DOI: 10.17117/na.2016.06.01.323

<http://ucom.ru/doc/na.2016.06.01.323.pdf>

Поступила (Received): 02.06.2016

Муратова Л.А.**Валидность и дискриминативность при исследовании и оценке качества теста «Интегральное исчисление»****Muratova L.A.****Validity and diskriminativnost at research and to assessment of quality of the Integral calculus test**

В статье проводится анализ качества теста «Интегральное исчисление» согласно классической теории тестирования. Рассматриваются два показателя: валидность и дискриминативность. Дается оценка качества теста на основании известных требований к предъявляемым характеристикам, предлагаются рекомендации по совершенствованию теста

Ключевые слова: тест, качество, валидность, дискриминативность, точечный бисериальный коэффициент корреляции

In article the analysis of quality of the Integral calculus test according to the classical theory of testing is carried out. Two indicators are considered: validity and diskriminativnost. The assessment of quality of dough on the basis of known requirements to the shown characteristics is given, recommendations about improvement of dough are offered

Key words: test, quality, validity, diskriminativnost, dot biserial coefficient of correlation

Муратова Лидия Александровна

Кандидат технических наук, доцент
Самарский государственный технический университет
г. Самара, ул. Молодогвардейская, 244

Muratova Lidiya Aleksandrovna

Candidate of Technical Sciences, Associate Professor
Samara state technical university
Samara, Molodogvardeyskaya st., 244

Объективный контроль знаний с помощью тестирования является необходимым условием оптимизации учебного процесса в вузе. В настоящее время тест представляет собой единственный, хоть и не идеальный технологический инструмент измерения результатов педагогического процесса [5].

Составление теста предполагает математическую обработку полученных результатов, определение качества теста и последующую коррекцию теста с учетом его недостатков. При оценке качества по классической теории тестирования, как правило, исследуют такие характеристики теста, как надежность, валидность и дискриминативность. Остановимся на анализе валидности и дискриминативности.

В качестве примера рассмотрим тест «Интегральное исчисление» (табл. 1) курса высшей математики. Тест охватывает все темы раздела, предусмотренные программой курса; включает в себя различные по сложности задания; используется для текущего контроля знаний студентов первого курса СамГТУ. Это

тест закрытого типа, он содержит 10 заданий с выбором единственного правильного ответа из пяти предложенных. Для обработки результатов тестирования были выбраны работы студентов нефтетехнологического факультета университета. В тестировании участвовали 235 студентов. Работы студентов, ответивших на все вопросы правильно и тех, кто ответил на все вопросы неправильно, были исключены из выборки. В результате выборка составила 203 работы.

Таблица 1. Интегральное исчисление

1. Вычислить: $\int_0^{\frac{\pi}{2}} \frac{\cos x}{1 + \sin^2 x} dx$.	2. Вычислить: $\int_1^3 x \cdot \ln x dx$.	3. Вычислить: $\int_0^{\frac{\pi}{4}} \sin^3 2x dx$.	4. Вычис-
лить: $\int \frac{x-3}{\sqrt{x^2-10x+2}} dx$.	5. Вычислить: $\int_0^{\frac{\pi}{2}} \frac{dx}{\sin x + \cos x + 1}$.		
6. Несобственный интеграл $\int_0^{\infty} \cos x dx$ (указать все правильные ответы)			
А) сходится; Б) расходится; В) $= \infty$; Г) $\neq \infty$; Д) от неограниченной функции.			
7. Изменить порядок интегрирования $\int_2^4 dy \int_y^4 f(x, y) dx$.			
8. Вычислить: $\int_0^2 dy \int_{\sqrt{2y-y^2}}^2 x dx \int_0^3 z^2 dz$.	9. Вычислить $\int_L \frac{y^2 dl}{9x}$, где L – отрезок прямой $y = 3x$		
между точками $A(1; 3)$ и $B(2; 6)$.	10. Найти работу силы $\vec{F} = y\vec{i} + (y + x^2)\vec{j}$ при перемеще-		
нии точки по дуге кривой $x = t, y = t^2, 0 \leq t \leq 1$			

Одним из важнейших показателей качества теста является его валидность – понятие, указывающее, что тест измеряет и насколько хорошо он это делает [1]. При этом единой характеристики валидности не существует. В данном случае рассмотрим числовую характеристику коэффициента валидности – точечный бисериальный коэффициент корреляции r_{pb}^j . Это коэффициент корреляции каждого задания с тестовым баллом студента (индивидуальным баллом). Его величину можно рассчитать по формуле [4]

$$r_{pb}^j = \frac{\bar{X}_1 - \bar{X}_0}{s_x} \sqrt{\frac{n_1 \cdot n_0}{n(n-1)}},$$

где n_1 – число студентов, выполнивших данное задание; n_0 – число студентов, не выполнивших его; $n = n_1 + n_0$ – общее количество студентов; $\bar{X}_1$ – средний индивидуальный балл студентов, справившихся с данным заданием; $\bar{X}_0$ – средний индивидуальный балл студентов, не справившихся с данным заданием; s_x – стандартное отклонение для индивидуальных баллов всех студентов.

Полученные в результате расчетов значения r_{pb}^j для каждого из заданий, расположенных в порядке убывания количества правильных ответов, представлены в табл. 2.

Таблица 2. Коэффициенты валидности и дискриминативности заданий

№	2	1	7	8	3	4	6	10	5	9
r_{pb}^j	0,535	0,533	0,53	0,57	0,494	0,526	0,417	0,511	0,314	0,547
$r_{диск}^j$	0,636	0,636	0,673	0,764	0,582	0,636	0,527	0,618	0,364	0,6

Задания, для которых $r_{pb}^j \leq 0$, считаются несостоятельными и непригодными для контроля знаний, поэтому подлежат удалению из теста; задания же, для которых $0 < r_{pb}^j < 0,3$, являются кандидатами на удаление из теста [5]. Таких заданий в тесте нет. Некоторые исследователи говорят о том, что задание можно считать валидным, когда $r_{pb}^j \geq 0,5$ [2]. В рассматриваемом тесте с точки зрения этого требования самым неудачным следует считать 5-е задание ($r_{pb}^j = 0,314$). Задание достаточно сложное для среднего студента, плохая корреляция, очевидно, связана с угадыванием ответа. 6-е задание также имеет низкий коэффициент корреляции с тестовым баллом ($r_{pb}^j = 0,417$), но если его исключить, пропадет возможность проверить степень усвоения такой темы, как несобственные интегралы. Существует мнение [5], что если тестовое задание уникально, не надо спешить удалять его из теста, даже если $r_{pb}^j = 0,1$. Тем более, что общий коэффициент валидности (среднее арифметическое коэффициентов r_{pb}^j) $r_{pb} = 0,498$ имеет среднее, близкое к высокому значение при градации [6]: 0,2 – 0,3 – низкое; 0,3 – 0,5 – среднее; свыше 0,5 – высокое. Не оценивая надежность исследуемого теста, отметим, что тест с высокой валидностью имеет и высокую надежность. Если тест имеет низкую валидность, то применять его нельзя, даже если он имеет высокую надежность [4].

Следующая характеристика для проверки качества теста – дискриминативность – способность задания дифференцировать студентов на лучших и худших [2, 3, 5]. Если сравнивать этот критерий с надежностью и валидностью, то он не является главным фактором, но если и коэффициент валидности данного задания, и его коэффициент дискриминативности не отвечают заданным критериям, то такое задание необходимо удалить из теста.

В качестве коэффициента дискриминативности можно принять точечный бисериальный коэффициент корреляции r_{pb}^j [2, 5]. Задания, для которых $r_{pb}^j < 0,2$, необходимо удалить из теста [2]. Различающая способность у качественного теста считается достаточной, если $r_{pb}^j > 0,3$ [3, 5]. В рассматриваемом тесте все значения r_{pb}^j соответствуют требованиям экспертов.

Второй способ вычисления коэффициента дискриминативности $r_{\text{диск}}^j$ состоит в применении формулы [2, 3, 5]:

$$r_{\text{диск}}^j = p_1^j - p_0^j,$$

где p_1^j – доля правильных ответов на j -задание у 27% «лучших» студентов по результатам выполнения теста; p_0^j – доля правильных ответов на j -задание у 27% «худших» студентов по результатам выполнения теста.

Значения индекса $r_{\text{диск}}^j$ заведомо лежат в интервале $[-1; 1]$. Если для некоторых заданий $r_{\text{диск}}^j \leq 0$, то такие задания подлежат рассмотрению к удалению из теста. Тест считают качественным, если $r_{\text{диск}}^j > 0,3$ [3]. Результаты вычислений, представленные в табл.2, дают основание считать, что рассматриваемый тест и по этому показателю имеет хорошую различающую способность.

Таким образом, в результате проведенных исследований качества теста «Интегральное исчисление» можно сделать следующие выводы:

общий коэффициент валидности $r_{pb} = 0,498$ имеет среднее значение и удовлетворяет требованиям экспертов;

не все точечные бисериальные коэффициенты корреляции, которые брались в качестве оценки коэффициента валидности, удовлетворяют требованиям экспертов, в частности, $r_{pb}^5 = 0,314$, $r_{pb}^6 = 0,417$;

различающая способность теста – дискриминативность оказалась в рамках допустимой.

В качестве рекомендаций по улучшению качества теста предлагается убрать 5-е задание и изменить 6-е, сохранив его тематику.

Список используемых источников:

1. Анастаси А. Психологическое тестирование. Кн.1. М.: Педагогика, 1982. 320 с.
2. Звонников В.И., Чельшкова М.Б. Современные средства оценивания результатов обучения. М.: Академия, 2007. 224 с.
3. Карпенко А.П., Домников А.С., Белоус В.В. Тестовый метод контроля качества обучения и критерии качества образовательных тестов // Наука и образование: электронное научно-техническое издание. Вып. №04. 2011. 28 с.
4. Ким В.С. Тестирование учебных достижений. Уссурийск: Изд-во УГПИ, 2007. 214 с.
5. Олейник Н.М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения. Донецк, ДонГУ, 1991. 168 с.
6. Психологическая диагностика. М.: Изд-во УРАО, 1997. 304 с.