

УДК 378.091.33:004.9

Л. С. Мороз, ассистент (БГТУ)

МЕТОДЫ ОПРЕДЕЛЕНИЯ НАДЕЖНОСТИ И ВАЛИДНОСТИ ТЕСТОВ ДЛЯ КОНТРОЛЯ ЗНАНИЙ

В статье рассмотрены такие основные качественные характеристики тестов, как надежность и валидность, описаны методы их определения по результатам тестирования. Особое внимание уделено оценке надежности теста. Представлены результаты расчета коэффициента надежности теста по формулам Кьюдера – Ричардсона, Спирмена – Брауна, Рюлона с помощью корреляционной матрицы. Описан опыт использования этих методов для определения качества разработанных автором тестов. Данная статья представляет интерес для специалистов, которые занимаются вопросами разработки и стандартизации тестовых измерительных материалов, используемых для контроля знаний студентов.

The article describes the main characteristics of test – reliability and validity; methods of determining the results of testing. Particular attention is paid to estimation of the reliability of the test. In the article presented the results of the estimation of the reliability of the test by Kyudera – Richardson's formula, Spearman – Brown's formula, Ryulona's formula, with the help of the correlation matrix. The article describes the experience of using these methods to determine the quality of tests developed by the author. This article is of interest to professionals engaged in the development and standardization of test measurement of materials used to monitor students knowledge.

Введение. Тестирование как инструмент контроля знаний становится неотъемлемой составляющей процесса обучения. Несмотря на определенные недостатки, оно является единственным по-настоящему технологичным средством для измерения уровня знаний, позволяет реализовать качественный педагогический контроль, что необходимо для эффективного управления учебным процессом.

По мнению А. Н. Майорова, тест – это инструмент, состоящий из квалитетически выверенной системы тестовых заданий, стандартизированной процедуры проведения и заранее спроектированной технологии обработки и анализа результатов, предназначенный для измерения качества и свойств личности, изменение которых возможно в процессе систематического обучения [1].

Чтобы тест адекватно оценивал знания испытуемого, он должен обладать необходимыми характеристиками. Самые важные из них надежность и валидность. Часто для контроля знаний используются тесты, не имеющие данных качеств, а значит, не совсем точно выполняющие свою функцию, и те, кто использует тесты в своей практике, даже не подозревают о необходимости анализа качества тестов.

Цель настоящей статьи – представить методы, позволяющие определить надежность и валидность тестов, используемых для контроля знаний. Данные методы применялись для оценки качества измерительных материалов при проведении итогового контроля знаний студентов I курса по дисциплине «Информатика и компьютерная графика».

Основная часть. Надежность – один из критериев качества теста, обеспечивающий

точность измерений, а также устойчивость результатов теста к действию посторонних случайных факторов.

Надежность теста тем выше, чем более согласованы результаты одного и того же человека при повторной проверке знаний посредством того же теста или эквивалентной его формы (параллельного теста).

Валидность – способность теста получать результаты, соответствующие поставленной цели. Валидность теста показывает, насколько хорошо тест делает то, для чего он был создан.

Для оценки надежности тестов используются следующие методы:

- ретестовый метод;
- метод параллельных форм;
- метод расщепления теста.

Оценка надежности строится на подсчете корреляции между двумя наборами результатов выполнения одного и того же теста или двух его параллельных форм. Чем выше корреляция, тем надежнее тест. Хорошим коэффициентом надежности теста считается тот, когда показатель колеблется в пределах $0,8 < r < 1$.

Ретестовый метод оценки надежности предполагает двукратное использование одного и того же теста в одной группе испытуемых. Он основан на подсчете корреляции индивидуальных баллов испытуемых по результатам выполнения первого и второго тестирования. Данный метод не очень удобен, так как временной фактор играет большую роль: временной промежуток не может быть большим – изменится уровень учебных достижений, но и не должен быть слишком коротким – испытуемые могут помнить задания теста и ответы на них.

Метод параллельных форм предполагает двукратное тестирование одной и той же группы испытуемых тестами, которые идентичны по содержанию, структуре, включают задания, равные по трудности, дифференцирующей способности, т. е. параллельными формами теста. Главная трудность применения этого метода в том, что разработчику надо создать новый тест, тождественный первому, и доказать идентичность, а это очень трудоемкая процедура.

Метод расщепления теста удобен в практическом применении, так как ограничивается однократным тестированием. Он основан на допущении параллельности двух половин теста и предполагает деление результатов тестирования на две части: данные по нечетным заданиям теста (X) и по четным (Y).

Коэффициент надежности вычисляется как коэффициент корреляции Пирсона по формуле

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \sqrt{\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}}}, \quad (1)$$

где x_i – индивидуальный балл i -го студента при ответе на тест по нечетным заданиям; y_i – индивидуальный балл i -го студента при ответе на тест по четным заданиям; n – число студентов.

Поскольку подсчет надежности ведется по расщепленному тесту, который в 2 раза короче, то оценка надежности корректируется по формуле Спирмена – Брауна [2]:

$$r = \frac{2r_{\text{расщ}}}{1 + r_{\text{расщ}}}. \quad (2)$$

В случае, когда выполнение задания оценивается дихотомически (1 балл – правильно; 0 баллов – неправильно), можно использовать способ оценки надежности с применением формулы Кьюдера – Ричардсона [3].

Коэффициент надежности находится по следующей формуле:

$$r = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m p_j q_j}{S_x^2} \right), \quad (3)$$

где m – число заданий в тесте; p_j – доля правильных ответов на j -е задание теста; q_j – доля неверных ответов на j -е задание, ($q_j = 1 - p_j$); S_x^2 – дис-

персия индивидуальных баллов, рассчитываемая из соотношения

$$S_x^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1},$$

где n – число студентов; x_i – индивидуальный балл i -го студента.

Еще один способ оценки надежности расщепленного теста основан на формуле Рюлона [3]:

$$r = 1 - \frac{S_d^2}{S_y^2}, \quad (4)$$

где S_y^2 – дисперсия суммарных баллов результата; S_d^2 – дисперсия разностей между результатами каждого испытуемого по обеим половинам теста. Она вычисляется по формуле

$$S_d^2 = \frac{\sum_{i=1}^n \left(x_i - y_i - \frac{\sum_{i=1}^n (x_i - y_i)}{n} \right)^2}{n-1},$$

где x_i – индивидуальный балл i -го студента при ответе на тест по нечетным заданиям; y_i – индивидуальный балл i -го студента при ответе на тест по четным заданиям; n – число студентов; $(x_i - y_i)$ – разность сумм баллов в строках с номером i субматриц с нечетными и четными заданиями [3].

Оценить надежность теста можно и с помощью следующей формулы:

$$r = \frac{m\bar{R}}{1 + (m-1)\bar{R}}. \quad (5)$$

Здесь m – количество заданий в тесте; \bar{R} – средний коэффициент корреляции заданий теста между собой [4].

Для определения валидности требуется независимый внешний критерий, т. е. оценка эксперта (преподавателя). За коэффициент валидности принимают коэффициент корреляции результатов тестовых измерений и критерия. Коэффициент валидности теста может быть рассчитан по формуле

$$V = \frac{\frac{\sum_{i=1}^n (Y_i y_i)}{n} - \bar{Y}\bar{y}}{S_Y S_y} \frac{n}{n-1}, \quad (6)$$

где Y_1, Y_2, \dots, Y_n – экспертные оценки знаний испытуемых; \bar{Y} – среднее арифметическое

экспертных оценок; S_y – стандартное отклонение этих оценок [4].

Валидность проверяется сравнением результатов тестирования испытуемого с экспертными – независимыми от этих результатов оценками уровня усвоения им материала другими методами: устным опросом, традиционной контрольной работой, экзаменом или сопоставлением этих результатов с оценками текущей успеваемости [2].

Для итогового контроля знаний 30 студентов был разработан тест, содержащий 22 тестовых задания. Математико-статистическая обработка результатов тестирования осуществлялась с помощью пакета MS EXCEL, применялась дихотомическая шкала оценивания. По результатам тестирования была сформирована матрица размерностью 30×22 , состоящая из 0 (неверный ответ) и 1 (верный ответ). Далее для расчета надежности теста потребовались следующие действия:

1) подсчитать индивидуальные баллы испытуемых (сумма баллов по строкам) X_i ;

2) найти количество правильных ответов на каждое задание (сумма баллов по столбцам) R_j ;

3) сначала упорядочить строки матрицы по убыванию величины X_i , затем столбцы – по убыванию величины R_j ;

4) подсчитать количество неправильных ответов на каждое задание W_j ($q_j = 1 - p_j$); долю правильных ответов p_j и долю неправильных ответов на каждое задание q_j ;

5) рассчитать коэффициент надежности по формуле (2).

Фрагмент матрицы представлен в табл. 1.

Таблица 1

Результаты тестирования

Порядковый номер студента	Ответы на вопросы					X_i
	1	2	3	4	5	
1	1	1	1	1	1	5
2	1	1	1	1	0	4
3	1	0	0	1	1	3
4	1	1	0	0	0	2
5	0	1	1	0	0	2
6	1	0	0	0	0	1
R_j	5	4	3	3	2	
W_j	1	2	3	3	4	
p_j	0,833	0,667	0,5	0,5	0,333	
q_j	0,167	0,333	0,5	0,5	0,667	

Далее надежность теста рассчитывалась по формулам Спирмена – Брауна и Рюлона. Бинарная матрица результатов тестирования

была разбита на две части: первую половину матрицы составили ответы по нечетным заданиям, вторую – по четным. Были подсчитаны индивидуальные баллы студентов по нечетным заданиям x_i и индивидуальные баллы по четным заданиям y_i , а также их разность в соответствующих строках. Подсчитан коэффициент корреляции Пирсона для двух половин теста по формуле (1), и полученное значение коэффициента надежности теста скорректировано по формуле Спирмена – Брауна (2). Чтобы использовать формулу Рюлона (4), вычислена дисперсия разностей между результатами каждого испытуемого по обеим половинам теста.

И, наконец, надежность теста была определена по формуле (5) с помощью корреляционной матрицы тестовых заданий, фрагмент которой приведен в табл. 2. Корреляционная матрица представляет собой квадратную матрицу размерности 22×22 , симметричную относительно главной диагонали. В ячейках таблицы содержатся коэффициенты корреляции тестовых заданий между собой, вычисленные по формуле (1). По таблице был определен средний коэффициент корреляций для каждого задания, затем средний коэффициент корреляций по всему тесту \bar{R} .

Таблица 2

Корреляционная матрица

Порядковый номер студента	Значения коэффициентов корреляции				
	1	2	3	4	5
1	1,000	0,024	0,024	0,279	0,398
2	0,024	1,000	0,148	0,373	-0,146
3	0,024	0,148	1,000	0,053	0,167
4	0,279	0,373	0,053	1,000	-0,098
5	0,398	-0,146	0,167	-0,098	1,000

Результаты расчета коэффициента надежности теста с помощью разных методов представлены в табл. 3.

Таблица 3

Коэффициенты надежности теста

Используемая формула	Значение
По формуле Кьюдера – Ричардсона	0,7023
По формуле Спирмена – Брауна	0,7544
По формуле Рюлона	0,7538
С помощью корреляционной матрицы	0,7238

Нижним пределом значения коэффициента надежности принято считать показатель 0,7. Если значение коэффициента надежности ниже, то надежность теста считается неудовле-

творительной, так как возникает большая погрешность измерений.

Заключение. Анализ результатов свидетельствует о том, что надежность разработанного теста является минимально допустимой, некоторые тестовые задания необходимо заменить.

Качество диагностических материалов является одним из ключевых моментов контроля знаний. Во избежание неадекватной оценки знаний необходимо использовать только статистически обоснованные тестовые материалы, обладающие достаточным уровнем надежности и валидности.

Литература

1. Майоров, А. Н. Теория и практика создания тестов для системы образования / А. Н. Майоров. – М.: Интеллект центр, 2001. – 296 с.

2. Крисилов, В. А. Методика анализа педагогических тестов по результатам тестирования / В. А. Крисилов, Т. В. Онищенко, Н. В. Русинова [Электронный ресурс]. – Режим доступа: http://www.nbu.gov.ua/Portal/natural/Popu/2007_2/5-3.pdf. – Дата доступа: 06.01.2010.

3. Материалы к семинару по обработке результатов тестирования [Электронный ресурс]. – Владивосток, 2001. – Режим доступа: http://kpmit.wl.dvgu.ru/library/aschepkova_testing17102001/index.phtml. – Дата доступа: 12.01.2010.

4. Ким, В. С. Тестирование учебных достижений: монография / В. С. Ким. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.

Поступила в редакцию 31.03.2010