

СОВРЕМЕННАЯ ТИПОЛОГИЯ ПЕДАГОГИЧЕСКИХ ТЕСТОВ

Джефри Вялфорд

*Перевод и адаптация к отечественной терминологии —
Кузнецов В. Г., кандидат педагогических наук.*

Теория и практика современного педагогического тестирования имеют не более чем столетнюю историю. Развитие психологии и педагогики с конца XIX века до начала Первой мировой войны было тесно связано со становлением тестовых технологий. Социальные и научные условия конца XIX и начала XX веков заставляли педагогов и психологов рассматривать результаты тестирования с точки зрения сравнения относительных способностей испытуемых, что стало со временем называться нормативно-ориентированным подходом к интерпретации результатов тестирования

Когда индивидуальные баллы интерпретируются по отношению к подготовленности самого экзаменуемого на данной области содержания, а не сравниваются с результатами других экзаменуемых из группы, то в этом случае говорят о критериально-ориентированном подходе к тестированию¹.

Идея критериально-ориентированного тестирования и сам термин появились в 60-е годы нашего столетия. К этому времени нормативно-ориентированное тестирование развивалось уже более полувека. Начиная с 70-х годов, критериально-ориентированные тесты получили всеобщее признание и стали широко применяться в сфере образования и профессиональной аттестации кадров. Они широко применяются в образовании, индустрии и вооружённых силах США, Великобритании, Голландии, Японии и в других развитых странах. Они обеспечивают

пользователей тестов ценной информацией, причём отличной от той, которая может быть получена в результате применения нормативно-ориентированных тестов.

Критериально-ориентированные тесты (domain-referenced tests — тесты, ориентированные на область содержания, и mastery tests — квалификационные тесты) используются во многих образовательных учреждениях развитых стран и решают различные задачи.

За последнее время стало обычным для государственных образовательных институтов определять совокупность знаний, умений и навыков, которые студенты должны усвоить, чтобы перейти с одного курса на другой или чтобы получить соответствующий диплом. В России эта совокупность получила своё воплощение в создании государственных образовательных стандартов. Использование критериально-ориентированных педагогических тестов, осно-

————— 1 —————

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; *Keeves J.P.* (Ed.) Educational Research, Methodology and Measurement: An International Handbook. Oxford, Pergamon press, 1988.

ванных на государственных образовательных стандартах, для итоговой и поэтапной аттестации студентов чрезвычайно перспективно.

Одно из самых важных приложений критериально-ориентированных тестов — сертификация и лицензирование в сфере профессиональной деятельности. В США, например, профессиональные организации часто проводят аттестационный экзамен в виде критериально-ориентированного теста (или батареи тестов), который испытуемые должны успешно пройти, чтобы получить возможность практиковаться в выбранной сфере профессиональной деятельности. Многие из этих профессиональных организаций применяют критериально-ориентированные тесты для повторной сертификации испытуемых. Типичный профессиональный экзамен, основанный на критериально-ориентированных тестах, измеряет знания, умения и навыки, необходимые для профессиональной деятельности, а результаты тестирования интерпретируются по отношению к минимальному установленному стандарту оценивания.

Понятие нормативно-ориентированного и критериально-ориентированного педагогического теста

Существует два основных подхода к разработке тестов для конкурсного, а также профессио-

нального отбора, и для аттестации учащихся и специалистов: нормативно-ориентированный и критериально-ориентированный. Эти два подхода различаются на всех этапах разработки с момента выбора содержания до момента интерпретации результатов.

Самая общая их характеристика заключается в следующем².

Нормативно-ориентированный подход (norm-referenced) позволяет сравнивать учебные достижения (уровень подготовки, уровень профессиональных знаний и умений) отдельных испытуемых друг с другом.

Критериально-ориентированный подход (criterion-referenced) позволяет оценивать, в какой степени испытуемые овладели необходимым учебным материалом

Для аттестационного контроля уровня обученности более привычной и естественной является задача, решаемая в рамках критериально-ориентированного подхода. Тем не менее, в процессе тестового контроля оба подхода необходимы.

Тесты измеряют степень усвоения учебного материала, овладения необходимыми знаниями, умениями и навыками, уровень профессиональной квалификации кадров. В результате измерения степень проявления какого-либо свойства испытуемых выражается в тестовом балле, представляющем собой определённое число. Тестовые баллы располагаются на

Теория

2

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980.

специальных шкалах, выбранных разработчиками.

Нормативно-ориентированные педагогические тесты используются для того, чтобы получить надёжные и нормально распределённые баллы для сравнения тестируемых.

Критериально-ориентированные педагогические тесты применяются для того, чтобы интерпретировать результат тестирования в соответствии с уровнем обученности испытуемых на хорошо определённой области содержания.

Хотя различия нормативно-ориентированных и критериально-ориентированных тестов существенны, они имеют много общего. С первого взгляда на сам тест довольно трудно определить, какой из этих двух типов имеет место: используют те же формы тестовых заданий и подобные инструкции для испытуемых.

Первое различие — цель создания теста. Нормативно-ориентированные тесты составлены специально для того, чтобы получить возможность сравнивать испытуемых в той области содержания, для которой тест предназначен. Для этого используются нормативные или стандартные шкалы.

Критериально-ориентированные тесты составлены для того, чтобы аттестовать результаты испытуемого в соответствии с областью определённых знаний, умений и навыков. Результаты критериально-ориентированного тестирования могут быть использованы для:

1) оценки итогового уровня обученности испытуемого, соответствия этого уровня государственным образовательным стандартам или программе обучения;

2) отбора студентов, достигших необходимого уровня обученности, в том числе и профессиональной компетенции;

3) оценки эффективности программы обучения в том или ином учебном заведении.

Результаты критериально-ориентированного тестирования также могут быть использованы для сравнения между собой уровней подготовки студентов, однако в этом случае можно получить сравнительно низкую надёжность, если распределение баллов однородно и имеет малую вариацию.

Второе различие этих двух типов тестов состоит в уровне детализации области содержания. От разработчиков обоих типов тестов обычно требуется формирование спецификации его содержания. Создатели критериально-ориентированных тестов должны в типичных случаях подготовить значительно более детализированную спецификацию содержания, чем составители нормативно-ориентированных тестов. Это необходимо для того, чтобы пользователи теста были уверены в адекватной интерпретации тестовых баллов.

Третье различие лежит в плоскости проведения статистической обработки результатов тестирования. Обработанные (или шкалированные) баллы по результатам

нормативно-ориентированного тестирования базируются на статистических данных нормативной группы, то есть специфической достаточно большой выборке испытуемых. В большинстве случаев для этого типа тестов применяются специальные нормативные шкалы. Каждый индивидуальный балл для данного теста имеет однозначное соответствие с процентильным эквивалентом, определённым на нормативной группе. Если для индивидуального балла данного испытуемого процентильный эквивалент равен 75, то это означает, что у 75% тестируемых из нормативной группы результаты тестирования были такие же или хуже. Обработанные индивидуальные баллы по результатам критериально-ориентированного тестирования не относятся к какой-либо нормативной группе или выборке испытуемых. Индивидуальный балл испытуемого интерпретируется по отношению к доле учебного материала, успешно им освоенного. Чаще всего балл студента отражает процент правильно выполненных заданий и выражается в шкале процентов.

Четвёртое основное различие лежит в плоскости технологии анализа и отбора тестовых заданий. В нормативно-ориентированных тестах статистические показатели тестовых заданий (уровень трудности и различающая способность) играют важную, а часто определяющую роль в отборе заданий. В общем случае зада-

ния со средним уровнем трудности и высокой различающей способностью имеют больший шанс быть отобранными для использования в тесте этого типа, так как они вносят существенный вклад в вариацию индивидуальных баллов студентов. Надёжность теста, в общем случае, будет выше, если вариация тестовых баллов увеличится. Напротив, в критериально-ориентированных тестах статистические характеристики (уровень трудности и различающая способность) заданий не является основной причиной для их включения в состав теста или, наоборот, исключения из него. Главное условие отбора заданий в критериально-ориентированный тест — это их соответствие (их конгруэнтность) спецификации и элементу содержания. Статистические характеристики тестовых заданий используются для составления параллельных форм критериально-ориентированных тестов и для выбора оптимального стандарта оценивания.

Эти два типа тестов различаются и по методам конструирования, и по особенностям применения. Однако основным критерием для разделения тестов на нормативно- и критериально-ориентированные является подход к интерпретации тестового балла.

Существует и ряд других важных различий между ними. Характерные особенности этих двух типов тестов можно свести в таблицу.

Теория

Типы тестов, их различия

ПЕД диагностика
ПЕД диагностика

3

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989; Educational Testing Service. Standards for quality and fairness. Princeton, NJ. 1987; *Keeves J.P.* (Ed.) Educational Research, Metodology and Measurement: An International Handbook. Oxford, Pergamon press, 1988.

4

Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989.

Нормативно-ориентированные тесты

В большинстве тестовых программ за рубежом использовались и используются в настоящее время различные виды шкалированных (обработанных) баллов. В практике массового педагогического тестирования существует более или менее стандартный набор шкалированных баллов, которые информируют пользователей теста об относительных способностях испытуемых³.

К таким шкалированным (обработанным) баллам относятся: процентильные эквиваленты; стандартные линейные шкалы; стандартные нормативные шкалы и другие виды педагогических шкал⁴. Шкалированные баллы помогают пользователям тестов получить информацию о позиции экзаменуемого по отношению к хорошо определённой группе, определённой специфической выборке испытуемых.

Эта специфическая выборка испытуемых часто называется **нор-**


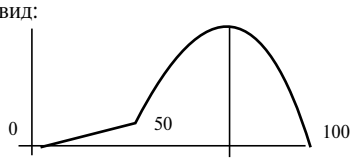
мативной группой, и так как обработанные баллы ориентированы на баллы в нормативной группе, они называются **нормативно-ориентированными баллами**.

В основе нормативно-ориентированной интерпретации лежит сопоставление индивидуального тестового балла данного студента с баллами, полученными другими испытуемыми. Самый простой способ провести нормативно-ориентированную интерпретацию тестового балла состоит в том, чтобы сообщить, сколько процентов учащихся из учебной группы, в которой проводилось тестирование, выполнили тест хуже (получили более низкий балл) и сколько процентов выполнили тест лучше (получили более высокий балл). Однако при такой интерпретации возникает следующая трудность: оценка индивидуальных учебных достижений оказывается зависимой от уровня подготовленности всей группы студентов, в которой проводилось тестирование. Гипотетически преодолеть эту проблему

Таблица

Характерные особенности нормативно-ориентированных и критериально-ориентированных тестов

Нормативно-ориентированные тесты	Критериально-ориентированные тесты
<p>1. Цель тестирования: возможность сравнения уровня подготовки испытуемых друг с другом в той области содержания, для которой тест предназначен. Пример использования: конкурсный отбор кандидатов на обучение</p>	<p>1. Цель тестирования: возможность аттестации испытуемого в соответствии с его уровнем усвоения определённой области содержания. Пример использования: итоговая аттестация уровня обученности студентов, уровня профессиональной подготовки кадров</p>

Нормативно-ориентированные тесты	Критериально-ориентированные тесты
<p>2. Используемые шкалы: нормативные (или стандартные) шкалы. Необходимо указание среднего значения и стандартного отклонения в выбранной шкале</p>	<p>2. Используемая шкала — в основном, шкала процентов с выбранным одним (или несколькими) критериальным баллом (баллами). Особое внимание уделяется методике оптимального выбора критериального балла (или баллов)</p>
<p>3. Распределение индивидуальных баллов: близко к нормальному, в большинстве случаев имеет вид:</p> 	<p>3. Распределение индивидуальных баллов: произвольное, в большинстве случаев асимметричное и имеет вид:</p> 
<p>4. Уровень детализации области содержания — не существенен. Авторы теста выбирают наиболее значимые элементы содержания</p>	<p>4. Уровень детализации области содержания — подробный. Авторы теста разрабатывают спецификацию (план) теста, включающую все элементы содержания. Затем по этой спецификации разрабатываются задания</p>
<p>5. Нормативная группа испытуемых обязательна. Обработанные (или шкалированные) баллы по результатам нормативно-ориентированного тестирования базируются на статистических данных нормативной группы, то есть специфической достаточно большой выборке испытуемых. В большинстве случаев применяются специальные нормативные таблицы, где каждый индивидуальный балл для данного теста имеет однозначное соответствие с процентильным эквивалентом, определённым на нормативной группе</p>	<p>5. Нормативная группа испытуемых не является необходимой. Индивидуальный балл испытуемого интерпретируется по отношению к доле учебного материала, успешно им освоенного. Чаще всего балл студента отражает процент правильно выполненных заданий и выражается в шкале процентов</p>
<p>6. Статистический анализ и отбор тестовых заданий. Статистические показатели тестовых заданий (в основном это уровень трудности и различающая способность) играют важную роль в отборе заданий. Выбираются задания со средним уровнем трудности (от 0,3 до 0,7) и высокой различающей способностью (большей 0,3). Существует ряд других важных статистических показателей качества заданий</p>	<p>6. Статистический анализ и отбор тестовых заданий. Уровень трудности и различающая способность заданий не являются существенными факторами включения в состав теста или, наоборот, исключения из него. Главное условие отбора заданий — это их соответствие (их конгруэнтность) спецификации и элементу содержания. Статистические характеристики тестовых заданий используются для составления параллельных форм (вариантов) теста и для выбора оптимального критериального балла</p>
<p>7. Надёжность теста оценивается либо путём нахождения корреляции между результатами двух тестирований, либо методом расщепления теста на две половины при однократном тестировании</p>	<p>7. Надёжность теста оценивается степенью постоянства принятия решения «зачёт — незачёт» при двукратном тестировании</p>

Нормативно-ориентированные тесты	Критериально-ориентированные тесты
<p>8. Валидность. Наряду с содержательной валидностью для тестов конкурсного отбора учащихся особое внимание уделяется высоким показателям прогностической валидности</p>	<p>8. Валидность. Особое внимание уделяется содержательной валидности. В случае принятия важных решений по результатам тестирования исследуются критериальная и конструктивная валидность</p>

можно было бы следующим образом: протестировать всю популяцию, для которой разрабатывался тест, и сравнить индивидуальный балл с полученными результатами. Под «популяцией» понимается всё множество учащихся, которые обучаются данной дисциплине и по данной программе. Если тест разрабатывается для применения в каком-либо одном учебном заведении, то в популяцию включают все учебные группы, обучающиеся по одной программе. Если тест разрабатывается для применения в рамках региона или страны, то популяция принимает весьма внушительные размеры. Протестировать такое количество испытуемых практически невозможно. Поэтому, чтобы получить информацию с том, как выполнили тест все испытуемые, прибегают к формированию так называемой выборки стандартизации. Это специально подобранная группа испытуемых, которая адекватно репрезентирует популяцию, для которой разрабатывается этот тест.

В выборке стандартизации пропорционально представляются все возрастные, социальные и другие группы испытуемых, из которых состоит популяция. Результаты проведения теста на образованной таким образом выборке назы-

ваются «тестовыми нормами». Отсюда происходит термин «нормативно-ориентированное тестирование». По нормам судят о том, как выполнили бы тест все испытуемые, для которых он предназначен. Если выборка стандартизации адекватно репрезентирует популяцию, то, сопоставляя индивидуальный балл с нормами, можно получить оценку уровня учебных достижений испытуемого по сравнению с уровнем учебных достижений других учащихся. Такая оценка уже не зависит от группы, в которой проводилось тестирование⁵.

Педагогические тесты, специально разработанные для того, чтобы провести нормативно-ориентированную интерпретацию, называются нормативно-ориентированными тестами.

Нормативно-ориентированные тесты имеют свою специфику, связанную с их основной задачей — **дифференцировать испытуемых**. При разработке этого типа стараются добиться высокой **вариативности тестовых баллов**. Если большинство испытуемых получает только низкие, или только средние, или только высокие баллы, то труднее установить между ними различия. Если же вариативность баллов высокая (встречаются и низкие, и средние, и высокие бал-

лы), то дифференцировать испытуемых легко. Нормативно-ориентированные тесты разрабатываются таким образом, чтобы распределение баллов было близко к нормальному. Математические методы, применяемые при конструировании и оценке нормативно-ориентированных тестов, рассчитаны на модель нормального распределения и основываются в основном на корреляционном анализе⁶.

Нормативно-ориентированный подход чрезвычайно полезен, когда необходима информация об испытуемом, связанная с его относительными способностями или относительной обученностью по сравнению с другими испытуемыми. Естественно, потенциальная ценность интерпретации нормативно-ориентированного подхода зависит и от соответствия нормативной группы целям тестирования, и от чистоты и тщательности, с которой была составлена нормативная группа.

Для того чтобы определить способности абитуриента по математике, очень важно знать, что нормативная группа репрезентативна для всей группы абитуриентов. Иногда для принятия важных решений необходимо ориентироваться на более специфическую нормативную группу. Например, нормативная группа абитуриентов на инженерные специальности будет более подходить для интерпретации относительной способности абитуриента на инженерные специальности по результатам нормативно-ориентированного тестиро-

вания, чем просто нормативная группа всех абитуриентов.

Отметим основные недостатки нормативно-ориентированного подхода к педагогическому тестированию. Для большинства педагогических тестов необходимо знать гораздо больше об испытуемом, чем даёт нормативно-ориентированный подход. Например, для некоторого абитуриента мы получили 65 процентильных эквивалентов в результате тестирования. Это означает, что данный абитуриент ответил на задания данного теста не хуже, чем 65% испытуемых из всей, достаточно большой нормативной группы. Но мы не знаем, какие именно понятия усвоены данным абитуриентом и какие задачи он может решать. Требовались ли в данном тесте от данного абитуриента только легко запоминаемые для решения алгоритмы, или от него требовалось творческое решение, творческий подход для решения заданий. Более того, мы не знаем, какую долю от всей программы твёрдо усвоил данный испытуемый. Подобного рода вопросы становятся важными тогда, когда намечено использование теста выходит за рамки определения относительных способностей экзаменуемого. Определение относительных способностей абитуриентов может быть важным для конкурсных экзаменов, но после учебного процесса преподавателям необходимо узнать, какого вида понятия каждый студент усвоил, какие типы задач он может решать, какую долю от изученного материала усвоил.

Теория

100000

6

Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989; Educational Testing Service. Standards for quality and fairness. Princeton, NJ. 1987.

ПЕД диагностика
ПЕД диагностика**Критериально-ориентированные тесты**

Существует, однако, и подход, отличный от нормативно-ориентированного, когда **обработанные баллы интерпретируются по отношению к подготовленности самого экзаменуемого на данной области содержания**. В этом случае говорят о **критериально-ориентированном подходе**⁷.

Такие тесты называются критериально-ориентированными. Однако педагогический тест может при специальной технологии составления и интерпретации предоставлять как нормативно-ориентированную, так и критериально-ориентированную информацию для пользователей.

В литературе существует много определений критериально-ориентированного теста. Авторы используют этот термин различно, поэтому необходимо однозначно определить критериально-ориентированный подход. Каждый вид критериально-ориентированного теста составлен для того, чтобы интерпретировать результаты тестирования по отношению к определённой области содержания. Эта интерпретация может совершаться различными способами, которые зависят от цели тестирования. Определение теста должно служить базисом для установления границ данной концепции. Критериально-ориентированный тест — это такой тест, который намеренно составлен, чтобы производить педагогические измерения и напрямую интерпретиро-

ваться в терминах точно определённого стандарта. Данные педагогические измерения производятся путём выбора репрезентативного набора тестовых заданий на соответствующей области содержания.

Критериально-ориентированные тесты специально составлены таким образом, чтобы обобщить результаты индивидуальных баллов относительно точно определённого домейна или области содержания⁸. Причём тестовые задания могут включать как проверку содержания, так и проверку навыков и умений. Домейн, или область содержания, называется хорошо определённой, если совершенно ясно, какая категория тестовых заданий потенциально принадлежит данной области содержания. Основное использование критериально-ориентированного тестирования — определение статуса испытуемого по отношению к некоторому количеству тестовых заданий выбранных или из определённой области содержания, причём таким образом, который позволяет пользователю теста сделать заключение об учебных достижениях испытуемого. Таким образом, хорошо определённая область содержания является необходимым, но недостаточным условием критериально-ориентированного тестирования.

Часто можно встретить довольно узкие определения, подразумевающие только один из видов критериально-ориентированных тестов. Более широкие определения нередко страдают неточнос-

7

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989.

8

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980.

тью формулировок. В работе⁹ приведено определение, которое, пожалуй, наиболее точно раскрывает сущность и специфику этого типа тестов.

«Критериально-ориентированный педагогический тест представляет собой систему заданий, позволяющую измерить уровень учебных достижений относительно полного объёма знаний, умений и навыков, которые должны быть усвоены учащимися».

Мы будем называть областью содержания теста тот полный объём знаний, умений и навыков, которые должны быть усвоены студентами в результате определённого курса обучения и овладение которыми измеряется критериально-ориентированным тестом. Выполнение критериально-ориентированного теста описывается, как правило, в терминах содержания учебной дисциплины. Например, результаты проведения теста позволяют судить о том, какую долю области содержания освоил студент, какие задачи и какого уровня сложности может решать испытуемый, и т.п.

Два основных вида критериально-ориентированных тестов

В литературе по тестологии (или педагогическим измерениям) выделяют два основных вида критериально-ориентированных тестов, отличающихся друг от друга по ряду характеристик. Установившихся названий для этих видов на русском языке пока не существует. В зарубежной литературе терминология окончательно также

ещё не установилась. Критериально-ориентированные тесты намеренно составляются таким образом, чтобы аттестовать уровень индивидуальных достижений студентов по отношению к заранее определённой области содержания. Принципиальный интерес при их разработке представляет строгая и точная спецификация области содержания теста. Это необходимо для того, чтобы увеличить до максимального уровня обоснованность интерпретации индивидуальных баллов студентов¹⁰. Именно акцентирование внимания на области содержания (по англ. — domain) внушило некоторым исследователям идею наименования таких тестов *domain-referenced*, что можно перевести как «тесты, ориентированные на область содержания».

Создание заданий для теста, ориентированного на область содержания, требует более детальных и точных правил. Для этого требуется точная и строго определённая спецификация области содержания теста. Этот тип критериально-ориентированных тестов предназначен для оценки доли от полного объёма учебного материала, которую усвоил испытуемый. В зарубежной литературе его обычно называют **domain-referenced test**. Мы будем называть этот тип критериально-ориентированных тестов термином: «тесты, ориентированные на определённую область содержания». Предполагается, что в результате обучения учащиеся

Теория

9

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980.

10

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; *Educational measurement* (Ed. by Linn R.) N.Y. Macmillan, 1989; *Keeves J.P.* (Ed.) *Educational Research, Methodology and Measurement: An International Handbook*. Oxford, Pergamon press, 1988.

ПЕД диагностика
ПЕД диагностика**11**

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989.

12

Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989.

13

Berk R.A. Criterion-referenced measurement: The state of art. Baltimore, MD: Johns Hopkins University Press, 1980; Educational measurement (Ed. by Linn R.) N.Y. Macmillan, 1989; Educational Testing Service. Standards for quality and fairness. Princeton, NJ. 1987.

должны приобрести определённую сумму знаний, умений и навыков, которая является областью содержания теста и условно принимается за 100 процентов. Уровень учебных достижений каждого испытуемого можно выразить в проценте от полного объёма области содержания теста. Для итогового контроля наиболее естественно применять именно этот тип тестов.

На практике существует и другой тип критериально-ориентированных тестов. Называются такие тесты **mastery tests**, то есть «тесты на мастерство» или «квалификационные тесты»¹¹. Они применяются для определения в группе испытуемых квалифицированных и неквалифицированных (masters and nonmasters) в какой-то области. Эмпирическая процедура анализа тестовых заданий в данном случае рекомендована для того, чтобы определить «чувствительность» тестовых заданий к различию между обученными и необученными испытуемыми.

Этот вид используется для классификации испытуемых и разделения их на группы на основании заранее выбранного стандарта оценивания (критериально-го балла). Чаще всего такие тесты используются для разделения испытуемых на две группы: **усвоивших и не усвоивших необходимый навык на базе определённой области содержания**. Иногда происходит деление на большее количество групп. Например, на тех, которые полностью овладели навыком, и

на тех, которым лучше закрепить его, а также на группу учащихся, которым надо начать изучение материала с самого начала. Для отнесения к определённой группе испытуемый должен достичь необходимого минимального стандарта оценивания. Этот стандарт устанавливается разработчиками теста и является критерием, на основании которого принимается решение относительно испытуемого. В тесте данный стандарт выражается определённым количеством правильно выполненных заданий. Важными элементами разработки таких тестов являются: методы установки стандарта оценивания квалификации по результатам тестирования, оценка уровня ошибки этого стандарта, определение последовательности принятия решения о квалификации или о недостаточной подготовке студента¹².

Многие учёные¹³ отмечают, что термин «критериально-ориентированное тестирование» неудачен. Он создаёт впечатление, будто существует некий критерий, с которым сопоставляются результаты тестирования. Одно из заблуждений относительно критериально-ориентированного тестирования — это неправильное понимание того, что данный термин означает использование так называемого критерия или «критериального (проходного) балла». Критериально-ориентированное тестирование не всегда требует наличия проходного балла, хотя для определённых практических при-

ложений при данном виде тестирования могут потребоваться так называемые стандарты оценивания или проходные баллы. Такой критерий, а точнее стандарт оценивания, в самом деле используется, но его наличие не обязательно. Ошибочная трактовка критерия как уровня подготовленности — это одно из заблуждений относительно критериально-ориентированного тестирования. Это неправильное понимание того, что этот термин означает использование так называемого «проходного балла».

Пользователь теста может использовать или установить проходной балл для любого теста — критериально-ориентированного или не критериально-ориентированного. Если просто установить 80% правильно решённых заданий как «проходной балл» для любого теста, это не означает автоматически, что вы используете хорошо определённую область содержания и описываете результаты тестирования по отношению к этой области содержания. Способность самого теста давать данную информацию об испытуемых, то есть давать статус испытуемых по отношению к данной области содержания, и делает тест критериально-ориентированным, а не тот факт, что вы использовали проходной балл или другой критерий оценивания.

Существует и другое заблуждение, а именно о связи критериально-ориентированного теста с критериальной интерпретацией педагогического измерения. Иногда тест составляется таким образом, чтобы прогнозировать статус испытуемого для другого теста или для других педагогических измерений, таких как экзамены, зачёты и т.д. Например, некоторые тесты, такие как широко известные SAT-I и SAT-II, составлены для того, чтобы помочь предвидеть, прогнозировать учебные успехи абитуриентов в университетах США¹⁴. Та переменная величина, которая определяется для данного прогноза, называется критериальной величиной.

Таким образом, итоговые оценки студента в университете являются критериальным измерением для вступительного теста. Данное педагогическое исследование, например тест SAT, проводится для того, чтобы установить полезность, обоснованность и фактически критериальную валидность самого вступительного теста, как прогностического теста установления статуса испытуемых, по отношению к критериальной переменной. Использование термина «критериальный» в данном случае отличается от использования этого же термина в связи с критериально-ориентированным тестированием.

Теория

14

14

Educational measurement (Ed. by Linn R.)
N.Y. Macmillan, 1989.