась в наших исследованиях, проведённых на базе современной психолого-педагогической науки.

Из всех ныне различаемых видов тестов нас будут интересовать только тесты непосредственно обслуживающие учебный процесс, так называемые тесты **достижений**.

Тесты **достижений** выявляют **только** качество усвоения учеником учебного материала, который ему по условиям обучения (программа, учебник) должен быть заведомо известным (*содержательная валидность*); эти тесты легко можно соотнести с параметрами и критериями, отражающими этапы процесса усвоения знаний учащимися (*функциональная валидность*).

Теория построения и применения тестов базируется на представлении (**постулате**) о том, что процесс усвоения знаний и умений совершается постепенно во времени и является последовательным восхождением ученика по уровням мастерства (α) в изучаемом предмете. Следовательно, задачей теста *становится выявление, на каком уровне достижений ученик находится в данное время*. Эта задача приводит нас к выводу о том, что не может существовать единый тест для выявления качества усвоения деятельности учащимся: тест должен соответствовать,

быть валидным уровню диагностируемого мастерства.

К сожалению, в мировой практике делается попытка выявить качество усвоения мастерства учеником в учебном предмете с помощью единственного теста, содержащего неупорядоченную смесь инвалидных тестов разного уровня. Полученный таким путём показатель качества усвоения не даёт правильного представления о действительном качестве подготовки учащегося. Он (показатель), кроме того, не обладает свойством стабильности, так как при другой смеси, казалось бы, тех же тестов может получиться результат, прямо противоположный предыдущему тестированию. Такие, кстати, «непонятные» вещи происходят в настоящее время в системах образования Америки и России. Обе страны внедряют итоговое тестирование выпускников общеобразовательной школы. Однако учащиеся, успешно преодолевшие школьные тесты, проваливаются на тестах при поступлении в высшие учебные заведения. Школьные и вузовские педагоги в недоумении обвиняют друг друга, а ларчик открывается очень просто: в школьной смеси тестов больше тестов первого уровня, а в вузовской — второго и даже третьего уровней. Вот поэтому-то даже успешные школьники проваливаются на экзаменах в вузы.

Выход из положения состоит, полагаю, в том, чтобы, изучив излагаемую здесь теорию тестирования, установить требуемый и достижимый уровни усвоения предметов школьниками и затем содержательно и функционально применять валидные тесты для диагностики конечного уровня усвоения в школе, а начального — в вузе.

5. Структура теста

Любой тест должен состоять из двух отделённых друг от друга частей:

Задания и Эталона.

Символически сказанное можно записать в виде следующей формулы:

Тест (Т) = Задание (З) + Эталон (Э).

Испытуемый получает задание и, выполняя его, формулирует своё решение. *Ответ*. Эталон находится в распоряжении экзаменатора или заложен в некоторый экзаменатор (компьютер). Ответ испытуемого сравнивается с эталоном и по принятой для данного теста методике выносятся суждение о тестируемом качестве.

Процедуру тестирования можно изобразить в виде следующей схемы (рис. 2):



Рис. 2.

Приведённые формула и схема тестирования выглядят тривиально, и это действительно так. Нетривиален, однако, их дидактический смысл, который состоит, во-первых, в стремлении создать у читателя прочное представление о том, что тест без эталона это не тест, а обычное контрольное задание, подверженное всем случайностям глазомерной оценки. Во-вторых, столь же прочно утвердить в сознании читателя мысль о том, что в тестировании сравнение ответа ученика с эталоном и подсчёт $K\alpha$ **обязательны**. Суждение о качестве выполнения теста без сравнения с эталоном превращает объективную процедуру контроля в субъективную со всеми присущими последней недостатками. Конечно, сопоставление с эталоном, если тестирование осуществляется вручную, создаёт для экзаменатора дополнительную нагрузку по сравнению с глазомерной оценкой того же ответа ученика, но зато делает тестирование качества знаний объективным.

Задание теста создаётся в полном соответствии с требованиями содержательной и функциональной валидности тестирования. Это значит, что задание теста безусловно выполнимо на основе изученного в классе содержания предмета и по конструкции соответствует определению диагностируемого уровня усвоения. Отсюда следует, что тесты должны различаться по **четырёх уровням** мастерства. Другие параметры качества знаний характеризуют их различные особенности и учитываются при создании тестов, но на конструкцию теста не влияют. Так, можно говорить о наборах тестов на различных ступенях абстракции или степени осознанности, различных по надёжности или автоматизации, но это не изменит дифференциацию тестов по четырём уровням усвоения.

Для удобства пользования эталоном при его сопоставлении с ответом испытуемого в эталоне указываются **существенные операции** теста, т.е. те операции, выполнение которых становится полным и правильным решением данного задания. Только существенные операции теста

учитываются при вычислении коэффициента усвоения $K\alpha = n / p$. Число «n» означает число *правильно выполненных существенных операций* теста; число «p» означает *общее число существенных операций* в тесте.

В зависимости от требуемой надёжности тестирования создаются соответствующие **батареи** тестов **данного уровня усвоения**. Общее число существенных операций в такой батарее выбирается по специальному графику надёжности тестирования, пример которого будет показан ниже.

6. Создание тестов различного уровня

При создании тестов успешности обучения следует руководствоваться тремя наиболее общими требованиями к ним: *быть содержательно валидными; быть функционально валидными; быть простыми*.

Требование *содержательной* валидности теста состоит в том, что можно контролировать только то знание, которому ученик был обучен. Другими словами, при проверке знаний и умений ученика нельзя выходить за пределы того содержания, которое он изучал в учебном процессе.

Если вузы считают, что содержательная подготовка учащихся недостаточна, им следует винить в этом Министерство образования, а не наказывать абитуриентов, поскольку это тупиковый путь получения лучшего абитуриента.

При существующих методах субъективного контроля качества знаний нередко можно услышать от ученика в ответ на вопрос учителя реплику: «А мы этого не изучали...». На что немедленно следует категорическое: «Попробуй сообразить (догадаться)». Такой диалог с большой вероятностью свидетельствует о нарушении учителем содержательной валидности контроля знаний учащихся. Скорее всего учитель просто не замечает, когда он переступает грань между своим знанием предмета, позволяющим ему «догадываться», и знанием предмета учеником, кому такие догадки не по силам. Чтобы избежать ошибок содержательной валидности теста, необходимо содержание обучения представить в виде логической структуры или таблицы учебных элементов, а затем уже строить тесты, сопоставляя предлагаемые в них задания с наличием в содержании обучения соответствующей информации о тестируемых учебных элементах (свойствах объектов, сущности явлений, алгоритмах и методах деятельности).

Под функциональной валидностью теста понимают его соответствие тому уровню деятельности, который тестируется. Ниже, рассматривая конструирование тестов разного уровня, будет показано, при каких обстоятельствах нарушается это требование к тесту.

Требование простоты теста состоит в том, что каждый тест должен состоять только из одной задачи данного уровня деятельности. Трудно объяснить почему, но часто наблюдается, вероятно, неосознанное стремление создателя теста усложнить его, соединяя несколько задач в одном тестовом кадре. В этом случае возникает часто непреодолимая трудность понять происхождение ошибки в ответе ученика: незнание предмета, непонимание вопроса или невниманность к деталям.

С учётом названных требований к тестам перейдём к примерам их конструирования. Подчеркнём: **конструирования**, а не придумывания или изобретения!

6.1. Тесты первого уровня

По определению, первый уровень усвоения — это деятельность по узнаванию ранее изученной информации при повторном её восприятии. Следовательно, тесты первого уровня должны предъявлять ученику ранее изученный учебный материал в различных видах для узнавания. Различают три наиболее типичных способа конструирования задания для тестов первого уровня: опознание, различение и классификация (соотнесение).

Самый простой способ — это задание на **опознание**, когда ученику необходимо лишь подтвердить (да — нет) узнавание. Вот примеры тестов на опознание ранее изложенного в этих лекциях материала:

1) **Задание (З):** Является ли педагогика наукой о педагогических системах?

Эталон (Э): Да.

Число существенных операций (р): 1 — выбор из альтернативы «да — нет».

Этот тест содержательно валиден, т.к. в первой лекции обсужден этот УЭ. Тест также функционально валиден деятельности первого уровня, так как не требует от ученика никаких действий, кроме опознания ранее изученного материала. Тест простой, т.к. в тесте содержится только одна задача.

2) **З:** Можно ли сказать, что учебник является техническим средством обучения?

Э: Можно.

Р: 1.

3) **З:** Является ли формула $K_t = T_{пр} / T_{уч}$ коэффициентом усвоения?

Э: Нет.

Р: 1.

Заметьте: В этом тесте напрашивается продолжение: «Что означает эта формула?» Включение этого продолжения в тест немедленно нарушит его функциональную валидность первому уровню деятельности (продолжение это уже второй уровень) и нарушит требование простоты теста (появляется вторая задача).

4) **З:** Можно ли по результатам тестирования на первом уровне судить о качестве усвоения деятельности на втором уровне?

Э: Можно.

Р: 1.

5) **З:** Являются ли врождёнными социальные свойства личности?

Э: Нет.

Р: 1.

Легко установить содержательную и функциональную валидность этих тестов: они построены на изложенном выше материале и деятельность по их разрешению — это узнавание ранее изученного материала, т.е. деятельность первого уровня. Тесты сопровождаются эталоном, и в них указано число существенных операций.

Не будут содержательно валидны следующие тесты, построенные на заведомо не знакомом ученику материале, например:

Может ли учащийся самостоятельно преодолеть сложный тест?

Может ли учащийся самостоятельно преодолеть трудный тест?

Не будет функционально валиден первому уровню и следующий тест, поскольку для его разрешения необходима деятельность более высокого уровня:

Усвоил ли ученик учебный предмет, если в батарее тестов с 22 существенными операциями он допустил шесть ошибок?

В этом тесте нечего узнавать. Чтобы получить ответ на поставленный вопрос, надо выполнить несколько промежуточных действий и по их результату дать ответ. По условиям же деятельности на первом уровне узнавание ответа должно

произойти с места и сразу, поскольку ответ должен содержаться в готовом виде в самом задании. Этот тест функционально не валиден тесту первого уровня.

Тесты, похожие по внешнему виду на тесты опознания, но содержательно или функционально не валидные им, называют *псевдо-тестами* на опознание.

Тесты различения — также тесты первого уровня: в них требуется выполнить многократные действия опознания правильных ответов (указание на неправильный ответ — это правильный ответ). Вот пример теста на различение:

Укажите, по какой формуле вычисляется коэффициент усвоения:

а) $K\alpha = n/p$; б) $K\beta = n/p$;

в) $K\tau = T_{пр} / T_{уч}$; г) $K\pi = a/b$.

При работе с тестом по методу «бумага — карандаш»:

Э. а) + б) — в) — г) -

$p = 4$.

При использовании компьютера для тестирования:

Э. а) +

$p = 1$.

В тестах на различение следует учитывать метод предъявления теста ученику: бумажный ли это вариант или машинный. При бумажном варианте предъявления теста студенту обычно выдаётся страница с написанными на ней тестами и таблица (матрица), в которой он заштриховывает клетки, соответствующие правильным ответам. Затем на таблицу накладывается специальный шаблон с отверстиями, в которых будут видны клетки, которые надо штриховать. Подсчитав число заштрихованных клеток, находят число правильных ответов — «n». Если в этом варианте работы с тестом считать существенными операциями теста на различение только правильные выборы, то можно исказить оценку знаний учащихся.

Первый вариант искажения получится, если учащийся, кроме правильного квадрата, заштриховал ещё и неправильный и его ошибка никак не учитывается. Возможен и второй вариант, найденный самими учеником: заштриховываются все квадраты, и он получает высший балл. Чтобы избежать ошибки оценивания в бумажном варианте теста, а ученику объясня-

ют, что при заполнении матрицы следует *плюсом* (+) отмечать правильные ответы и *минусом* (-) — неправильные. Числом существенных операций в тесте (p) является сумма всех альтернатив, а числом правильных ответов — сумма правильно поставленных плюсов и минусов.

При машинном тестировании существенной операцией считается выбор правильного ответа, поскольку после любого выбора, правильного или неправильного, экзаменатор отключается. Общее число существенных операций в батарее тестов (p) в этом случае равно сумме всех правильных ответов в батарее тестов.

Вот ещё пример теста на различение:

З: *Что является предметом педагогической науки:*

а) процесс воспитания подрастающих поколений;

б) процессы в сознании человека при обучении;

в) система взаимодействия учителя и ученика;

г) педагогическая система;

д) образовательная система.

Э. (бумажный):

а) — ; б) — ; в) — ; г) + ; д) —.

$p = 5$.

Э. (машинный): г) + .

$p = 1$.

При создании тестов различения также возможны ошибки, которые ведут к конструированию псевдо-тестов первого уровня. Например:

Каков коэффициент усвоения знаний учеником, если в батарее тестов с 22 существенными операциями он допустил шесть ошибок?

а) 3/11; б) 11/3; в) 8/11; г) 11/8.

Объясните, усвоил ли ученик учебный предмет?

Этот тест, кроме своей функциональной «инвалидности» первому уровню усвоения, содержит две задачи второго уровня. Псевдо-тестами первого уровня вперемешку с функционально валидными тестами первого уровня пе- стрят все, без исключения, тестовые сборники на сегодняшнем тестовом рынке. Всё бы ничего, но никаких корректных суждений о знаниях учащихся или о качестве процесса обучения по таким тестам вынести нельзя, а следовательно, нельзя и принимать правильные решения о направлениях совершенствования педагогических систем.

Тесты классификации (сопоставления)

Их смысл понятен из приведённого ниже примера теста сопоставления (классификации):

3. Укажите соответствие понятий и их обозначений:

- | | |
|----------------------|-------------|
| 1) Уровень усвоения, | а) N, |
| 2) Число УЭ | б) ψ , |
| 3) Степень освоения | в) а, |
| 4) Осознанность | г) г. |

Э. 1 — в; 2 — а; 3 — г; 4 — б.

p = 4.

Деятельность ученика в тестах классификации является многократным узнаванием, а число существенных операций равно числу соответствий, которые надо распознать. Более кропотливый анализ деятельности ученика покажет, что на самом деле он выполняет намного больше сканирований задания теста, чем мы считаем в нём операций. Последнее, однако, много проще и определённое для суждения, и мы принимаем число существенных операций равным числу соответствий.

Некоторые авторы тестов классификации добавляют в одну из колонок избыточную альтернативу, которую не с чем сопоставлять. Эта альтернатива ничего не меняет в структуре деятельности ученика, но иногда вызывает его недоумение и искусственно провоцирует на ошибку. По-видимому, подобное стремление запутать учащегося — это следствие недостаточной педагогической этики составителя тестов.

Тесты второго уровня мастерства

В соответствии с определением второго уровня усвоения деятельность испытуемого должна состоять в репродуцировании по памяти ранее усвоенной информации и алгоритмов деятельности при решении типовых задач. Наиболее важное правило, которого создатель тестов второго уровня должен строго придерживаться: чтобы задачи, включаемые в тест, могли быть решены путём **прямого** и **буквального** применения ранее усвоенных алгоритмов решения, без какой бы то ни было трансформации задач или алгоритмов.

Различают три конструкции тестов второго уровня:

- *тесты-подстановки*,
- *конструктивные тесты*,
- *тесты — типовые задачи*.

Вот примеры таких тестов:

Тесты-подстановки

Задания в таких тестах содержат пропуски текста или символов в формулах, которые испытуемый должен восполнить. Например:

1) 3. Назовите основные элементы ПС:

- 1) _____; 2) _____; 3) _____; 4) _____;
5) _____; 6) _____.

Э. 1) Студенты; 2) Цели обучения; 3) Содержание; 4) Процесс обучения; 5) Орг. Формы; 6) Учитель/Учебник.

p = 6.

Здесь требуется буквальное перечисление элементов ПС, на число которых *намекает* форма теста. Намёк это не ответ — подсказка, как в тестах первого уровня.

2) 3. Дополните формулу: $K\alpha = /$.

Э. n / p .

p = 2.

Если: *Дополните формулу: $K\alpha =$* , то **p = 3**, т.к. надо ещё воспроизвести знак деления.

3) 3. Приведите названия уровней усвоения:

- 1) _____; 2) _____; 3) _____; 4) _____.

Э. 1. Ученический; 2. Исполнительский; 3. Экспертный; 4. Творческий.

p = 4.

Однако следующий тест не является тестом подстановки второго уровня:

4) 3. Образовательный процесс является _____ педагогической науки.

а) объектом; б) предметом.

Это тест различения первого уровня и, следовательно, псевдо-тест подстановки второго уровня.

В тестах подстановки даётся только *намёк* на форму и состав требуемого ответа, а в тестах первого уровня различения даётся *подсказка* в виде полного и правильного ответа. Намёки подобного рода не являются подсказками, а поэтому не снижают уровня необходимой деятельности для разрешения теста.

Конструктивные тесты

Задания конструктивных тестов — это хорошо известные вопросы по пройденному материалу, требующие от ученика его воспроизведения по памяти. В этих тестах число существенных операций также определяется по эталону, содержащему *полный и правильный ответ* на вопрос, поставленный в задании теста. Например:

1) **3.** Если в тесте содержится «р» существенных операций и ученик выполнил «п» из них правильно, то как записать коэффициент усвоения?

Э. $K_b = п/р$.

р = 2 (1. Кб; 2. п/р). Но можно также считать **р = 5** (1. Кб; 2 п; 3. /; 4. р; 5. п/р).

2) **3.** Назовите уровни усвоения.

Э: *ученический; исполнительский; экспертный; творческий.*

р = 4.

В этих тестах, в отличие от тестов — подстановок с теми же вопросами, не содержится намека на форму и состав ответа.

Тесты «Типовые задачи»

Что мы понимаем под термином «Типовая задача»? В определении понятия второго уровня усвоения под типовой задачей понималось такое задание, когда усвоенные методы деятельности (алгоритмы) могут быть применены к его разрешению без какой бы то ни было трансформации как условия задачи, так и алгоритма деятельности. К примеру, ниже приведены две, как будто одинаковые задачи, но относятся они к разным уровням усвоения:

31: *Подсчитайте коэффициент усвоения, если общее число существенных операций в батарее тестов равно 22, а число правильно выполненных студентом операций равно 16.*

32: *Подсчитайте коэффициент усвоения, если общее число существенных операций в батарее тестов равно 22, а число допущенных студентом ошибок равно 6.*

В этих примерах выполнение 31 требует деятельности на втором уровне, а 32 — на третьем, т.к. требует трансформации исходной задачи на две подзадачи: сначала надо найти число правильно выполненных студентом существенных операций, а затем уже подсчитать коэффициент усвоения.

А к какому уровню деятельности отнести такой тест?

3. *Сколько времени Тст надо дать учащемуся на выполнение батареи тестов, если*

$K_t = 0,5$, а $T_{пр} = 10$ мин?

Э. $T_{ст} = T_{пр} / K_t = 10 \text{ мин} : 0,5 = 20 \text{ мин}.$

р = 1.

В этой задаче надо преобразовать формулу расчёта K_t ($K_t = T_{пр}/T_{ст}$) относительно

$T_{ст}$ ($T_{ст} = T_{пр} / K_t$). Это математическое преобразование не меняет ни педагогической задачи, ни педагогического алгоритма, а поэтому и не меняет вида задачи, она остаётся задачей второго уровня. Математические же операции не считаются существенными операциями этого теста.

В тестах по математике все математические действия являются существенными операциями теста даже если тест по геометрии, а операция арифметическая.

Типовые задачи называют ещё алгоритмическими задачами, чтобы подчеркнуть существование чёткого и однозначного правила (формулы), с помощью которой задача всегда решается. Не следует забывать о требовании простоты теста: в задании должна содержаться лишь одна задача данного уровня. К примеру, если в задании в приведённом выше примере теста второго уровня добавить, казалось бы, простое и естественное продолжение: «Сделайте вывод об усвоении учащимся предмета» или «Оцените знания учащегося по 12-балльной шкале», то будет нарушено требование простоты теста.

Различение задач второго и третьего уровней усвоения имеет существенное значение для правильной оценки мастерства учащегося в данной области деятельности. Поскольку такие задачи реально существуют, умение их решать надо оценивать отдельно. Для этого и различают второй и третий уровни деятельности и, соответственно, мастерства.

Тесты третьего уровня мастерства.

В соответствии с определением третьего уровня усвоения испытуемый должен выполнить нестандартную (*эвристическую*) деятельность, когда не существует заранее заданного алгоритма решения задачи. Ситуация, предлагаемая тестом третьего уровня, состоит в том, чтобы, преобразовав некоторым способом её условия, подвести её под известные правила действия, которые могут привести к требуемому решению. Задачи третьего уровня (так же как и других уровней усвоения) не обязательно требуют математических решений, т.е. они не обязательно сформулированы или поддаются формулировке на количественной ступени абстракции. Они могут быть и феноменологическими, как это сплошь и рядом происходит с задачами педагогической деятельности. Поэтому в любом учебном предмете попытки применить его содержание к реальной деятельности людей порождают задачу третьего уровня деятельности.

К примеру, следующая педагогическая задача является заданием третьего уровня: «Создайте тесты первого и второго уровней деятельности в вашем предмете на одном и том же учебном элементе и на качественной ступени абстракции». Почему это задание валидно третьему уровню деятельности? Потому, что нет готового алгоритма, как эту деятельность выполнить с места и безошибочно. Испытуемому предстоит составить план её выполнения, выбрать учебные элементы, сделать (или найти) их описание на качественной ступени абстракции и лишь затем создать тесты, для чего пока также нет однозначных алгоритмов.

Возникает естественный вопрос: если для решения задач третьего уровня нет готовых алгоритмов — однозначных правил деятельности, гарантированно приводящих к успеху, то как к таким тестам создавать эталоны? Действительно, готовых алгоритмов для решения всей задачи третьего уровня нет, но содержательно и функционально валидная задача третьего уровня всегда может быть сведена к нескольким последовательным задачам второго уровня, для которых есть готовые алгоритмы. Так, выше сформулированную педагогическую задачу третьего уровня можно представить следующими задачами второго уровня:

- 1) выделить некоторый учебный элемент из данного предмета;
- 2) сформулировать учебный текст по данному УЭ на качественной ступени абстракции;
- 3) создать тест опознания или различения на данный учебный элемент;
- 4) создать тест-подстановку или конструктивный тест.

Итого, четыре существенные операции. Действия внутри этих операций (построение графа для выделения УЭ, безошибочность формулирования учебного текста) — это несущественные операции для этого теста третьего уровня.

В каждой области деятельности есть эксперты; первый и главный из них — сам учитель. Каждая задача третьего уровня и её эталон проходит экспертизу признанных экспертов (профессионалов), которые должны ответить на вопросы:

- а) содержательно ли валиден тест; знаком ли ученик, решающий задачку, с различными методиками, которые могут быть использованы в тесте;
- б) функционально ли валиден тест, т.е. требует ли задание теста выполнения деятельности тре-

тьего уровня, т.е. эвристической деятельности (переформулировки задачи, разработки плана решения, подбора алгоритма);

в) прост ли тест, т.е. состоит ли задание теста из одной-единственной задачи данного уровня.

Последнее требование в тестах третьего уровня может вызвать некоторые трудности и недоумения, поскольку задача третьего уровня решается при сведении её к нескольким задачам второго или даже первого уровня, для которых известен разрешающий алгоритм. Эти составляющие задачу третьего уровня промежуточные задачи являются её существенными операциями, тогда как операции решения промежуточных задач в оценке деятельности ученика могут не учитываться. При диагностике третьего уровня усвоения даже не обязательно требовать детального решения задачи, можно ограничиться планом решения. Это допущение понятно, т.к. проверка третьего уровня усвоения предполагает, что предыдущие уровни учеником освоены. Такое допущение может существенно сэкономить время тестирования без ущерба для точности суждений.

Существует множество задач с ситуациями, сформулированными на количественной ступени абстракции. С большинством из них учащиеся знакомятся при изучении математики, физики или химии. Другие же предметы чаще всего ограничиваются тестами второго уровня на качественной ступени абстракции, хотя и на этой ступени существуют неограниченные возможности для формулирования тестов третьего уровня. В таких тестах могут содержаться задачи, требующие для своего решения относительно развитого логического анализа (из-за неочевидности решения или постановки эксперимента для получения искомого ответа). Вот пример такого теста:

3. При каких обстоятельствах две надёжные батареи тестов по одному и тому же предмету, выполненные одним и тем же испытуемым, покажут разные качества усвоения?

Э. 1) если предположить, что учебный предмет состоит из N учебных элементов и

2) что батареи выполнены: первая преимущественно на УЭ первой половины предмета, а вторая — на УЭ второй половины предмета и

3) что ученик усвоил только первую половину предмета; но большая часть тестов в первой батарее относится ко второй половине предмета, а во второй батарее — к первой половине предмета, то

4) оба теста покажут разное качество усвоения учеником предмета: худшее по первому тесту и лучшее по второму тесту.

p = 4.

В предметах, излагаемых на количественной ступени абстракции, тесты третьего уровня требуют многоходовых вычислений или искусственных преобразований. Например, тест для изучающих экономику:

З. Если вы сдали свой товар перекупщику за 40% его первоначальной цены, а перекупщик поставил его на продажу, прибавив 20% от суммы, которую он вам заплатил, то сколько процентов составляет цена, назначенная перекупщиком от исходной цены товара?

Вот как может выглядеть эталон к этой задаче в предположении, что ученик уже усвоил методику расчёта простых процентов:

Э. 1) Обозначить первоначальную цену товара буквой *A*;

2) Найти сумму, за которую перекупщик купил товар (40% *A*);

3) Найти сумму, которую перекупщик прибавил к купленному им товару (20% (40%*A*) = 8%*A*);

4) Найти сумму за которую перекупщик продаёт товар (40%*A* + 8%*A* = 48% *A*);

Следовательно: новая цена товара составляет 48% от его первоначальной цены.

p = 4.

Тесты, в которых требуются искусственные преобразования, относятся, чаще всего, к предметам, базирующимся на математике:

Например: **З.** В приведённой ниже фигуре (Рис. 11) *CD* = 4; *ED* = 5; и *ABCE* — квадрат. Чему равен периметр треугольника *BED*?

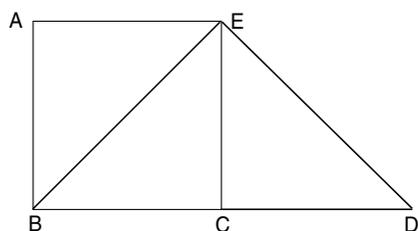


Рис. 3

Здесь эталон — это план решения, исключая сами вычисления. Если вычисления не учитывать в задачах третьего уровня, то лучше задавать их в общем виде, без цифр.

Э. 1) Применить теорему Пифагора для нахождения *EC*;

2) Сделать то же для нахождения *BE*;

3) *BC* = *EC*;

4) Найти периметр треугольника *BED*: *ED* + *DC* + *BC* + *BE*.

p = 4.

Для тестов третьего уровня характерны применение комбинации различных действий и алгоритмов. Так, в последней задаче дважды используется теорема Пифагора, извлекается квадратный корень, суммируются смешанные числа. При решении задач третьего уровня ученик познаёт новые возможности известных ранее алгоритмов и таким образом обогащает свой опыт. В противоположность этому, при решении задач на первом и втором уровнях лишь автоматизируются усвоенные умения.

Тесты четвёртого уровня мастерства

По определению, четвёртый уровень усвоения деятельности — это уровень творчества, который является венцом всякого обучения. Не каждому человеку, жившему и живущему на планете Земля, удалось достичь такого уровня мастерства в какой-либо области деятельности. Те же, кому это удалось, хорошо известны, одни лишь в кругу своих коллег, другие — всему миру. В результате творческой деятельности порождается объективно новое знание, т.е. знание, обогащающее науку и общечеловеческий опыт. Возникает, казалось бы, непреодолимая проблема диагностики достижения учеником четвёртого уровня мастерства, поскольку то знание и умение, усвоение которого предстоит диагностировать, ещё не существует и ученик в процессе тестирования должен его открыть. Возникает замкнутый круг: знания ещё нет, но на его основе должен быть построен тест. Выходом из положения становится искусственная процедура построения теста четвёртого уровня, пригодного для контроля качества обучения. В качестве заданий для тестов четвёртого уровня можно использовать уже исследованные ранее проблемы, выполненные и запатентованные изобретения, защищённые диссертации, принёсшие объективно новую информацию в данную отрасль науки.

Почему можно использовать эти источники?

Во-первых, потому, что всякое новое знание не возникает на пустом месте, наука или техника в определённое время и при определённых условиях бывает «беременна» этим новым знанием, и достаточно одарённый человек, хорошо подготовленный в данной области деятельности (на α_3) однажды усматривает возможное новое продолжение этой деятельности. Теории относительности не было бы и мы не знали бы Альберта Эйнштейна, если бы до него не существовали бы Бор, Максвелл и многие другие творцы всей предшествующей Эйнштейну физики.

Если кому-то до испытуемого удалось сделать открытие или изобретение, то и испытуемый, претендующий на звание творца, должен суметь переоткрыть открытое.

Во-вторых, уже выполненные исследования четвёртого уровня можно использовать в учебно-диагностических целях, если они не вошли ещё в учебники и хрестоматии по предмету. В противном случае это будут тесты от силы второго уровня.

Эти источники для постановки задач в тестах четвёртого уровня могут использоваться в их оригинальной форме для диагностики достижений студентов вузов. Для диагностики четвёртого уровня усвоения в общеобразовательной школе названные источники проблем и решений необходимо адаптировать в соответствии требованиям содержательной валидности теста.

От испытуемого следует требовать схемы или плана возможного решения проблемы или эскиза изобретения. В эталоне схематично отображается уже выполненное предшествующими исследователями решение проблемы и подсчитывается число существенных операций теста — «р». Ниже показан тест четвёртого уровня по педагогике, материал для которого заимствован из кандидатской диссертации Д. Гельфанда, защищённой на факультете психологии МГУ.

3. Разработайте критерии для диагностики моральных качеств личности.

- Э. 1) Главным показателем моральных качеств личности является его поведение в обществе;
- 2) оно может быть приемлемым (положительным) или неприемлемым (отрицательным);
- 3) для суждения об усвоенных навыках поведения требуется наблюдение за публичным поведением испытуемого в различных социальных ситуациях;

в итоге наблюдения можно обнаружить следующие формы поведения:

- 4) положительное, под непосредственным наблюдением;
- 5) положительное вне наблюдения, из-за опасения наказания;
- 6) положительное по внутреннему убеждению.
- 7) эти формы поведения могут быть приняты за **три уровня моральной воспитанности личности.**

p = 7.

Вот ещё один тест четвёртого уровня, построенный на основе наших исследований (см. Лекцию № 5):

2) **3. Разработайте план эксперимента по определению скорости усвоения учеником учебного предмета с заданными параметрами.**

- Э. 1) выделить законченный фрагмент учебного предмета;
- 2) сосчитать число УЭ (N) в этом фрагменте;
- 3) определить цель (α, β) изучения предмета;
- 4) провести обучение группы студентов до достижения ими усвоения на заданном уровне ($K\tau > 0,7$);
- 5) зафиксировать время, затраченное на учебный процесс;
- 6) создать формулу для расчёта скорости (C) усвоения;
- 7) рассчитать скорость усвоения СУЭ/мин.

p = 7.

Итак, выше показана методика тестирования качества усвоения знаний и действий учащимися, которая опирается на объективную теорию процесса восхождения человека от незнания к знанию. На этой основе чётко определены четыре уровня усвоения и построена методика создания тестов разного уровня. Эта методика покрывает весь диапазон возможного качества усвоения информации человеком: от первоначального ознакомления с новой информацией (уровень «знакомства») до продуцирования новой информации (уровень «творчества»).

Для контроля качества усвоения знаний учащихся используются батареи тестов, состоящие из тестов одного и того же уровня определённой надёжности. О них речь пойдёт в следующей лекции. □