

## ПРАКТИКА ЭКСПЕРТНОЙ ОЦЕНКИ ПРИГОДНОСТИ ТЕСТА ДЛЯ КОНТРОЛЯ ЗНАНИЙ СТУДЕНТОВ В САМАРСКОМ ГОСУДАРСТВЕННОМ МЕДИЦИНСКОМ УНИВЕРСИТЕТЕ

© 2015 Е.В. Мензул, Н.М. Рязанцева

Самарский государственный медицинский университет

Поступила в печать 14.03.2015

Использование тестовых заданий в учебном процессе помогает в мониторинге оценки знаний студентов. Для повышения качества тестовых заданий необходима их экспертиза. Проведение экспертизы повышает качество составления тестов и качество оценки знаний студентов.

*Ключевые слова:* экспертиза тестовых заданий, валидность, надежность, индекс трудности, индекс дискриминативности.

Тестовые задания в современном образовательном процессе уже давно заняли прочное место. Они используются на всех этапах контроля знаний студентов: текущего, рубежного и итогового.

Разработкой тестовых заданий занимаются практически все преподаватели университета, разработанные тестовые задания сразу же используются в учебном процессе. Однако при разработке не всегда выполняются все требования, предъявляемые к ним, что влечет за собой возникновение ошибок.

Как показала практика, разработчики тестов в своей работе допускают одни и те же ошибки, а именно:

- содержание тестов тематически соотносится с содержанием изучаемой темы, однако не полностью соответствует формулировкам заявленных учебных целей;
- выбранный тип теста не всегда соответствует заявленному уровню освоения учебного материала;
- формулировки тестовых заданий не всегда корректны;
- предлагаемые тестовые задания являются очень легкими или очень трудными для студентов;
- не учитывается психология студентов [2].

Для составителей тестов в СамГМУ была разработана методика экспертной оценки пригодности теста для контроля знаний студентов, в которой были прописаны все этапы работы по формированию тестов – от создания до оценки их качества, определена организационная сторона деятельности, предъявлены требования к оформлению документации.

Разработка тестов по учебным темам, дисциплинам поручается наиболее опытным сотрудникам кафедры, преподающим дисциплину. Для оценки пригодности теста к использованию

*Мензул Елена Владимировна, кандидат психологических наук, доцент, заведующая кафедрой педагогики, психологии и психолингвистики. E-mail: lenamenzul@rambler.ru*  
*Рязанцева Наталья Михайловна, старший преподаватель кафедры педагогики, психологии и психолингвистики.*

в учебном процессе на кафедре создается экспертная комиссия. В состав комиссии могут входить заведующий кафедрой, профессора, доценты и преподаватели с ученой степенью. Количественный состав комиссии – не менее трех человек (при этом число экспертов не может быть четным). Состав комиссии утверждается на заседании кафедры. Целью работы комиссии является определение качества контрольно-измерительных материалов, их соответствие требованиям, предъявляемым к фонду оценочных средств.

Первым этапом работы является составление теста. Разработчиками должны быть выполнены следующие шаги:

- 1) определить назначение теста (предварительный, текущий, рубежный, итоговый контроль);
- 2) осуществить отбор учебных элементов, подлежащих контролю в соответствии с заявленными целями;
- 3) определить вид тестовых заданий для каждого учебного элемента, подлежащего контролю в соответствии с уровнем усвоения, заявленного в целях:
  - а) открытого типа (дополнения, свободное изложение),
  - б) закрытого типа (альтернативные ответы, множественный выбор, восстановление соответствия, восстановление последовательности);
  - 4) сформулировать тестовые задания.

Тестовое задание состоит из трех основных частей: 1) инструкция; 2) текст задания (вопрос); 3) эталон ответа.

*Инструкция* содержит указания на то, что испытуемый должен сделать, каким образом выполнить задание (выбрать, сгруппировать, дописать и т.п.). Для однотипных заданий допускается одна инструкция, которая помещается в начале данной группы заданий.

*Текст задания* или вопроса представляет собой содержательное наполнение задания. Оно может быть представлено в виде текста, рисунка, таблицы, графика, символа, цифры и т.д.

Эталон ответа предназначен для осуществления анализа правильности ответа испытуемого. Тестовое задание должно иметь однозначный правильный ответ;

5) сгруппировать тестовые задания в тест.

Тестовое задание должно располагаться на одной странице, шрифтовое и полиграфическое оформление должны быть удобны для чтения. Кроме перечисленных частей в состав теста может входить специально подготовленный бланк ответов.

Следующим этапом работы является предварительная экспертиза теста. Предварительная экспертиза является обязательным условием для составления тестов, т.к. позволяет устранить ряд недостатков до апробации теста.

Задача экспертной группы определить валидность (содержательную, конструктивную), надежность теста. Определение данных показателей качества теста достигается коллегиальностью обсуждения экспертами.

При определении **содержательной валидности** теста выявляется, насколько содержание контрольных заданий соответствует содержанию учебных целей. При определении **конструктивной валидности** выявляется соответствие тестового задания уровню учебных целей, т.е. уровню усвоения, определяемому учебными целями.

На данном этапе экспертизы осуществляется анализ **надежности** тестового инструментария, т.е. оцениваются возможные источники неудовлетворительной надежности теста. Источниками неудовлетворительной надежности, как правило, являются: величина теста, непонятность и двусмысленность заданий или инструкции, возможность случайного угадывания правильных ответов, случайных ошибок и т.п. [1].

Эксперты по пятибалльной шкале оценивают тестовые задания по следующим параметрам:

1. соответствие заявленной теме (учебные элементы);
2. соответствие уровню усвоения (целям);
3. качество формулировок инструкции и текста тестового задания;
4. качество иллюстративного материала (графиков, рисунков и т.п.);
5. правильность эталона;
6. качество комплектования (последовательность) тестовых заданий.

Средний балл экспертной оценки теста по критерию «надежность» должен быть не менее 3 баллов. Если экспертной группой выявлены недостатки, они подлежат исправлению до начала следующего этапа работы.

Третий этап работы - первичная апробация тестовых заданий.

Целями первичной апробации тестовых заданий являются сбор информации для определения показателей качества теста, трудности и

дискриминативности, а также выявления случайных ошибок (опечатки, пунктуация), уточнения времени, необходимого для решения теста, формулировки правильных ответов, посредством анализа ответов студентов на открытые вопросы, выявления недостатков в инструкции.

Для проведения апробации формируется выборка. Объем выборки определяется по формуле:

$$n = t^2 W (1 - W) N / d^2 N + t^2 W (1 - W),$$

где:  $n$  – численность выборки (табл. 1);

$t$  – коэффициент кратности ошибки, связанный с вероятностью ( $P_t$ )=1,98, с которой требуется гарантировать результаты выборочного наблюдения ( $t$ -критерий);

$d$  – размер допустимой ошибки выборки (0,05);

$W$  – доля (частность) данного признака (0,8);

$1 - W$  – доля противоположного признака (0,2);

$N$  – численность генеральной совокупности.

На полученной выборке проводится апробация тестовых заданий (тестирование).

Ответ на тестовое задание, полностью совпадающий с эталоном считается единственно правильным.

По результатам тестирования создается матрица ответов студентов (табл. 2).

Четвертый этап работы – определение показателей качества теста (трудности и индекса дискриминативности).

Трудность теста определяется статистически с помощью индекса трудности (Итр). Итр – показатель измерения легкости и трудности тестового задания с учетом всех правильных ответов, данных испытуемыми.

Из всей выборки испытуемых выбирается «сильная» и «слабая» группа: в «сильную» группу отбираются испытуемые (1/3 от общего числа) с наиболее высокими индивидуальными баллами, в «слабую» группу (1/3 от общего числа) – с наименьшими индивидуальными баллами.

Ответы, данные в «сильной» и «слабой» группе, сводятся в табл. (табл. 3).

Вычисление индекса трудности (Итр) производится для каждого тестового задания.

Индекс трудности (Итр) вычисляют следующим образом:

$$\text{Итр} = (H + L)/n \times 100,$$

где:  $H$  – число правильных ответов в «сильной» группе;

$L$  – число правильных ответов в «слабой» группе;

$n$  – общее число испытуемых в обеих группах (1/3+1/3).

Оптимальное значение индекса трудности тестового задания составляет 50-60%, а приемлемые колебания – от 30 до 70%.

Тестовые задания со значением Итр ниже 30% и более 70% из тестовой программы исключаются (или перерабатываются с последующей дополнительной экспертизой и апробацией).

Таблица 1. Рекомендуемая численность репрезентативной выборки\*

Объем генеральной совокупности	Объем репрезентативной выборки
25	25
50	25
100	72
200	112
300	136
400	154
500	167
1000	201

\*Примечание: объем репрезентативной выборки не может быть менее 25 человек вне зависимости от объема генеральной совокупности.

Таблица 2. Матрица ответов студентов

№ тестового задания Ф.И.О. студента	1	2	3	4	5	6	7	8	..	n	Всего правильных ответов
	1. Иванов И.И.	1	1	1	0	0	0	1	1	0	
2. Сидоров С.С.	1	1	1	1	1	1	1	1	1	1	
3. Петров П.П.	0	0	0	1	0	1	1	0	0	0	
4. Николаев Н.Н.	1	0	1	0	0	1	0	1	1	0	
...											
n											
Всего ответов	правильных										
	неправильных										

Таблица 3. Сводная таблица ответов в сильной и слабой группе

Номер тестового задания	1	2	3	4	5	6	7	8	9	10
Сильная группа										
Слабая группа										

Индекс дискриминативности (Ид) показывает, насколько данное тестовое задание может отличить более подготовленных специалистов от менее подготовленных:

$$\text{Ид} = 2 \times (\text{H} - \text{L})/n$$

Значение Ид:

0,35 и выше – отличное тестовое задание

0,25-0,34 – хорошее тестовое задание

0,15-0,24 – спорное тестовое задание

ниже 0,15 - тестовое задание составлено плохо и подлежит исключению из теста.

После вычисления индексов трудности и дискриминативности тестовая программа пересматривается – из нее исключаются и перерабатываются тестовые задания с неудовлетворительными индексами Итр и Ид, и прежде всего, тестовые задания со значением Итр более 70% (легкие тесты) и Ид менее 0,25. Формулировки тестовых заданий подлежат повторной экспертизе.

Пятый этап – повторная апробация теста, которая проводится с целью окончательного определения сопряженной и прогностической валидности теста.

**Сопряженная валидность** характеризует соответствие результатов тестирования студен-

тов результатам оценки, полученным ими при использовании других форм контроля (практической, устной, письменной). **Прогностическая валидность** показывает, насколько результаты тестового контроля могут предсказывать успешность дальнейшего обучения и повышения эффективности профессиональной деятельности в разных ее аспектах.

Последний этап работы – **оформление экспертного заключения**.

Экспертное заключение включает в себя собственно лист заключения и приложение, в котором отражена информация о полученных результатах экспертизы надежности теста, индекса трудности и индекса дискриминативности по тестовым заданиям.

Прошедший экспертизу тест утверждается на заседании кафедры. Пересмотр экспертной комиссией набора, содержания тестовых заданий и эталонов ответов проводится с периодичностью в 1-2 года с целью привести в соответствие содержание теста с рабочей программой по дисциплине (при внесении в нее изменений), а также с достижениями научной и практической медицины.

**СПИСОК ЛИТЕРАТУРЫ**

1. Майоров А.Н. Теория и практика создания тестов для системы образования. М.: Интеллект-Центр, 2010. 296 с.
2. Теория и методика обучения в высшей медицинской школе: учебное пособие / И.Э. Есауленко, А.Н. Пашков, И.Е. Плотникова. Воронеж: ИПЦ «Научная книга», 2011. 384 с.

**THE PRACTICE OF EXPERT REVIEW OF TEST APPLICABILITY  
FOR CONTROL OF STUDENTS' KNOWLEDGE  
IN SAMARA STATE MEDICAL UNIVERSITY**

© 2015 E.V. Menzul, N.M. Ryazantseva

Samara State Medical University

The use of test tasks in educational process helps in monitoring the students' knowledge. Expert assessment is required in order to increase the quality of testing tasks. Expert review contributes to the quality of tests and to the quality of evaluation of students' knowledge.

*Keywords:* examination tests, validity, reliability, discriminatory power index, index of difficulty