

ТРЕБОВАНИЯ К ТЕСТАМ ДЛЯ ОПРЕДЕЛЕНИЯ УРОВНЯ ОБРАЗОВАТЕЛЬНЫХ ДОСТИЖЕНИЙ СТУДЕНТОВ

Ветрова В.Т., Полушкина С.И.

Белорусский государственный аграрный технический университет

Педагогические тесты критериально-ориентированного типа — это комплекс специальных контрольных заданий, направленных на выявление факта усвоения деятельности определенного уровня в сочетании с определенной системой измерения и оценки качества усвоения.

Основные отличия теста от обычного контрольного задания:

1. Тесты — это инструмент объективного контроля (тестирование — метод), поэтому они являются необходимым компонентом проведения контроля в обучении.

В отличие от обычных проверочных заданий тесты ориентированы на диагностические параметры целей образования и измеряют достижение именно этих целей. Благодаря тестированию достигается объективность измерения и оценки качества обучения. Описание целей образования по диагностическим параметрам позволяет распознавать степень их реализации посредством тестов точно, однозначно и воспроизводимо, т.е. объективно.

2. Тест представляет собой систему заданий, составляемых по стандартной форме и определенным правилам.

3. Тестирование является одним из наиболее технологичных методов проведения автоматизированного контроля.

4. Каждое тестовое задание обеспечивается эталоном.

Современный тест содержит три основных компонента:

- тестовое задание;
- эталон ответа;
- весовой коэффициент.

Исчерпывающее раскрытие понятия «тест» требует ознакомления с такими понятиями как диагностичность целей, тестовое задание, эталон, уровень образовательных достижений, надежность и валидность тестов.

Диагностическое задание цели обучения с учетом качества усвоения знаний и умений состоит в определении необходимого уровня усвоения. Говоря иначе, уровень усвоения выступает как цель.

Диагностическое целеполагание включает в себя:

- вполне определенное (конкретное) и однозначное описание целей;
- указание способов выявления, измерения и оценки степени их реализации (объективный контроль).

Эталон (полный и правильный ответ с указанием существенных операций теста) разрабатывается экспертным методом к каждому заданию для измерения и оценки результата выполнения этого задания. По эталону определяется число существенных операций, ведущих к решению тестового задания.

Сопоставляя ответ студента с эталоном, делают выводы о качестве выполнения теста. Тест, лишенный эталона, превращается в обычное контрольное задание, решение о качестве выполнения которого подвержено всем возможным особенностям субъективного суждения.

Результаты тестирования могут представляться в балльных или рейтинговых системах оценивания. Наибольшее распространение находит балльная система оценки в виде оценок по 100-балльной или 10-балльной шкале.

Применительно к педагогическим измерениям в качестве эмпирических индикаторов выступают тестовые задания. В отличие от обычно принятых заданий, имеющих в экзамене

национных билетах, здесь речь идет о системе более дробных, коротких, взаимосвязанных между собой общей логикой заданий, отвечающих определенным научно обоснованным критериям качества.

Характер заданий в значительной степени определяется спецификой и логикой дисциплины, по которой осуществляется тестирование.

В зависимости от целей контроля тестовые задания подразделяют на пять уровней образовательных достижений, представленных в таблице

Уровни образовательных достижений	Процент правильно выполненных основных мыслительных операций	Оценка по 10-балльной шкале
рецептивный	70–90%	3
	90–100%	4
репродуктивный	70–90%	4
	90–100%	5
репродуктивно-продуктивный	70–80%	5
	80–90%	6
	90–100%	7
продуктивный	70–80%	7
	80–90%	8
	90–100%	9
продуктивно-творческий	70–80%	9
	80–100%	10

Рецептивный уровень знаний соответствует узнаванию и различению объектов изучения программного материала, предъявляемых в готовом виде.

Репродуктивная деятельность — это узнавание и воспроизведение программного материала, применение приобретенных знаний в стандартных ситуациях.

Продуктивная деятельность — это получение на основе ранее приобретенных знаний субъективно или объективно новых знаний; самостоятельное применение полученных знаний в нестандартной ситуации.

Под валидностью (адекватностью) теста понимают точное соответствие содержания выявляемому признаку. Так, если речь идет о тестировании деятельности на продуктивном уровне, то задания, предлагаемые в тесте, должны быть решаемы только в ходе деятельности продуктивного уровня и не решаемы путем использования способов деятельности более низкого уровня.

Валидность теста в этом отношении называется функциональной валидностью.

Функциональная валидность теста — соответствие задания теста выявляемому уровню усвоения, заданному образовательным стандартом.

Оценка функциональной валидности теста может быть осуществлена как путем его теоретического анализа, так и экспериментально.

При теоретическом анализе функциональной валидности теста проводится экспертная оценка структуры и содержания каждого действия или операции эталона и исследуется возможный путь ее осуществления студентами. Если оказывается, что все или подавляющее число существенных операций выполнения теста осуществляются только на основе деятельности данного уровня, делается заключение о функциональной валидности теста.

Деятельность студента по тесту сопоставляется с другими способами оценки его деятельности на данном уровне, по отношению к которым установлена их функциональная валидность.

Под содержательной валидностью теста понимают его соответствие содержанию опыта, приобретенного обучающимися. Содержательно валиден такой тест, с помощью которого предъявляемые задания на деятельность (по содержанию и уровню деятельности) обязательно должны быть выполнимы студентами по условиям предшествующего обучения. Содержательная валидность теста также оценивается сначала с помощью теоретического анализа (экспертизы), а затем — экспериментально. Теоретический анализ при этом состоит в сопоставлении содержания теста учебным программам и учебным пособиям, а его уровня — задачам обучения.

Экспериментальная оценка содержательной валидности теста легко может быть получена в естественном эксперименте. Он состоит в сопоставлении результатов тестового испытания качества усвоения опыта с другими способами (в том числе и экспертными) определения этого же качества, по отношению к которым имеются доказательства их содержательной валидности.

Третий показатель валидности теста — его прогностическая (предсказательная) валидность. О прогностической валидности теста судят по его связи с успешностью последующей учебной или производственной деятельности студента. Показатель прогностической валидности, очевидно, имеет существенное значение для оценки качества или смысла применения теста успешности. Если его предсказательная сила достаточно велика, то очевидны выводы из этого факта.

Выделим внутритестовую прогностическую валидность — характеристику, относящуюся только к тестам-лестницам. Эта характеристика показывает, насколько по выполнению тестов более низкого уровня можно предсказать качество знаний студентов на более высоких уровнях деятельности.

Все три показателя валидности теста — функциональная, содержательная и прогностическая — имеют количественную меру выражаемого ими свойства. Математический способ описания этой меры одинаков и исходная посылка для ее получения для всех признаков также общая — сопоставление оценки качества усвоенной деятельности по тесту с оценкой этого же качества другим, но определенно объективным, способом. Это сопоставление выражается с помощью коэффициента корреляции.

Надежность — устойчивость результатов теста, получаемых при его применении. Существуют следующие методы оценки надежности:

- параллельное тестирование в двух группах;
- ретестовая надежность (повторное тестирование);
- расщепление теста (тест расщепляют на две части по четным и нечетным номерам заданий — чем выше корреляция между двумя частями, тем выше надежность).

Под длиной теста понимается количество заданий, входящих в тест. Наряду с «нормальными» тестами существуют тесты и очень короткие, состоящие из 7–15 заданий, и очень длинные, содержащие не более пятисот заданий. В классической теории тестов применяется следующее утверждение: при прочих равных условиях надежность теста зависит от числа заданий (чем длиннее тест, тем он надежнее). Но практика показывает, что если тест очень длинный, то ухудшается мотивация и внимание у испытуемых. Противоречие между теорией и практикой решается компромиссом в ту или другую сторону в каждом конкретном случае. Практически длину теста следует определять опытным путем с учетом валидности и ряда факторов, таких как время, организационные возможности, стоимость и др. Оптимальная длина теста, как показали теория и практика, — 40–60 заданий.