

Т е о р и я

Педагогический тест — основной метод педагогической диагностики

В.С. Аванесов,
testolog@mail.ru

Всякое определение
есть отрицание.
Спиноза

A totality does not consist
of things but of relationships.
L. Elmslev

Проблемная ситуация

На текущий момент все определения педагогического теста можно признать спорными. Даже такое очень общее определение, что тест — это *метод* педагогической и психологической диагностики. Потому что из него не ясно — какой это метод, какова его структура, каков состав?

Другое общее и неконкретное определение: *тест* — это стандартизованный метод. В литературе есть и много других примеров определения теста как «пробы», «проверки», «стандартизованной процедуры» «набора заданий, вопросов...» и т.п.

Путь к достижению единых и согласованных понятий долг и труден, особенно в нынешних условиях. Потому что нет научного общества в вопросах диагностической деятельности, нет общих взглядов, нет согласия

исследователей по ключевым вопросам, нет дискуссий и конференций по вопросам теории тестов и методики тестирования. Соответственно, нет и позитивной динамики в развитии тестовой культуры. Вместо этого есть неметрический ЕГЭ, и всё, что с ним связано. Это и есть причины удручающего состояния понятийного аппарата теории тестов.

Определение педагогического теста

Между тем, от правильного определения основного понятия зависит состояние соответствующей науки и практики тестирования, а также педагогической диагностики. Поэтому определение теста можно признать основной проблемой теории тестов и теории педдиагностики. Главная цель применения педагогических тестов — установить отношение порядка между испытуемыми по уровню проявляемых при тестировании знаний, умений навыков и др. И на этой основе определить место (или рейтинг) каждого по измеряемому признаку.

Для достижения этой цели можно создать бесчисленное количество тестов, и все они могут соответствовать некоторым требованиям.

В традиционном (неадаптивном) тесте все испытуемые отвечают на одни и те же задания, в течение одинакового времени, в одинаковых условиях. Ко всем испытуемым применяются одинаковые правила оценивания ответов.

В теории тестов выделяется три структурных понятия: задание в тестовой форме, тестовое задание и тест.

Основное определение

Тест определяется как система заданий равномерно возрастающей трудности, специфической формы;

это система заданий, которая позволяет качественно и эффективно измерить уровень и структуру подготовленности испытуемых.

Здесь дано определение гомогенного теста, т.е. теста, измеряющего только одно интересующее свойство личности. Это, казалось бы, естественное требование для гомогенного теста очень часто нарушается из-за стремления разработчиков увязать проверку знаний по отдельной дисциплине со стремлением проверить что-то ещё. Нарушение этого требования приводит к ухудшению качества измерения, к потере так называемого свойства одномерности теста или, короче, шкалы, если держать в уме ещё одно очень короткое определение: тест — это шкала.

Истолкование основных терминов определения теста

Система означает, что в тесте собраны такие задания, которые обладают системообразующими свойствами. Здесь, в первую очередь, надо выделить общую принадлежность заданий к одной и той же системе знаний, т.е. к одной учебной дисциплине, их связь и упорядоченность. Для итоговой аттестации выпускников школ нужно разрабатывать интегративные задания, содержание которых охватывает систему знаний. В педагогическом тесте задания располагаются по мере возрастания трудности — от самого лёгкого до самого трудного. Иначе говоря, главным формальным системообразующим признаком теста является различие заданий по степени их трудности. Слово «формально» появилось здесь для того, чтобы ограждать читателя от соблазна трактовать тест единственным образом — только как систему заданий возрастающей трудности, без учёта их содержания.

Специфическая форма тестовых заданий отличается тем, что зада-

ния теста представляют собой не вопросы и не задачи, а задания, сформулированные в форме высказываний, истинных или ложных, в зависимости от ответов. Традиционные вопросы, напротив, истинными или ложными не бывают, а ответы на них нередко настолько неопределённые и многословны, что для выявления их правильности требуются заметные, в суммарном исчислении, затраты интеллектуальной энергии преподавателей. В этом смысле традиционные вопросы и ответы нетехнологичны, и потому их лучше не включать в тест.

Определённое содержание означает использование в тесте только такого контрольного материала, который соответствует содержанию учебной дисциплины; остальное в педагогический тест не включается ни под каким предлогом.

Например, проверка уровня интеллектуального развития является предметом психологического тестирования. Содержание теста существует, сохраняется и передаётся в одной из четырёх основных форм заданий. Вне тестовых форм ни тест, ни его содержание не существуют.

Теоретически оправданным критерием упорядочения содержания так называемого гомогенного теста является согласованная концепция знаний, и критерий *трудности заданий*. Такую концепцию сейчас согласовывать в России негде и не с кем. Внеучебное содержание (например, проверка уровня интеллектуального развития) в педагогический тест не включается. Это предмет психологического измерения.

Возрастающую трудность заданий можно образно сравнить с разновысокими барьерами на беговой дорожке стадиона, где каждый следующий выше предыдущего. Пробежать дистанцию и успешно преодолеть все барьеры сможет только тот, кто лучше подготовлен.

Поскольку в педагогическом тесте задания упорядочиваются по принципу возрастающей трудности, одни испытуемые «заваливаются»

уже на самом лёгком, первом задании, другие — на последующих заданиях. Учащиеся среднего уровня подготовленности могут ответить правильно только на половину заданий теста и, наконец, только самые знающие в состоянии дать правильный ответ на задания самого высокого уровня трудности, расположенные в конце теста.

Трудность задания может определяться двояко: а) умозрительно, на основе предполагаемого числа и характера умственных операций, необходимых для успешного выполнения заданий, и б) после эмпирической апробации заданий, с подсчётом доли неправильных ответов. В классической теории тестов многие годы рассматривались только эмпирические показатели трудности. В новых вариантах психологических и педагогических теорий тестов больше внимание стало уделяться характеру умственной деятельности учащихся в процессе выполнения тестовых заданий различных форм [1].

Показатель трудности задания рассматривается как важный системно-, и одновременно, структурообразующий фактор теста. К этому можно добавить ещё один критерий — это критерий логической определённости тестового задания. Его можно сформулировать словами, близкими к формулировке Х. Карри: задание является определённым, если на него можно ответить утвердительно или отрицательно, и если существует эффективный процесс для нахождения такого ответа [2; с. 69].

Ответ на задание педагогического теста представляет собой краткое суждение, связанное по содержанию и по форме с содержанием задания. Каждому заданию ставятся в соответствие ответы правильные и неправильные. Критерии правильности заранее определяются авторами теста. Вероятность правильного ответа на любое задание зависит от соотношения уровня знаний испытуемого и уровня трудности зада-

ния. При наличии сопоставимых шкал, эту вероятность выражают значением от нуля до единицы.

Оценка ответов по степени их правильности проводится довольно редко, но при необходимости создаются задания с такими ответами, которые правильны в различной степени [3]. Инструкция для испытуемых в таких случаях может быть такой: «Вашему вниманию предлагаются задания, в которых могут быть один, два, три и большее число правильных ответов. Нажимайте на клавиши с номерами всех правильных ответов!»

Посредством тестирования чаще других признаков проверяются знания, умения, навыки и представления. С точки зрения педагогических измерений полезно ввести два основных показателя качества знаний — уровень и структура знаний. Они оцениваются посредством регистрации баллов как за знание, так и за незнание всех требуемых компонентов проверяемого материала. Для объективизации этого процесса все компоненты должны быть одинаковы. Одинаковыми являются и правила выставления оценок испытуемым. Эти условия открывают дорогу для объективного сравнения индивидуальных структур знания и не знания.

Уровень знаний выявляется при анализе ответов каждого ученика на все задания теста. Чем больше правильных ответов, тем выше индивидуальный тестовый балл испытуемых. Обычно этот тестовый балл ассоциируется с понятием «уровень знаний» и проходит процедуру уточнения на основе той или иной модели педагогического измерения. Один и тот же уровень знаний может быть получен за счёт ответов на различные задания. Например, в тесте из тридцати заданий ученик получил десять баллов. Эти баллы, скорее всего, получены за счёт правильных ответов на первые десять, сравнительно лёгких заданий. Присущую для такого случая последовательность единиц, а затем нулей

можно назвать правильным профилем знаний ученика.

Если тестовый балл ниже требуемого уровня (измеряемого критерия), то проявленные при этом знания, умения, навыки и представления указывают на *докритериальный* уровень подготовленности испытуемого. Этот уровень является самым распространённым и массовым. Для многих он оказывается вполне преодолимым по мере изучения ими наук и овладения мастерством. Однако некоторые, преимущественно в силу недостатка способностей и прилежания, так и остаются на этом уровне. В России на докритериальном уровне оказывается довольно много выпускников школ.

В истории науки и искусств известно немало случаев, когда имевшийся в соответствующие времена критериальный уровень требований к знаниям (или мастерству) мешал общественному признанию отдельных талантов. Поэтому признание нередко приходило только по мере изменения критериев оценки. Чаще всего после смерти авторов многих бессмертных произведений в сфере науки, музыки, живописи и др. Яркий пример такого рода — изменение критериев оценки творчества Ван Гога, творчества многих художников-импрессионистов, формалистов и др. Известны и такие, например, факты. Будущий лауреат Нобелевской премии Альберт Эйнштейн летом 1895 года не смог сдать экзамены в Цюрихский политехникум, а всемирно известный, впоследствии, пианист С. Рихтер дважды отчисляется из консерватории за общую неуспеваемость.

Структура знаний оценивается на основе последовательности правильных и неправильных ответов на задания возрастающей трудности. Формой представления индивидуальной структуры знания и незнания является *профиль* знаний испытуемого, представляемый последовательностью единиц и нулей, получаемых каждым студентом. Про-

филь знаний представляет собой упорядоченный набор оценок (вектор-строку) в матрице тестовых результатов. Если испытуемый отвечает правильно на первые, сравнительно лёгкие задания, можно говорить о правильной структуре знаний. Профиль называется правильным, если в строке баллов у испытуемого все нули следуют за всеми единицами.

Если же обнаруживается противоположная картина, (испытуемый правильно отвечает на трудные задания и неправильно — на лёгкие), то это противоречит логике теста и потому такая структура знаний может быть названа инвертированной. Она встречается редко, и чаще всего, из-за нарушения требования располагать задания по мере возрастающей трудности. При условии, что тест сделан правильно, каждый профиль свидетельствует о структуре знаний. Эту структуру можно назвать элементарной (поскольку есть ещё факторные структуры, которые выявляются с помощью методов факторного анализа).

Роль структуры знаний многократно подчеркивалась выдающимся педагогом А. Дистервегом, а также психологом Д. Брунером. Последний считает, что «изложение структуры знаний, овладение этой структурой, а не просто усвоение фактов и технических приёмов является центральным моментом» [4].

Каждое учебное заведение должно стремиться, в первую очередь, к формированию правильных индивидуальных структур знаний, в которых не было бы пробелов (разрывов в знаниях), и на этой основе повысить уровень подготовки.

Уровень знаний в значительной степени зависит от личных усилий и способностей учащихся, в то время как структура знаний заметно зависит от правильной организации учебного процесса, от индивидуализации обучения, от мастерства педа-

гога, от объективности контроля — в общем, от всего того, чего обычно не хватает.

Уровень знаний в значительной степени зависит от личных усилий и способностей учащихся. В то время как структура знаний зависит от правильной организации учебного процесса, от индивидуализации обучения, от мастерства педагога, от объективности контроля. В общем, от всего того, чего обычно у нас не хватает.

В тестовом задании внимание педагога привлекает, в первую очередь, содержание и форма. Содержание определяется как отображение фрагмента учебной дисциплины в тестовой форме, форма — как способ связи, упорядочения элементов задания. Содержание теста существует, сохраняется и передаётся в одной из четырёх основных форм заданий. Вне тестовых форм ни тест, ни его содержание не существуют.

Один из главных вопросов теории тестов — вопрос выбора наилучшего теста из практически неограниченного множества всех возможных тестов.

Такой выбор проводится на основе критериев эффективности и качества, исследованных автором этой статьи [5–6] и опубликованных в прежних номерах прежнего журнала «Педагогические измерения»¹.

Каждый тест может отличаться от других по числу заданий и другим характеристикам. С прагматической точки зрения полезнее делать тест, имеющий сравнительно меньшее число заданий, но обладающий достоинствами, присущим тестам с большим числом заданий. Вот почему в тест стараются отобрать достаточное количество заданий, которое позволяет сравнительно точно определить уровень и структуру подготовленности.

¹ С 2016 года автор перешёл работать в «Педагогическую диагностику», а «Педагогические измерения» теперь редактируются другим составом редколлегии.

Статистическая интерпретация результатов тестирования ведётся преимущественно с опорой на среднюю арифметическую, показатели вариации тестовых баллов и на так называемые процентные нормы, показывающие, — сколько процентов испытуемых имеют тестовый результат худший, чем у интересующего испытуемого. Такая интерпретация тестовых результатов называется нормативно-ориентированной.

Измерение уровня знаний

Знания приходится измерять опосредовано, через эмпирически фиксируемые проявления признаков (индикаторов) знания. Поэтому каждое задание теста желательнее рассматривать как *индикатор*, выявляющий какой-то один фрагмент знаний у тех испытуемых, у которых знания есть.

Сколько индикаторов (заданий) нужно иметь в одном тесте? Здравый смысл подсказывает, что судить о знаниях всего проверяемого материала по ответу только на одно задание довольно опрометчиво, хотя в каждой учебной дисциплине есть задания, правильные ответы на которые говорят о многом.

Тем не менее, устойчивые выводы лучше делать по результатам применения достаточного числа индикаторов, обычно от 20 до 40. Индикаторы в форме тестовых заданий, сведённые в одну систему заданий возрастающей трудности, образуют тест.

Тесты являются сейчас наиболее развитой в научном отношении частью методического арсенала педагогики, позволяющей адекватно скреплять теорию с эмпирией, в соответствии с такими известными стандартами качества результатов педагогического измерения, как критерии надёжности и валидности.

Литература

1. *Tatsuoka K.K.* Item construction and psychometric models appropriate for constructed response. Prinstone, N-J, 1993. — 56 pp; *Frederiksen N., Mislevy R.J., Bejar I.J.* (Eds). Test theory for a new generations of tests. Lawrence Erlbaum Ass. Publ. 1993, Hillsdale, N-J, 404 pp.
2. *Kapru X.* Основания математической логики: Пер. с англ. /Под ред. Ю.А. Гастева. — М.: Мир, 1969. — 568 с.
3. *Wright B.S., Masters G.N.* Rating Scale Analysis: Rasch Measurement. Chicago, MESA Press, 1982. — 206 pp.
4. *Брунер Дж.* Процесс обучения. — М., 1962. — С. 15.
5. *Аванесов В.С.* Проблема объективности педагогических измерений. //Педагогические измерения. — № 3, 2008. — С. 3–25.
6. *Аванесов В.С.* Проблема эффективности педагогических измерений // Педагогические измерения. — № 4, 2008. — С. 3–24. См. также в прошлых изданиях этого же журнала статьи автора по надёжности и валидности результатов тестирования.