

# КАК ИЗМЕРИТЬ КОМПЕТЕНЦИИ?

**Александр Владимирович Могилев,**

*профессор Воронежского государственного педагогического университета,  
доктор педагогических наук*

**Наталья Михайловна Ткачева,**

*аспирантка кафедры новых информационных технологий и средств обучения  
Воронежского государственного педагогического университета*

- *тестовые задания* • *образовательная компетенция* • *компетентностный подход* • *знаниевая педагогика* • *модель образовательных компетенций*
- *таксономия Б. Блума*

## **Качество образования и тесты**

Ответ на вопрос, насколько качественное образование получают выпускники, можно искать по-разному. Один из подходов, что называется, «от печки», состоит в том, что предварительно можно попытаться определить понятие качества как такового и далее качества образования. Этот подход может далеко завести в рассуждения о том, какие требования предъявляет к образованию общество и государство, зачем образование нужно членам общества, какой подготовки требует жизнь и успешная деятельность в каких-то конкретных условиях. Он требует разработки понятийного аппарата качества образования и моделей образования, связанных с футурологическими прогнозами о развитии тех или иных аспектов общества. В итоге на этом пути мы тонем в дефинициях, так и не добравшись до самих процедур измерения качества образования и, главное, результатов этих измерений.

В педагогической среде есть и другой, совершенно противоположный, упрощённый подход к проблеме качества образования как чисто инструментальной задаче: берём и измеряем, особенно не вдаваясь в суть измеряемого понятия. Действительно, в рамках физики и техники, мы, например, так и не знаем досконально, что такое расстояние между телами или размер тела, хотя прекрасно умеем их измерять путём сравнения с эталоном. Для измерения просто берётся эталон длины и прикладывается к расстоянию или размеру столько раз, сколько нужно — и мы получаем это расстояние или размер: длину, ширину, высоту.

Такой инструментальный подход в измерениях качества образования известен уже очень давно (так, ещё в 1864 г. появились шкалированные книги англичанина Джоржа Фамера и были опубликованы таблицы по проверке орфографических знаний учащихся американца Дж. М. Райса). Чтобы оставаться до конца корректными и не претендовать на оценку таких не до конца определённых и понятных явлений, как качество образования и, тем более, вторгаться в оценку параметров личностного развития, американские тестологи

использовали в своей работе такие узкие и точные термины, как «оценка учебных успехов» учащихся.

Как бы то ни было, но в рамках этого упрощённого подхода легко можно составить для учащихся набор заданий, контрольную работу и по результатам её выполнения высказывать суждения о качестве образования выполнивших её учащихся.

Однако при проведении такой контрольной для большинства учащихся возникает проблема объективности и стабильности её проверки и оценивания результатов. Разные учителя могут по-разному оценивать сходные ошибки и недостатки или наоборот одинаково принципиально разные погрешности. Даже один учитель может по-разному оценить одну и ту же работу в зависимости от своего настроения, усталости, состояния здоровья или личности выполнившего её ученика. Как сделать проверку работ учащихся и оценивание менее трудоёмкими и, по возможности, обеспечить их объективность и стабильность?

Так появляются задания в стандартизированной форме: ответ выражается в выборе одного из альтернативных вариантов или указании набора значений, сумма которых является правильным ответом, указании порядка и некоторых других — настоящие тестовые задания в их современной форме. Такие тестовые задания допускают полностью формализованную проверку с помощью ключа, не требуя участия специалиста, и обеспечивают диагностику структуры и уровня подготовленности тестируемого.

Творческая сторона педагогической деятельности при тестовом оценивании переносится с проверки решений и ответов на процесс составления заданий. Именно составляя задания теста, предусматривая определённые реакции — действия тестируемого учащегося в ответ на полученные задания, педагог фактически формулирует некие предположения о том, как качество подготовки учащегося выразится в ответах теста.

Однако, не всё так просто. С появлением тестов как инструмента педагогических измерений возникает вопрос о качестве тестов, их адек-

ватности тем требованиям и ожиданиям, которые к ним имеются, а также о методах и приёмах их разработки.

### **Традиционные методики составления тестов и требования к формам заданий**

Итак, тестовые задания по своей форме существенно отличаются от традиционных вопросов и заданий. Необходимо понимать: эти отличия по форме возникают в результате стремления добиться отличий и стандартизации по содержанию. Так, взамен традиционных заданий, предполагающих ответ или решение в открытой форме (когда тестируемый формулирует ответ в виде слова, фразы или предложения, числа или выражения), большая часть тестовых заданий рассчитана на ответы в закрытой форме, т.е. выборе из предложенных вариантов. Открытая форма ответа слишком трудна для оценивания ответа по существу, кроме того, сами вопросы и эталоны ответа могут отражать субъективные пристрастия и мнения составителя тестов, не учитывать наличия эквивалентных, синонимичных и родственных ответов, влияние контекста того учебного курса или даже конкретного занятия, для которого разрабатывается вопрос. Однако такой вопрос может быть преобразован в тестовой форме, которая устранил неопределённость, связанную с субъективностью составителя, влиянию контекста конкретного курса или занятия.

Открытый традиционный вопрос может быть таким: «Каким животным является корова?» Ясно, что, не зная предпочтений учителя или контекста конкретного занятия, к которому относится задание, ответить корректно невозможно. Здесь подойдут ответы «крупным», «рогатым», «домашним», «жвачным» и т.д. Автор же задания подразумевал ответ «травоядным». Таким образом, многие открытые вопросы и задания некорректны

и непригодны для оценки инвариантных знаний, не привязанных к конкретному курсу или даже уроку.

Преобразование этого вопроса — выберите правильный вариант ответа: всеядное, травоядное, плотоядное».

Как видим, тестовое задание состоит из: 1) содержательного блока, сформулированного в виде вопроса или утверждения («Каким животным является корова?»); 2) блока инструкции («Выберите правильный вариант ответа»); блока вариантов ответов («всеядное, травоядное, плотоядное»).

В блоке вариантов ответов правильным должен быть только один вариант («травоядное»), иначе тестовое задание станет некорректным. Неверные варианты ответов иногда называют дистракторами (отвлекающими вариантами, похожими на правильный, но не являющимися таковыми). Важно, чтобы варианты ответов были однородными, были бы оформлены сходно, относились бы к одной классификации и контексту. Лучше выделить содержательный блок прописными буквами или жирным шрифтом, а блок инструкций — строчным. Некорректно, например тестовое задание про корову с такими вариантами ответа: «жвачное, травоядное, млекопитающее», так как каждый является верным в своём контексте.

Если все задания однотипны, блок инструкций можно вынести из задания, сделав его общим для всего теста. Это повышает удобочитаемость теста и делает его более простым и успешным для тестируемых.

Нельзя считать вполне корректным и такой подход: тестовые задания формально усложняют за счёт дополнительных уровней операций, которые должен выполнить тестируемый, например: «Укажите номер ответа на вопрос, сколько делителей имеет число 16: 1) 3; 2) 2; 3) 4; 4) 1. Число 16 имеет 3 несобственных делителя, поэтому верен вариант 1), однако некоторые тестируемые укажут как верный вариант 3), поскольку

поддадутся на уловку составителя теста, старающегося формально усложнить тест и запутать тестируемого, заставить его допустить ошибку.

Общее требование к тестовым заданиям — предотвратить возможность простого угадывания правильного варианта ответа, однако это требование приводит к удлинению набора заданий и соответственно времени его выполнения. Дистракторы должны тщательно продумываться и быть правдоподобными. Неэффективны задания, в которых верные ответы могут быть получены путём отсеивания неправдоподобных дистракторов, как например в задании: «В каком году произошла Куликовская битва? Выберите правильный вариант ответа: 1) 1812; 2) 1941; 3) 1380.

Если говорить об угадывании, то в тестах с 2 вариантами ответа вероятность угадывания  $1/2$  в каждом задании, следовательно, каждое дополнительное задание всего лишь вдвое уменьшает влияние угадывания на результаты выполнения теста. В задании с 4 вариантами вероятность случайного угадывания  $1/4$ , т.е. в каждое дополнительное задание в 4 раза снижает влияние угадывания на результаты, но требует увеличения времени как в случае с заданиями с 2 вариантами ответа. Если увеличивать число вариантов ответов, то вместе с уменьшением вероятности угадывания ответа будет расти вероятность ошибки, связанная с объёмом восприятия информации человеком, который лежит между 5 и 9 вариантами (более 9 объектов человеку трудно удерживать в сознании одновременно). Поэтому тесты с более чем 6 вариантами ответа применяются редко, на результаты тестирования будут влиять психофизиологические возможности тестируемых, а не качество их подготовки. Оптимальным является 4 варианта ответа для тестового задания в закрытой форме с выбором одной из альтернатив, что и используется в заданиях типа А ЕГЭ.

Помимо рассмотренной формы закрытых тестовых заданий с выбором одной из альтернатив имеется ещё 5 видов заданий:

- с множественным выбором;
- на установление соответствия;
- на установление порядка;
- на классификацию;
- на указание геометрической области.

Однако различия между этими видами заданий непринципиальны, любое из них можно свести к заданию выбора одной из альтернатив или другому. Блок инструкции или содержательный блок при этом, как правило, получается громоздким, трудным для восприятия, поэтому тестируемые показывают хорошие результаты при тестировании с различными видами заданий при условии дополнительной тренировки в выполнении заданий этих видов на любом предметном материале.

Можно утверждать, что тестовые задания на выбор альтернативы лучше использовать при оценке знаний, в рамках знаниевого подхода. Использование других видов заданий в тесте становится оправданным при более сложном целеполагании, лежащем в основе теста, и оценке не только знаниевой, но и других компонент подготовки, которые относятся к компетентностному подходу (и когнитивной компетенции в частности).

### **Проблема подхода при интерпретации тестовых измерений**

До сих пор мы рассматривали форму тестовых заданий саму по себе, без связи с методикой интерпретации результатов тестирования. Подходов к интерпретации существует два: критериальный и нормативный.

При критериальном подходе составитель теста исходит из того, что тестовые задания отражают обязательные элементы знаний для всех учащихся. Это значит, что при интерпретации результатов тестирования в рамках этой логики следует учитывать число выполненных заданий теста или сумму баллов за тест и интерпретировать результаты теста в соответствии с критерием типа: от 75% до 100% баллов, возможных в тес-

те — это «отлично», от 50% до 75% — «хорошо», от 25% до 50% — это удовлетворительно, а ниже 25% — неудовлетворительно. Мы намеренно привели такие критерии оценивания, которые сразу же иллюстрируют всю условность процедуры критериального оценивания с помощью теста. Возможны вариации при установлении этих критериев, позволяющие выполнить «настройку» теста как инструмента педагогических измерений. С помощью критериев можно, например, учесть сложность тестовых заданий (при наличии в тесте небольшого числа заданий повышенной сложности диапазон сумм баллов, соответствующих отметке «отлично», должен уменьшаться), или получить заданные доли тестируемых, которые не выполнили тест (получили отметку «неудовлетворительно») или, наоборот, отлично справились с ним (и должны быть поощрены из некоторого фиксированного поощрительного фонда). В основе критериального подхода к интерпретации результатов теста лежит представление о неких инвариантных предметных знаниях, которые должны иметь все учащиеся, чтобы их можно было считать освоившими учебный курс. Критериальный подход весьма близок к знаниевой модели образования, с его использованием вполне естественно представляется трактовка теста, как инструмента, позволяющего определить (измерить) «уровень знаний» учащихся, а самого учащегося представить в виде сосуда (ёмкости), наполняемого знаниями в процессе обучения. Очевидно, что при критериальной логике тестирования и его участники и организаторы стремятся получить максимально высокие результаты, поэтому имеется тенденция к снижению трудности и сложности заданий и пороговых критериев. Наиболее соответствует такому подходу использование тестов для оценки остаточных знаний, а заданий в тесте — на выбор ответа из многих альтернатив.

Противоположный подход к интерпретации результатов теста — нормативный. Цель теста при его нормативной интерпретации — дифференцировать тестируемых, как можно более уверенно отделить тех из них, кто имеет лучшую подготовку, от тех, кто подготовлен средне или плохо. Нормативный подход означает, что в результате выполнения теста на некоей контрольной выборке тестируемых, устанавливаются сложность и трудность тестовых заданий, тест нормируется. Затем нормированный тест используется для дифференциации и получения рейтинга тестируемого, а столь привычные нам 4-балльные шкалы — «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно» с результатом теста могут и не сопоставляться. Нормативному подходу соответствует представление о тестируемом как творческой личности, действующей с некой долей неопределённости, случайности, и в ходе теста проявляющей не столько знания, сколько близкие к компетенциям комплексы, позволяющие тестируемому выполнять задания различного вида и сложности. Задания теста должны быть более сложными, среди них естественно смотрелись бы задания разных видов, не только с выбором ответа из нескольких альтернатив.

При критериальном подходе длина теста определяется лишь степенью охвата тестом учебного материала, освоенной программы. Важно требование равномерного охвата тестом дидактических единиц, освоения которых проверяется. При нормативном подходе длина теста устанавливается требованием минимального числа заданий, необходимых для дифференцирования тестируемых, не требуется равномерный охват тестов освоенной программы, и от стандартизированных дидактических единиц также можно отойти. Нормативный подход в интерпретации результатов теста в большей степени отвечает компетентному подходу.

Если попытаться в свете вышесказанного как-то проанализировать технологии, сло-

жившиеся при проведении тестирования ЕГЭ, то становится ясно, что они являются эклектичным смешением этих совершенно различных подходов — и критериального, и нормативного. ЕГЭ имеет и черты знание-ориентированности, и в то же время может претендовать на элементы компетентного подхода. Вероятно, именно из-за этого смешения черт и подходов ЕГЭ как технология педагогических измерений вызывает такие большие претензии и критику.

Отметим, что надёжность и валидность тестов в ходе тестирования имеет два аспекта. Традиционный — качество заданий теста, о чём мы уже говорили выше. Наличие в тексте некорректностей приводит к снижению результатов, которые показывают тестируемые, и это снижение можно исключить, проведя редактирование теста, исправив имеющиеся в нём недостатки. Второй аспект надёжности и валидности теста — натренированность на выполнение тестовых заданий, вработанность тестируемых в ситуацию тестирования. Очевидно, что при отсутствии необходимой тренировки на прохождение тестов (неважно какого содержания) результаты тестов оказываются заниженными и отражают волнение участников, а также их недостаточное владение приёмами работы с тестами, восприятия их инструктивных и содержательных блоков, оформления результатов тестов. Тренировка позволяет снизить влияние этих несодержательных факторов на результаты тестирования и перейти к оценке собственно подготовки, качества образования. Глубинный смысл тестового контроля заключается не в одноразовых процедурах тестирования и получении индивидуальной оценки учащегося, а в создании целостной системы обучения, развития учащихся и контроля этих процессов, мониторинга и анализа качества образования, основанных на объективных методах педагогических измерений.

### Тесты компетенций и таксономия педагогических целей Б. Блума

В июне 2012 года на основании концепции модернизации российского образования и национальной образовательной инициативы произошёл переход к новым федеральным образовательным стандартам третьего поколения, которые подразумевают вместо простого освоения содержания обучения по программе и передачи знаний, умений и навыков от учителя к ученику развитие способности учащегося самостоятельно ставить учебные цели, проектировать пути их реализации, контролировать и оценивать свои достижения, работать с разными источниками информации, оценивать их и на этой основе формулировать собственное мнение, суждение, давать оценку.

Одно из важных условий формирования ключевых компетенций учащихся — сформированный аппарат педагогических измерений, набор методов измерения качества образования, в том числе к тестированию, адаптированный к компетентностному подходу.

Перед педагогом стоит проблема использования таких компетентностных моделей тестирования, которые позволят диагностировать и выявлять степень соответствия результатов обучения современным российским и международным требованиям.

Но, к сожалению, при хорошей общетеоретической разработанности положений тестирования сейчас недостаточно конкретны методик разработки тестовых заданий. Разработчикам тестов предлагаются ценные, но слишком общие указания. Проблема отбора содержания тестовых заданий остаётся в тени, что связано как с новизной проблематики, так и со сложностью анализа содержания образования ввиду отсутствия методик и технологий, гарантирующих соответствие проектируемого содержания тестов образовательному стандарту.

В предлагаемой методике используется системный подход к проектированию тестовых заданий, то есть подбор такого содержания заданий, который отвечал бы требованиям системности знаний, их связям с компетенциями. На наш взгляд, задания в рамках системного подхода должны диагностировать наличие вза-

имосвязей в знаниях тестируемого, их сложную многоуровневую структуру. Тест должен решать задачу оценки не только уровня, но и структуры знаний.

Тестовые задания должны проверять не только знание элементарных дидактических единиц (ЭДЕ), но и, прежде всего, понимание взаимоотношений между ними, а также свойств, приобретаемых ЭДЕ только в составе данной системы знаний, умение применять их для решения практико-ориентированных задач, использовать для анализа существующих и синтеза новых систем в рамках предметных областей, проводить критическую оценку и отбор информации. Данные обстоятельства наводят на мысль о том, что методика разработки системных тестовых заданий должна учитывать таксономию педагогических целей по Б. Блуму (знание-понимание-применение-анализ-синтез-критическая оценка).

Таксономия Блума оказывается весьма продуктивным инструментом при анализе наборов компетенций и конструировании тестовых заданий. Так, оказывается, что требования к знаниям относятся к нижним ступеням пирамиды Блума (знание-понимание-применение), а требования к компетенциям — к верхним ступеням (анализ-синтез-критическая оценка). На рисунке приведено соответствие компетенций, приведённых во ФГОС нового поколения, и определённым ступеням пирамиды Блума).

Для построения системных тестовых заданий необходимо выделить не только связи, существующие между понятиями в предметной области (они выражаются, во-первых, законами, теоремами, правилами и принципами рассматриваемой предметной области). Эти связи могут отражать более широкие отношения: функциональной зависимости, симметрии, включения или исключения, логического или хронологического следования, генетического наследования,

причинно-следственной связи, классификации и т. д. Кроме того, они могут иметь характер не только специфический для предметной области, но и более широкий, метапредметный. Действительно, по одной из аксиом теории систем, каждая система входит в качестве подсистемы в систему более высокого ранга, и потому её понимание требует знания и понимания взаимосвязей данной системы с другими системами.

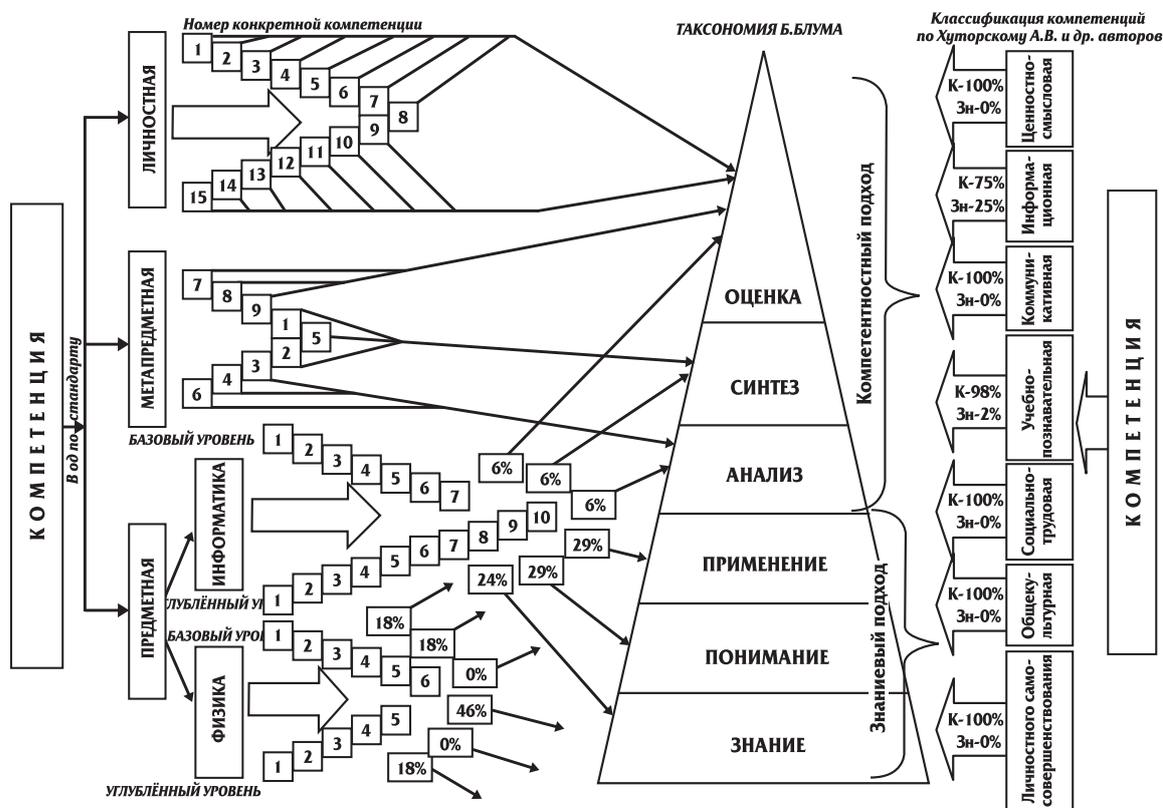
В соответствии с предлагаемой методикой, на первом этапе выделяются наиболее фундаментальные понятия данной предметной области. Фундаментальность определяется экспертом. Эти понятия рассматриваются как вершины графа, ребрами которого служат взаимосвязи между понятиями.

По мере построения графа связи понятий обычно выявляется необходимость в дополнительных понятиях, значимость которых на первых этапах была недооценена по тем или

иным причинам. У графа могут быть «висячие» ребра, связывающие понятия данной предметной области с понятиями смежных областей. Простейший вариант таких графов — деревья, которыми принято описывать состав систем в терминах отношений «часть — целое».

Основные принципы методики составления тестовых заданий, отражающие как формальные свойства «дерева знаний» предметной области, так и эффективные приёмы его конструирования:

- граф легче строить от вершин, а не от рёбер;
- граф должен иметь центр, состоящий из одной или двух смежных вершин;
- все вершины графа разбиваются на непересекающиеся яруса;
- вершины каждого яруса должны быть равноудалены от центра графа;
- наиболее значимые понятия располагаются в вершинах с наибольшим весом.



На втором этапе понятия ранжируются по весу: наибольший вес присваивается тем, из которых можно логически вывести остальные. Расположив таким образом фундаментальные понятия и приняв во внимание известные закономерности предметной области, можно определить «остов» дерева понятий.

На третьем этапе вершины графа соединяются ребрами, которые представляют собой взаимосвязи — отношение, закон, теорема, тождество. В зависимости от того, вершины каких ярусов оно соединяет, ребру приписывается вес. Наибольший вес, естественно, у ребер, соединяющих вершины верхних ярусов.

На четвертом этапе для каждого ребра/вершины составляются тестовые задания, соответствующие понятиям и взаимосвязям графа. Максимальное число заданий создаётся для ребер (вершин) с наибольшим весом.

На последнем этапе самим тестовым заданиям приписывается вес, соответствующий весу представляемых ими ребер/вершин графа, и определяется успешность прохождения теста по сумме весов успешно выполненных заданий.

При составлении заданий важно использовать формы вопросов, вопросительные слова, отвечающие сущности уровней (ступеней) таксономии. Например, уровень знаний — это припо-

минание конкретной, фактологической информации, уровень понимания — установление эквивалентности формулировок, уровень применения — решение простых задач, использование правил для получения конкретного ответа. Ступени анализа отвечают заданиям, требующим выделить компоненты и их взаимосвязи в составе целого, в структуре систем («из чего состоит?», «входит в состав?»). Ступени синтеза соответствуют заданиям, в которых требуется сформулировать некие новые правила, причинно-следственные связи, провести прогнозы. Наконец, ступени критической оценки отвечают созданию и выбору критериев, поиск основания для классификации и её проведение, поиск аргументов и логическое доказательство утверждений. Разумеется, данные задания не должны быть типовыми, повтором из учебников и пособий.

Тестовые задания, разработанные на основе данной методики, отвечают компетентностному подходу и позволяют провести оценку результатов обучения на соответствие современным российским и международным требованиям, в частности уровня сформированности компетенций в рамках компетентностного подхода и стандартов нового поколения. **НО**